JOIV : Int. J. Inform. Visualization, 6(4) - December 2022 771-777

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Application of ARIMA Kalman Filter with Multi-Sensor Data Fusion Fuzzy Logic to Improve Indoor Air Quality Index Estimation

Bayu Erfianto [a,*], Andrian Rahmatsyah [a]

[a] *Department of Informatics, Telkom University, Jl. Telekomunikasi No 1, Kabupaten Bandung, 40257, Indonesia*
*Corresponding author: *erfianto@telkomuniversity.ac.id*

***Abstract*— Air quality monitoring is a process that determines the number of pollutants in the air, one of which is indoor air quality. The Fuzzy Indoor Air Quality Index was developed in this research. It is a method for determining the indoor air quality index using sensor fusion and fuzzy logic. By combining several different time series determinants of air quality, a fuzzy logic-based sensor fusion method is used to build a knowledge base about indoor air quality levels. Without the use of complicated calculation models, fuzzy logic-based fusion will make it easier to determine indoor air quality levels based on various sensor parameters. The input for fuzzy-based data fusion is obtained from the ARIMA method with Kalman Filter's air quality parameter values estimation. The application of ARIMA with a Kalman Filter was used to improve the accuracy of indoor air quality estimation in this study. ARIMA(3,1,3) had a MAPE of 0.1 percent on the CO2 dataset, and ARIMA(1,0,1) had a MAPE of 0.63 percent on the TVOC dataset based on approximately three experimental days. ARIMA (3,1,3) estimation with a Kalman Filter results in a MAPE of 0.03 percent for the CO2 dataset and a MAPE of 0.24 percent for ARIMA(1,0,1) Kalman Filter estimation on TVOC dataset. As a result, the Fuzzy Indoor Air Quality Index (FIAQI) developed in this research reasonably estimates indoor air quality. This can be seen by examining the percentage of estimation errors obtained from the experiment.**

***Keywords*— Sensor data fusion; fuzzy logic; air quality index; prediction; time series.**

Manuscript received 16 Jun. 2022; revised 23 Jul. 2022; accepted 28 Aug. 2022. Date of publication 31 Dec. 2022.

I. INTRODUCTION

Indoor air quality has attracted the interest of researchers from various disciplines, including public health, environmental governance, and industry. They have conducted extensive research to improve the health, comfort, and overall well-being of building occupants in commercial buildings, single-family homes, or multi-family rooms [1]. Volatile organic compounds are one of the compounds that significantly contribute to the decline in indoor air quality (hereinafter abbreviated as VOC). VOCs are organic chemical compounds with a molecular structure that enables them to evaporate under normal indoor atmospheric temperature and pressure conditions [2], [3]. This is the most used definition of VOC in scientific literature and can also be used to describe indoor air quality. According to studies, the concentrations of certain organic matter in VOCs are, on average, two to five times higher indoors than outdoors. Numerous activities people perform in their homes generate VOCs that are hazardous to their health [4].

Low concentrations of VOC compounds are typically found in indoor air but can also be found in certain situations. It truly depends on the method used to determine the VOC concentration. All available measurement methods are insufficient to quantify all VOC compounds accurately. For example, formaldehyde HCHO and other similar compounds are quantified using a different method than benzene and toluene. The sensitivity of the measurement, as well as its selectivity and bias, are determined by the complexity of measurement instruments and analytical methods. The concept of total volatile organic compounds (TVOC) then attempts to overcome this practical constraint by providing simple measurements for the aggregate of all volatile organic compounds without regard for individual chemicals.

Air pollution was quantified using the Air Quality Index, which was calculated by fusing CO2 and TVOC data [5]. Time series analysis aims to model and predict physical systems using various models, and TVOC and CO2 are the primary pollutants contributing to serious indoor health problems. Time series analysis and prediction have become an integral part of system identification in everyday life due to the rapid advancements in the Internet of Things [6].

Forecasts are time-dependent predictions, which means they are more precise when dealing with time series data instead with data point prediction. The Autoregressive Integrated Moving Average (ARIMA) model is a class of time series models that produces accurate forecasting results and is widely used by researchers [7]–[9].

This research proposes an indoor air quality index (Fuzzy Indoor Air Quality Index) based on the ARIMA Kalman Filter-based Time Series method, which is used to estimate and forecast indoor air quality parameters over time. This research is motivated by [10] that uses a derivative of Kalman Filter (fractional order EKF) to fuse humidity sensors with temperature to identify the effect of the pollutant on both parameters. The advantage of combining the FIAQI sensor data with the ARIMA Kalman Filter is that it can accommodate a variety of sensor parameters with estimation outputs to improve time series-based air quality estimation based on these multi-parameter air quality data. It is hoped that the estimation and forecasting results will be more accurate when compared to using ARIMA only. When compared to using only the standard ARIMA model, the ARIMA Kalman Filter model (hereinafter referred to as ARIMAKF) can help reduce TVOC and CO2 estimation errors.

