Sensing Data Fusion for Enhanced Indoor Air Quality Monitoring

Q. P. Ha, S. Metia, and M. D. Phung

***Abstract*—Multisensor fusion of air pollutant data in smart buildings remains an important input to address the well-being and comfort perceived by their inhabitants. An integrated sensing system is part of a smart building where real-time indoor air quality data are monitored round the clock using sensors and operating in the Internet-of-Things environment. In this work, we propose an air quality management system merging indoor air quality index (IAQI) and humidex into an enhanced indoor air quality index (EIAQI) by using sensor data on a real-time basis. Here, indoor air pollutant levels are measured by a network of waspmote sensors while IAQI and humidex data are fused together using an extended fractional-order Kalman filter (EFKF). According to the obtained EIAQI, overall air quality alerts are provided in a timely fashion for accurate prediction with enhanced performance against measurement noise and nonlinearity. The estimation scheme is implemented by using the fractional-order modeling and control (FOMCON) toolbox. A case study is analysed to prove the effectiveness and validity of the proposed approach.**

***Index Terms*—Sensing fusion, Indoor air quality, Extended Fractional Kalman Filter,**

I. Introduction

With the increasing growth worldwide of active population working inside a building, the management of indoor air quality is becoming crucially important for human health and work efficiency [1]. In this regard, the development of smart buildings is aimed to provide comfort and improved indoor air quality (IAQ) for occupants. Common issues associated with IAQ include improper or inadequately-maintained heating and ventilation as well as pollution by hazardous materials [2] (olefins, aromatics, hydrocarbons, glues, fiberglass, particle boards, paints, etc.) and other contaminant sources (laser printers [3], tobacco smoke, excessive concentrations of bacteria, viruses, fungi (including molds [4]), etc.). Moreover, the increase in the number of building occupants and the time spent indoors directly impact the IAQ [5].

Air quality can be evaluated by such parameters as concentration of air pollutants including carbon monoxide (CO), carbon dioxide (CO_2), formaldehyde (HCHO), nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2), total volatile organic compounds (TVOCs), particulate matter ($PM_{2.5}$), total suspended particles (TSP), as well as temperature, relative humidity and air movement. For an indoor environment, air quality is affected also by household chemicals, furnishings, air contaminants emitted from outside, occupant activities (e.g., smoking, cooking, breathing) [6], as well as air infiltration, and manual/mechanical ventilation. Due to a large number of factors involved, the development of an accurate system for IAQ monitoring is of great interest. To this end, fusing heterogeneous data from a network with a multitude of sensor types is essential for calculating the IAQI and for monitoring different pollutants in a building.

For indoor air quality, season-dependent models have been developed in [7] for monitoring, prediction and control of the IAQ in underground subway stations, where the IAQ of a metro station is shown to be influenced by temperature variations in different seasons. As indoor air quality (IAQ) is affected by heating, ventilation and air-conditioning conditions, modelling and control strategies have been proposed for residential air conditioning [8] and ventilation systems [9] to improve the occupants' living environment. A recent work [10] has showed that IAQ is affected by both outdoor particle concentration and indoor activities (walking, cooking, etc.). In [11], IAQ is assessed by monitoring and analysing CO_2 levels at the building's foyer area taking into account also thermal comfort. While indoor thermal comfort can be predicted via himidity [12], it is known that an elevated level of humidity may have a positive impact on the perceived IAQ with some effects on human health [13]. In IAQ modeling, data fusion is an effective way to reduce the sensors measurement uncertainties and overcome sensory limitations [14]. Various strategies have been employed, among which Kalman filtering is quite popular and effective. Multi-sensor data fusion using Kalman filtering is adopted in [15] to estimate the mass and flow parameters of gas transport processes from their relation of to building energy consumption and indoor air quality. For improving the model accuracy and robustness, system identification and data fusion are implemented for on-line adaptive energy forecasting in virtual and real commercial buildings with filter-based techniques [16]. In [17], a Kalman consensus filter is also used to analyze aircraft cabin contamination data with state estimation.