Additionally, the ARIMAKF results are used as input for fuzzy logic-based sensor fusion to generate an Indoor Air Quality Index or FIAQI in this research (Fuzzy Indoor Air Quality Index). This research contributes to the generation of a dataset that records TVOC and CO2. However, other parameters, such as HCHO and indoor temperature levels, were recorded but not used for FIAQI purposes.

The structure of this paper is written as follows. The introduction section contains an introduction that discusses the context of the research and the research's contribution. The material and methods section discusses the methods and modelling used in the research, specifically the ARIMA KF development and the development of a fuzzy logic-based data fusion model. The results and Discussion section discuss the analysis of the results of the conducted experiments. The conclusion section presents the summary of the conducted research.

II. MATERIALS AND METHOD

A. Principle of Indoor Air Quality Index

Air quality can be monitored and evaluated using parameters such as carbon monoxide (CO), carbon dioxide (CO2), formaldehyde (HCHO), nitrogen dioxide (NO2), ozone (O3), sulfur dioxide (SO2), total volatile organic compounds (TVOCs), particulates (PM2.5), and total suspended particles (TSP), as well as temperature, relative humidity, and air movement [11]. Another indoor quality index is defined as an additional index to common IAQI, which is used for indoor thermal comfort related to health conditions [12]. Indoor air quality is also influenced by household chemicals, furnishings, outdoor air contaminants, and occupant activities (e.g., smoking, cooking, and breathing) [13]. Due to the high pollutant volume, developing an accurate method for monitoring IAQI presents an intriguing challenge. Combining heterogeneous data from networks with a variety of sensor types is critical for calculating IAQI and monitoring various pollutants, most notably indoor air quality. We summarized recent research on the classification of AQI and IAQI for a variety of pollutants, as well as several forecasting methods used to generate IAQI, in this paper. The latest research on fuzzy logic and forecasting techniques for monitoring AQI and IAQI is summarized in Table I. This research is relevant to the research presented in this article.

TABLE I
RELATED RESEARCH IN INDOOR AIR QUALITY INDEX

Author	Problem / Objective	AQI Classification Methods	Forecasting Methods
[13]	Measurement of everyday AQI of outdoor environment using 6 pollutants (O3, No2, SO2, CO, PM10, PM2.5)	Sigma operator, Fuzzy Classification	AR with Moving average smoothing
[14]	Measurement of everyday AQI of outdoor environment using 6 pollutants (O3, NO2, SO2, CO, PM10, PM2.5)	Gaussian Fuzzy number	
[5]	Monitoring and control EIAQI value expressed in four pollutants: **CO2**, Thermal, O3, **TVOC**, PM2.5	Fuzzy Logic controller	
[15]	Measurement of everyday AQI of outdoor environment using 6 pollutants (O3, No2, SO2, CO, PM10, PM2.5), **TVOC**	Fuzzy Logic	
[16]	Measurement of everyday AQI using CO,NH3,O3		Comparing ARIMA with LSTM
This paper	To monitor Indoor Air Quality (IAQ): **CO2, TVOC**	Fuzzy Logic	ARIMA with Kalman Filter

According to Javid et al. [15] and the remaining summary in Table 1, TVOC and CO2 are gases that are frequently used to construct IAQI. Additionally, some studies use Fuzzy Logic to classify the various types of gas parameters [14], [17], [18], but only a few IAQI studies employ forecasting techniques, either statistical or machine learning based. TVOC and CO2 were used as gas parameters in this study to calculate the IAQI. Another interesting technique is proposed by Colella et al. [19] to monitor indoor air quality for special purposes, i.e., for hospital surgical rooms. The combination of the ARIMA method and the Kalman Filter is novel in this study, which aims to reduce estimation error compared to the ground truth. Of course, some previous studies have used the ARIMA method in conjunction with the Kalman Filter, but not to develop IAQIs, as demonstrated in research on rainfall forecasting [20].

B. Methods

The methods used in this research can be seen in Fig 1. The description of each block in Fig. 1 is explained in the followings.

1) Acquiring Time Series Data for TVOC and CO2: A web-based application was developed to collect data from the WP6003 sensor device in this research. The IAQI is constructed using all raw data from sensors, but only total volatile organic compounds (TVOC) and CO2 or CO2 estimates.

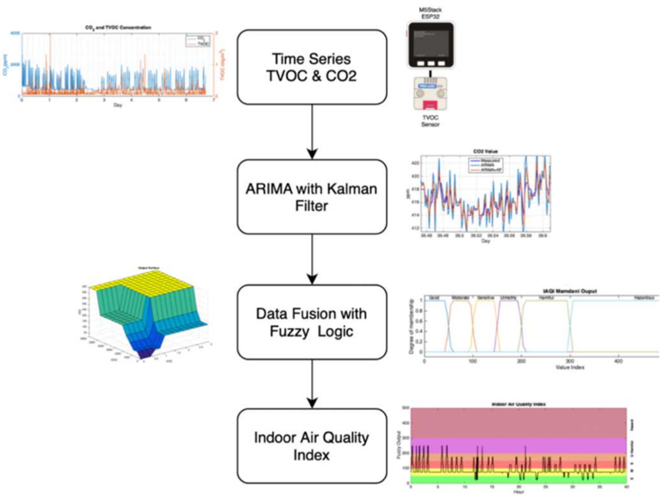

Fig. 1 Methods used in the research

The SGP30 series air quality sensor was used in this research because it is integrated with the WP6003 chip and comes in plastic packaging. It can measure temperature, humidity, TVOC, CO2, and HCHO. Only TVOC and CO2 were processed in this study with a sensing rate of at least 5 seconds to produce a reasonably representative indoor air quality index. The volatile organic compounds (TVOC) group is defined as the primary consideration. Although the air pollutant criterion is the most significant outdoor air pollutant, the indoor concentration of this compound is between 20% and 80% greater than the outdoor concentration. In the case of VOCs, it has been observed that indoor concentrations of these parameters are generally higher than corresponding amounts in the environment [21]. Fig. 2 shows the TVOC and CO2 plots dataset, recorded in 74 hours.

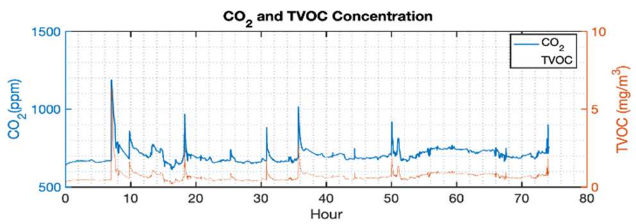

Fig. 2 Data Plot of 74 hours recording. Example of short experiment dataset was recorded on 13 -17 April 2022

2) ARIMA Model with Kalman Filter: The data is then processed using the ARIMA time series model as indoor air quality parameters. The ARIMA method generally employs the Box-Jenkins, which consists of four stages: identification, parameter estimation, scenario examination, and forecasting time series data for CO2 and TVOC values. After obtaining the estimation results, the Kalman Filter is used to combine the ARIMA output with the best model, resulting in a smoother forecast than using only the ARIMA model. This section focuses on TVOC and CO2 data modeling for indoor environments (homes) in the East Bandung area. The dataset was collected from within the house between April 13 and 17, 2022, using the WP6003 sensor device. Daily data are collected to create a time series data set. Two methods were used for offline data analysis: the ARIMA method and the Kalman Filter method. The description of ARIMA modeling results is described in the next section.

3) Data Fusion using Fuzzy Logic: The data fusion method with Fuzzy Logic begins with the creation of an input membership function based on the CO2 and TVOC gas parameters. Numerous studies have concluded that elevated levels of carbon dioxide in the home affect cognitive performance and human scenarios. The average indoor CO2 concentration is between 600 and 1000 parts per million but can exceed 2000 parts per million with increased occupancy and inadequate ventilation. Exposure to levels greater than 1000 parts per million causes declining cognitive capability, whereas levels greater than 2000 parts per million have been linked to inflammation, renal calcification, bone demineralization, and endothelial dysfunction. The following Table II summarizes the CO2 quality indexes used in the membership function scenario.

TABLE II
CO2 LEVEL OF CONCERN

Level of Concern	CO2 Level ppm	Description
Excellent	0 – 400	The air inside is as fresh as the air outside
Fine	400 – 1000	The air quality inside remains at harmless levels
Moderate	1000 – 1500	The air quality inside has reached conspicuous levels
Poor	1500 – 2000	The air quality inside has reached precarious levels
Very Poor	2000 – 5000	The air quality inside has reached unacceptable levels
Severe	> 5000	The air quality inside has exceeded the maximum workplace

Meanwhile, TVOC concentrations can be expressed in terms of grams per cubic meter (g/m3) of air (or milligrams per cubic meter (mg/m3), parts per million (ppm), or parts per billion (ppb). A TVOC concentration of less than 0.3 mg/m3 is low. Additionally, levels between 0.3 mg/m3 and 0.5 mg/m3 are considered acceptable. The TVOC Table II contains the values for the TVOC concentrations used to calculate the Air Quality index.

TABLE III
LEVEL OF CONCERN TVOC

TVOC Level mg/m3	Level of Concern
< 0.3	Low
0.3 to 0.5	Acceptable
0.5 to 1	Marginal
1 to 3	High

Both Table II and Table III are used to construct Fuzzy membership function, as depicted in Fig. 3 and Fig. 4.

Afterward, the Fuzzy Logic inference rule is constructed using the two level of concern in Table II and Table III. The followings are six examples of 24 constructed rules (rule number and conditions).