Motivated by [12]–[14], this paper proposes a data fusion strategy for the sensor network of a smart building to integrate the humidex and IAQI into an Enhanced Indoor Air Quality Index (EIAQI) with a weighting scheme to take into account also indoor humidity. Here, an Extended Fractional Kalman Filter (EFKF) incorporating the Matérn covariance function and a fractional order system is developed to deal with spatial distributions as well as the highly nonlinear, uncertain nature of indoor air quality data while merging humidex into IAQI for the proposed EIAQI. The Matérn covariance function is

Q. P. Ha, S. Metia, and M. D. Phung are with Faculty of Engineering and Information Technology, University of Technology Sydney, Broadway NSW 2007, Sydney, Australia (e-mail: Quang.Ha@uts.edu.au, Santanu.Metia@uts.edu.au, and ManhDuong.Phung@uts.edu.au).

Manuscript received July 14, 2019.

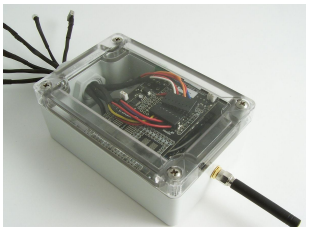

Fig. 1: Waspmote sensor for recording IAQ.

Fig. 2: Meshlium gateway router

know as smoothness parameter which controls the degree of smoothness of the random field. In [18], authors have used Matérn covariance function geostatistical applications. In [19], researchers have proposed various ways to produce cross-covariance models, which is based on the Matérn covariance model class, that are suitable for describing prominent nonstationary characteristics of the global environmental processes. Air quality monitoring across the globe is mainly based on monitoring stations, which are too sparse to accurately assess the exposure effects of air pollution for the world. In [20], the Matérn covariance function based predictive model is used to predict PM_{10} and $PM_{2.5}$. In this research work the Matérn covariance function is used to smooth indoor pollutant profile to predict accurately.

The remainder of the paper is organized as follows. Section II describes the sensor network system for IAQ management and the paper motivation. Section III presents the proposed framework for obtaining the enhanced indoor air quality index. In Section IV, the EFKF development is included together with results and discussion on real data obtained from the building network sensors. Rationale for data fusion with EFKF as well as IAQ assessment are given in Section V. Finally, the conclusion is drawn in Section VI.

II. System Description and Motivation

An office building is chosen as a testbed for the study in this paper. The building is embedded with numerous sensors for monitoring of its energy consumption as well as internal and external environment. For environmental monitoring, data are collected for such parameters as structural strain, people counting, vibrations and noise levels, as well as gas concentrations, weather, temperature, and meteorological conditions.

The building management system is installed on its top floor. In this paper, our focus is on its application to the monitoring of IAQ in the building only. The sensors used in the building's IAQ sensor network are the Waspmote sensors, as shown in Fig. 1. The sensor can measure levels of air pollutant (hydrogen (H_2), ammonia (NH_3), ethanol (C_2H_6O), hydrogen sulfide (H_2S) and toluene (C_7H_8)), carbon monoxide (CO), oxygen (O_2) and carbon dioxide (CO_2) in parts per million (ppm). Temperature and humidity are also recorded in $^{\circ}$C (centigrade) and %RH (relative humidity). For air quality monitoring, 16 sensors were implemented on a floor and more than 100 others were implemented throughout the building. Sensor data gathered by the Waspmote plug and sense nodes are sent to the cloud by the Meshlium, a gateway router specially designed to connect the Waspmote sensor networks to the Internet via Ethernet, Wi-Fi and 3G interfaces, as shown in Fig. 2. Fig. 3 shows location of the Waspmote sensors on the 10^{th} floor, building 11, University of Technology Sydney.

A load resistance is required at the output of each Waspmote sensor to maximize an amplification stage gain. The choice of amplification stage gain and of the sensor's load resistance can be carried out according to two parameters: the specific sensor available, since there may be significant variations between two different sensors of the same model, and the value and range of concentrations of gas to be monitored. When selecting load resistance and amplification it must be remembered that, although the sensors must be powered by a voltage of 5V to function appropriately, the Waspmote allows input between 0 and 3.3V, so it will be necessary to calculate the resistance, load and gain values to adapt the measurement range of the sensor to the Waspmote input. The accuracy which can be obtained in the sensor's output value will be dependent on the way in which it is supplied. This way, the longer the power time or duty cycle, as appropriate, the better accuracy will be obtained. The disadvantage of prolonged power is an increase in the mote's consumption, with the consequent decrease of the battery's life, so adjusting the power of each sensor to the requirements of the specific application being developed is recommended to optimize the equipment's performance.