Rule 1: If (**VOC is Low**) and (**CO2 is Excellent**) then (**IAQ is Good**))
Rule 2: If (**VOC is Low**) and (**CO2 is Fine**) then (**IAQ is Moderate**)
Rule 9: If (**VOC is Acceptable**) and (**CO2 is Moderate**) then (**IAQ is Sensitive**)
Rule 10: If (**VOC is Acceptable**) and (**CO2 is Poor**) then (**IAQ is Unhealthy**)
Rule 11: If (**VOC is Acceptable**) and **(CO2 is VeryPoor**) then (**IAQ is Harmful**)
Rule12: If (**VOC is Acceptable**) and (**CO2 is Severe**) then (**IAQ is Hazardous**)

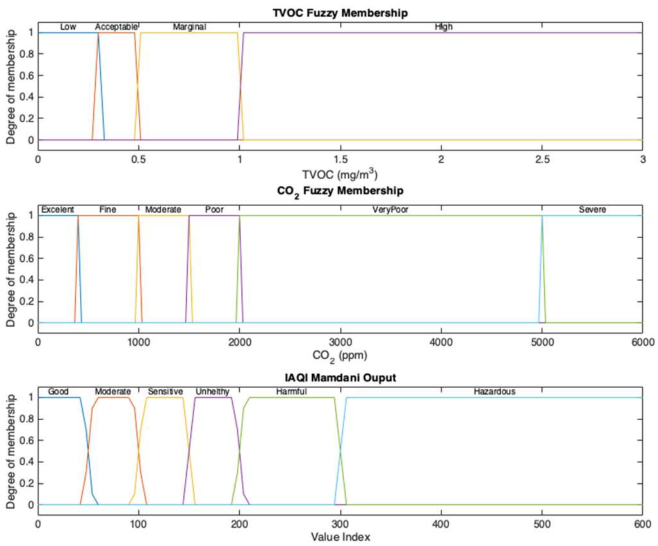

Fig. 3 Input membership function dan output membership function

The final step in the fuzzy inference system is defuzzification, which involves calculating the air quality index using the centroid method. The centroid function is the most frequently used and physically appealing method for defuzzification. Fig. 4 illustrates defuzzification process. The centroid method returns the center of the region defined by the fuzzy output function's area under the curve. The inference process is used to generate a decision about the indoor air quality index in the form of a fuzzy logic output.

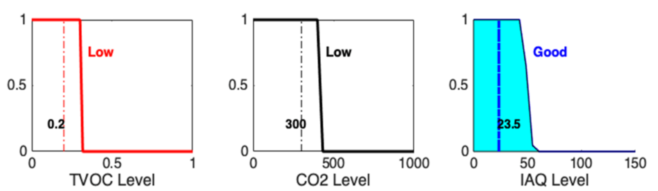

Fig. 4 Input membership function and the output of defuzzification

Fig. 4 shows the fuzzification process, where the input of TVOC is measured 0.2 mg/m3, CO2 level is 300 ppm, and the output of Fuzzy is 23.5 (classified as good air quality). Because the input dataset for fuzzy inference is a time series, the output is also a time series graph with a quality limit index that is color-coded according to EPA standards. The complete results are presented in the next section.

III. Result and Discussion

The previous chapter described the procedure for acquiring TVOC and CO2 data from the SGP30 sensor. A simple Javascript-based application has been developed and ported to the macOS web server to read data from sensors to facilitate data acquisition. The WP6003 is equipped with an SGP30 sensor chip and an ESP32 series microcontroller capable of wireless and Bluetooth communication. However, for the purposes of this study's data collection, it is sufficient to use only Bluetooth. Fig. 5 illustrates the architecture of the device's data reading or acquisition application.

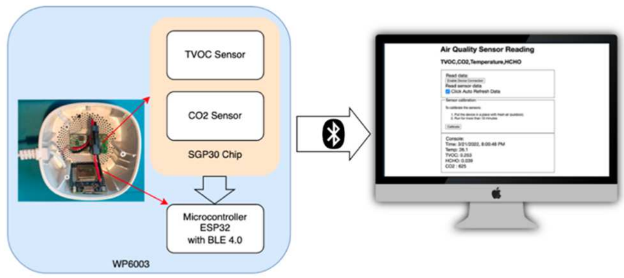

Fig. 5 Architecture and desktop application to read data from sensors

A. Data Preparation

As previously stated, commercial devices are used as a reference for data acquisition results from the SGP30 sensor in the WP6003 device. Fig. 6 illustrates the difference in reader results between the ground truth device and the device used (WP6003). The results of approximately 2.5 minutes of sensor readings indicate that the standard deviation of the error between the ground truth device and WP6003 for CO2 is approximately 13.18 ppm and for TVOC is approximately 0.0087 mg/m3. Because the error difference during the reading process is quite small, the WP6003 is sufficiently accurate to be used in this study in comparison to the commercial version of the device.

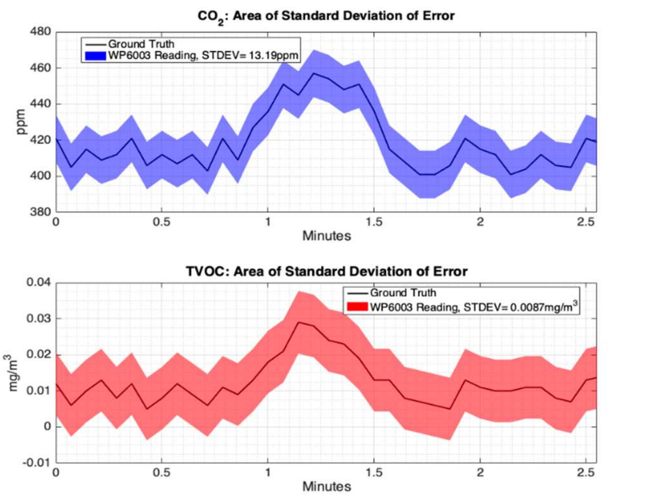

Fig. 6 Error in readings between the sensor devices used (WP 6003) compared to the commercial sensor IAQI JD032 series

After that, the data preparation process begins. After the sensor data is acquired, it is saved in CSV format for offline processing and analysis using Matlab. Fig. 7 illustrates the dataset fragments derived from TVOC and CO2 sensor readings. The dataset for this example is drawn from two scenarios: ideal or healthy conditions and unhealthy conditions. To create an unhealthy environment, insecticide aerosol is added. As a result, unhealthy air conditions will develop, and the sensor will be able to distinguish between healthy and unhealthy air conditions.

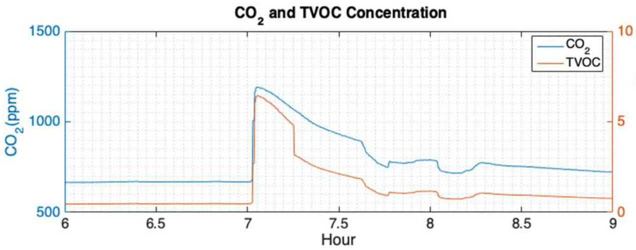

Fig. 7 Two condition applies: normal indoor air quality (hour 6 to 7), adding aerosol changes the normal condition to high polluted indoor air (hour 7 to 8).

When dealing with serially correlated data, time series models are extremely useful. Air pollution is quantified in this study using a parameter called the Air Quality Index. Time series analysis is concerned with modeling and physical forecasting systems using various models to gain control over the system. CO2 and TVOC are classified as major pollutants, among others, because they have the potential to cause serious health problems. In time series analysis, estimation entails determining the optimal parameter using historical data, whereas prediction utilizes the data to determine the random value of previously unseen data. As a result, estimation is associated with model construction, that is, determining the most appropriate parameter that best describes the distribution of historical data. The following section discusses how to estimate time series parameters with the ARIMA model.

B. ARIMA Model with Kalman Filter

After the time series data is acquired, it is saved in CSV format for offline processing and analysis using Matlab. Fig. 2 shows the dataset obtained from TVOC and CO2 sensor readings. The dataset for this example is drawn from two scenarios: ideal or healthy conditions and unhealthy conditions. To create an unhealthy environment, insecticide aerosol is added. As a result, unhealthy air conditions will develop, and the sensor will be able to distinguish between healthy and unhealthy air conditions. This section focuses on TVOC and CO2 data modeling for indoor environments (homes) in the East Bandung area. Data collection took place between April 13 and 17, 2022, using the WP6003 sensor device. Daily data are collected to create a time series data set. Two methods were used for offline data analysis: the ARIMA method and the Kalman Filter method. The ARIMA method is based on the Box-Jenkins methodology, which is divided into four stages: identification, parameter estimation, diagnostic examination, and forecasting.

The first step in the identification stage is to determine whether the time series data in Fig. 2 is stationary. The Augmented Dickey-Fuller test can be used to determine stationarity in the mean (ADF test). The ADF value is -0.27358, the critical value is -1.9416, indicating that the ADF value is less than the critical value and the p-value of 0.55167 is greater than the threshold (0.05). As a result, we can state categorically that the CO2 time series is not stationary. Meanwhile, the ADF value for the TVOC dataset is -3.4222, while the critical value is -1.9416, indicating that the ADF value is greater than the critical value, but the p-value of 0.001 is less than the threshold (0.05). As a result, we can conclude that the time series of the TVOC is stationary. The autocorrelation function (ACF) and partial autocorrelation function are then plotted (PACF). The ACF and PACF plots can be used to determine the autoregressive (AR) and moving average (MA) models. After obtaining the ARIMA model (p, d, q), the next step is to estimate the parameters of the AR and PCF models. Autocorrelation and partial autocorrelation plots can be used to visualize this process, as depicted in Fig. 8.