It is highly recommended to calibrate the Waspmote sensor in order to get an accurate value. The Waspmote sensor?s normal resistance and sensitivity may vary from one unit to another in a wide range. This calibration may not be necessary in all applications, for example if the Waspmote sensor is going to be used in a gas detection, where monitoring the variation of the Waspmote sensor output may be enough to have it working properly, and the normal operation conditions may be replicated without an specific equipment. The calibration procedure requires the capture of the Waspmote sensor response under different concentrations of gas in the target operation range (that should be comprised in the operation range of the Waspmote sensor), and, depending on the conditions of the application to be implemented, under controlled temperature and humidity. The larger the number of calibration points in that range the more accurate the calibration will be.

The case study in this paper was intrigued by a slight incident of a fainting student in a laboratory room. During the period, sensor measurements were recorded as depicted in Fig. 4 for temperature and humidity, as well as in Figs. 5-8 for the room air quality, where it can be seen from the logged data of a critical episode on the 24^{th} of August 2016 with an initial assessment as lack of oxygen. After a thorough investigation, the cause of the incident turned out to be high levels of indoor air pollutants on that day. This has motivated us of the development of an enhanced indoor air quality index to forecast to the occupants to avoid experiencing severe adverse health effects.

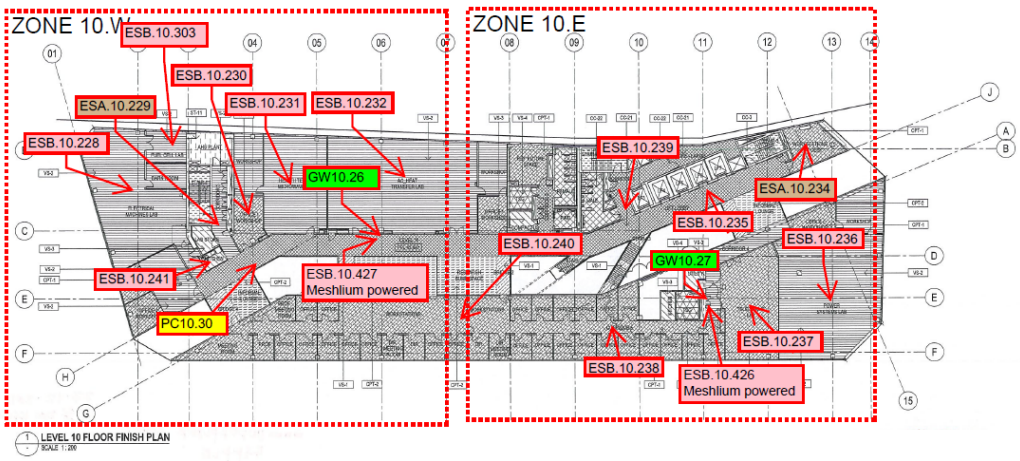

Fig. 3: Location of the Waspmote sensors.

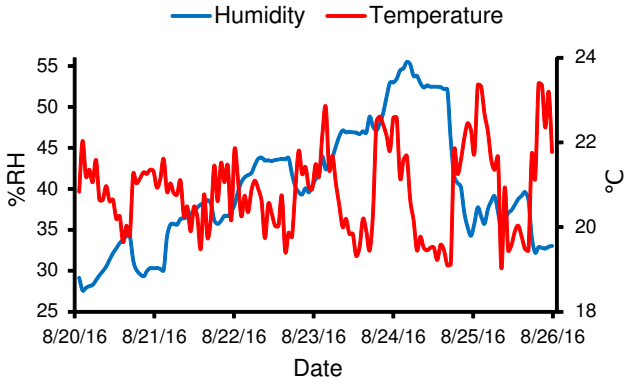

Fig. 4: Temperature (°C) and humidity (%RH) levels by Waspmote.

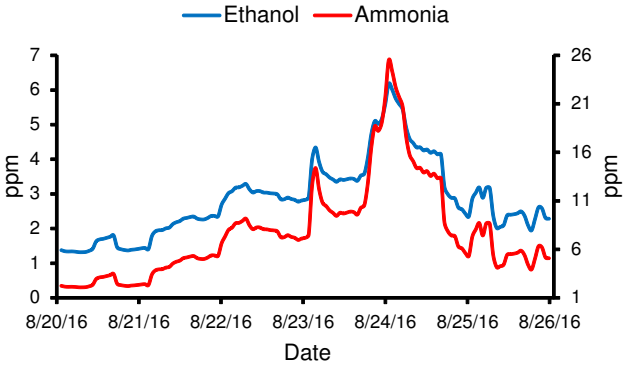

Fig. 6: Ethanol (ppm) and Ammonia (ppm) levels by Waspmote.