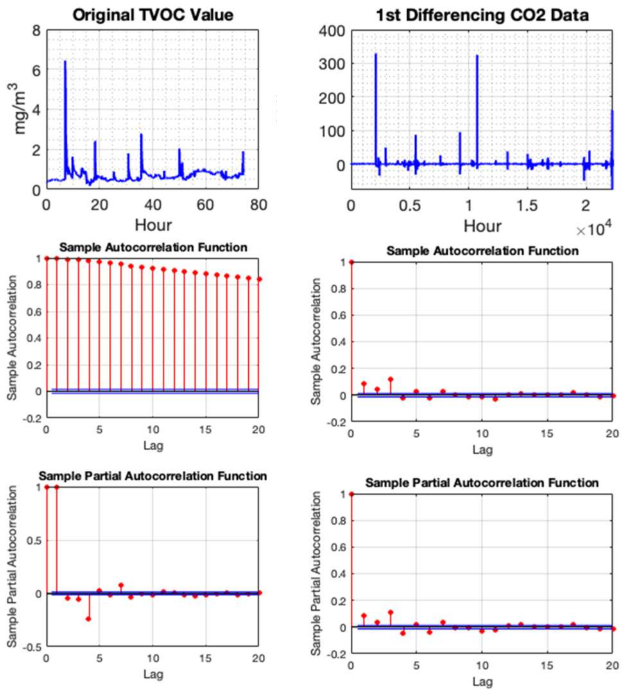

Fig. 8 Left: TVOC uses order 0 or data origin to determine AR and MA. The spike on PACF is at lag 1 and ACF is at lag 1, thus AR=1 and MA=1. Right: while for CO2, a differencing process must be carried out to obtain ACF and PACF. Spike on PACF is at lag 3 and ACF is at lag 3, thus AR=3 and MA=3

The results of the stationary process with constant mean and variance are shown in Fig. 8. Because the ADF test indicates that the TVOC time series dataset is stationary, the dataset used to determine the AR and MA is order 0 or the origin dataset. For the TVOS dataset, the PACF spike occurs at lag 1, and the ACF spike occurs at lag 1, implying that AR=1 and MA=1 with I=0, implying that the model obtained is ARIMA (1,0,1). After the first differencing process on the CO2 dataset, the spike in PACF occurs at lag 3 and the spike in ACF occurs at lag 3, resulting in AR = 3 and MA = 3 with I = 1, implying that the model obtained is ARIMA (3,1,3).

After obtaining the ARIMA (1,0,1) model, the estimation and forecasting algorithm must be run. This ARIMA estimation will be further processed using the Kalman Filter algorithm to improve the results of the ARIMA prediction estimation, as depicted in Fig. 9.

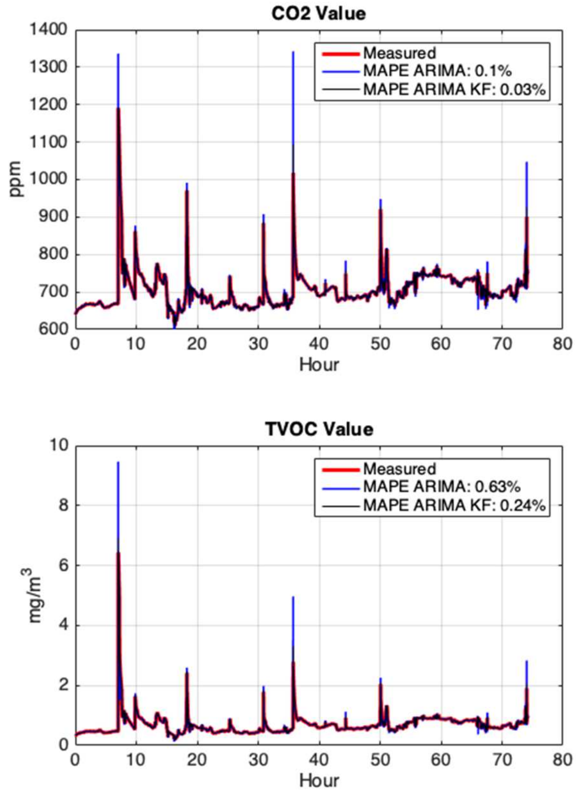

Fig. 9 Comparison between datasets with ARIMA and ARIMA Kalman Filter. It looks like ARIMA Kalman Filter is smoother than ARIMA

After completing all stages of the Kalman filter, the ARIMA-Kalman filter is used to forecast the next twelve months. The results of the ARIMA estimation with the ARIMA Kalman Filter are compared to the actual data in Figure 9. The figure demonstrates that the ARIMA-Kalman Filter estimation results are more accurate than the ARIMA estimation results. For the CO2 dataset, the MAPE value is 0.1 percent for the ARIMA (3,1,3) model and 0.03 percent for the ARIMA (3,1,3) Kalman Filter method. Meanwhile, the MAPE value for TVOC data is 0.63 percent for the ARIMA (1,0,1) method and 0.24 percent for the ARIMA (1,0,1) Kalman Filter method. Thus, as evidenced by the mean absolute percentage error, using ARIMA KF can help reduce the time series dataset estimation error (MAPE).

The following stage is a diagnostic examination, which is divided into two parts: a parameter significance test using the Q-test and a model suitability test that includes a test of the residual white noise and normal distribution assumptions. The Ljung-Box test in MATLAB can determine whether the residual white noise is autocorrelated, while the Kolmogorov-Smirnov test can determine whether the residuals are normally distributed or not. The Ljung-Box test results for the CO2 and TVOC datasets are h0 = 1, indicating that the null hypothesis is rejected with the explanation that at least one significant autocorrelation exists in lags 1 to 20. Meanwhile, the residual test with Kolmogorov-Smirnov indicates that h0 = 1 for CO and TVOC results. The test accepts a single hypothesis with a 95% confidence interval, assuming that the residuals for the two-time series are normally distributed.

C. The Output of Fuzzy Indoor Air Quality Index

As previously stated, the fuzzy inference technique used is Fuzzy Mamdani. The Mamdani inference system's output must be defuzzified to obtain crisp values. The mean of maxima method, the center-of-gravity (COG) method, the weighted average method, and the max method have all been used for defuzzification. The COG method was used in this study because it was the simplest to apply.

The output of fuzzy IAQI (FIAQI) data from the dataset obtained is shown in Fig. 10. G (Good or Excellent), M (Moderate), S (Sensitive), U (Unhealthy), Harmful, and Hazzard are the FIAQI scale or index. The monitoring results indicated that the air quality was at a hazardous level. Insecticide aerosols are sprayed into the room to prevent this. Meanwhile, hazards cannot be detected because the FIAQI output is between 30 and 250, which cannot be obtained using the scenario described previously. This is because the level of conditioned aerosol does not exceed the hazardous level as defined by the Fuzzy IAQI index.

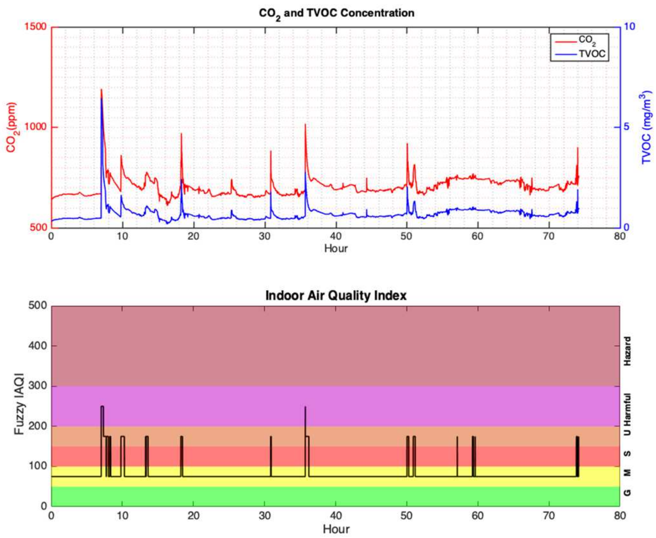

Fig. 10 TVOC and CO2 time series dataset obtained by the sensor for about 74 hours. Below: FIAQI outputs.

Now, less common IAQ assessments are conducted qualitatively and quantitatively. Quantitative assessments can be conducted using sensors that detect indoor bioaerosols and volatile organic compounds (VOCs), such as visually inspecting the indoor environment for indoor odour distribution. The FIAQI index proposed in this study includes the most significant indoor TVOC and CO2 air pollutants, which are necessary for the index value to accurately represent the IAQ in terms of the health of the room's occupants.

IV. CONCLUSION

The purpose of this paper is to propose an efficient indoor air quality index through the fusion of sensors measuring multiple parameters. This can be accomplished using the integration of ARIMA Kalman with Filter-based Time Series, which estimates and forecasts indoor air quality parameters over time. The sensor fusion using Fuzzy Logic is used to classify the index. Additionally, the FIAQI index incorporates references and expert knowledge when determining sensor fusion. This way, we can avoid the time and expense associated with designing traditional models. Another effective strategy for demonstrating the FIAQI index's validity is selecting significant indoor air quality parameters. According to the experiment, all the FIAQI's quality parameters significantly contribute to generating the final index. Therefore, if the indoor air quality index is to accurately reflect the indoor environmental air quality in terms of human health, it must incorporate all the parameters listed above. This FIAQI index result considers the most hazardous indoor TVOC and CO2 air pollutants, which are required for the index value to accurately reflect the IAQI in terms of the health of the room's occupants. However, the effect of each parameter on the final index must be evaluated in terms of its significance to long term of human health, as is the case with the FIAQI used in this research.

ACKNOWLEDGMENT

This research was funded by the 2nd term of Basic and Applied Research Grant 2021, Internal Research Grant Telkom University.

REFERENCE

[1] J. Saini, M. Dutta, and G. Marques, "A comprehensive review on indoor air quality monitoring systems for enhanced public health," *Sustainable Environment Research*, vol. 30, no. 1, 2020, doi: 10.1186/s42834-020-0047-y.

[2] J. Ahn, D. Shin, K. Kim, and J. Yang, "Indoor air quality analysis using deep learning with sensor data," *Sensors (Switzerland)*, vol. 17, no. 11, Nov. 2017, doi: 10.3390/s17112476.

[3] M. Hori, "Total Volatile Organic Compound (TVOC) as Index of Indoor Air Quality and Its Measuring and Evaluation," *Journal of the Human-Environment System*, vol. 1, no. 1, pp. 1–11, 2020.

[4] J. Fernández-Agüera, S. Dominguez-Amarillo, M. Fornaciari, and F. Orlandi, "TVOCs and PM 2.5 in naturally ventilated homes: Three case studies in a mild climate," *Sustainability (Switzerland)*, vol. 11, no. 22, Nov. 2019, doi: 10.3390/su11226225.

[5] B. W. Dionova, M. N. Mohammed, S. Al-Zubaidi, and E. Yusuf, "Environment indoor air quality assessment using fuzzy inference system," *ICT Express*, vol. 6, no. 3, pp. 185–194, Sep. 2020, doi: 10.1016/j.icte.2020.05.007.

[6] J. Saini, M. Dutta, and G. Marques, *Internet of Things for Indoor Air Quality Monitoring*. Springer, 2021.

[7] C. Guarnaccia, J. G. C. Breton, R. M. C. Breton, C. Tepedino, J. Quartieri, and N. E. Mastorakis, "ARIMA models application to air pollution data in Monterrey, Mexico," in *AIP Conference Proceedings*, Jul. 2018, vol. 1982. doi: 10.1063/1.5045447.

[8] Z. Ye, "Air Pollutants Prediction in Shenzhen Based on ARIMA and Prophet Method," 2019. doi: 10.1051/e3sconf/20191360.

[9] S. M. Abdullah, "To Predict Air Pollution using Machine Learning and Arima Model," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 11, 2021.

[10] Q. P. Ha, S. Metia, and M. D. Phung, "Sensing Data Fusion for Enhanced Indoor Air Quality Monitoring," *IEEE Sensors Journal*, vol. 20, no. 8, pp. 4430–4441, Apr. 2020, doi: 10.1109/JSEN.2020.2964396.

[11] J. Saini, M. Dutta, and G. Marques, "Sensors for indoor air quality monitoring and assessment through Internet of Things: a systematic review," *Environmental Monitoring and Assessment*, vol. 193, no. 2, Feb. 2021, doi: 10.1007/s10661-020-08781-6.

[12] S. M. Saad *et al.*, "Development of indoor environmental index: Air quality index and thermal comfort index," *AIP Conference Proceedings*, vol. 1808, no. August, 2017, doi: 10.1063/1.4975276.

[13] J. J. Carbajal-Hernández, L. P. Sánchez-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "Assessment and prediction of air quality using fuzzy logic and autoregressive models," *Atmospheric Environment*, vol. 60, pp. 37–50, Dec. 2012, doi: 10.1016/j.atmosenv.2012.06.004.

[14] Y. Wang, M. Zhao, Y. Han, and J. Zhou, "A fuzzy expression way for air quality index with more comprehensive information," *Sustainability (Switzerland)*, vol. 9, no. 1, 2017, doi: 10.3390/su9010083.

[15] A. Javid, A. Abbas Hamedian, H. Gharibi, and M. H. Sowlat, "Towards the Application of Fuzzy Logic for Developing a Novel Indoor Air Quality Index (FIAQI)," *Iran Journal of Public Health*, vol. 45, no. 2, pp. 203–213, 2016.

[16] G. Mani and R. Volety, "A comparative analysis of LSTM and ARIMA for enhanced real-time air pollutant levels forecasting using sensor fusion with ground station data," *Cogent Engineering*, vol. 8, no. 1, 2021, doi: 10.1080/23311916.2021.1936886.

[17] S. Ebrahim and F. Mofid, "Air Pollution Monitoring Using Fuzzy Logic in Industries," *Advanced Air Pollution*, 2011, doi: 10.5772/16947.

[18] S. Widodo, M. M. Amin, and A. Supani, "Design of Indoor Room Gas CO and SO2 Detection Based on Microcontroller Using Fuzzy Logic," *E3S Web of Conferences*, vol. 125, 2019, doi: 10.1051/e3sconf/201912523013.

[19] Y. Colella, A. S. Valente, L. Rossano, T. A. Trunfio, A. Fiorillo, and G. Improta, "A Fuzzy Inference System for the Assessment of Indoor Air Quality in an Operating Room to Prevent Surgical Site Infection," *International Journal of Environmental Research and Public Health*, vol. 19, no. 6, 2022, doi: 10.3390/ijerph19063533.

[20] Sukarna, E. Y. P. Ananda, and M. S. Wahyuni, "Rainfall Forecasting Model Using ARIMA and Kalman Filter in Makassar, Indonesia," in *Journal of Physics: Conference Series*, Dec. 2021, vol. 2123, no. 1. doi: 10.1088/1742-6596/2123/1/012044.

[21] E. J. Bardana, "Indoor pollution and its impact on respiratory health," *Annals of Allergy, Asthma and Immunology*, vol. 87, no. 6 SUPPL. 3, pp. 33–40, 2001, doi: 10.1016/S1081-1206(10)62338-1.