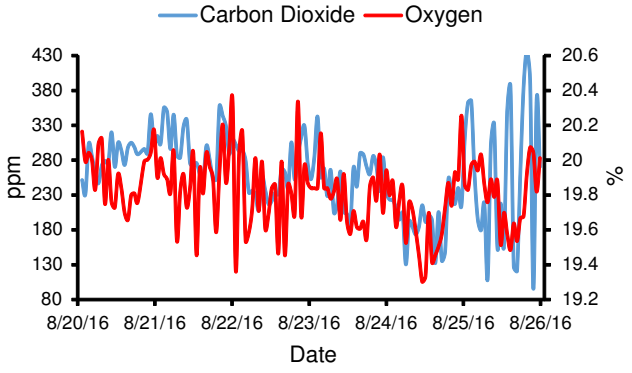

Fig. 5: CO_2 (ppm) and O_2 (%) levels by Waspmote.

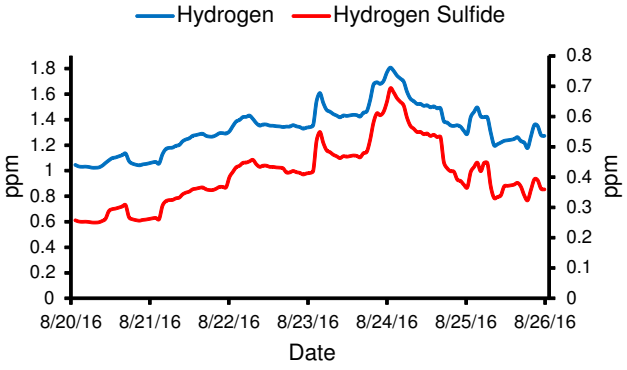

Fig. 7: Hydrogen (ppm) and Hydrogen Sulfide (ppm) levels by Waspmote.

III. IAQI DATA FUSION FRAMEWORK

In this section we provide a brief description of data fusion to calculate Indoor Air Quality Index (IAQI) and the proposed Enhanced Indoor Air Quality Index (EIAQI) incorporating also humidity.

A. Indoor Air Quality Index (IAQI)

Air quality index (AQI) has been used by environment protection agencies throughout the world. It is a scale of air pollution to indicate its levels to inform people around a region to adjust their outdoor activities in avoiding the health risk of getting polluted. The AQI is calculated on a real time basis to form a numerical scale with a colour code which is classified into several specific ranges. The information of AQI is very important especially to children, elderly people and people with pre-existing conditions such as cardiovascular and respiratory diseases. However, this index is usually applied to outdoor instead of indoor environments even though the indoors such as work places, hotels, homes, bedrooms and theater halls also have a certain impact on human health. For

TABLE I: INDOOR AIR QUALITY INDEX (IAQI)

CO (ppm)	**CO_2** (ppm)	**H_2** (ppm)	**NH_3** (ppm)	**C_2H_6O** (ppm)	**H_2S** (ppm)	**C_7H_8** (ppm)	**O_2** (%)	**IAQI**	**Health effects**
0-0.2	0-379	0-1	0-24	0-0.49	0-0.00033	0-0.0247	20.95	0-50	Good
0.21-2	380-450	1.1-2	25-30	0.5-10	0.00034-1.5	0.0248-0.6	19-20.9	51-100	Moderate
2.1-9	451-1000	2.1-3	31-50	11-49	1.6-5	0.7-1.6	15-19	101-150	Unhealthy for Sensitive
9.1-15.4	1001-5000	3.1-5	51-100	50-100	6-20	1.7-9.8	12-15	151-200	Unhealthy
15.5-30.4	5001-30000	5.1-8	101-400	101-700	21-50	9.9-12.2	10-12	201-300	Very Unhealthy
30.5-50.4	30001-40000	8.1-10	401-500	701-1000	51-100	12.3-100	<10	301-400	Hazardous

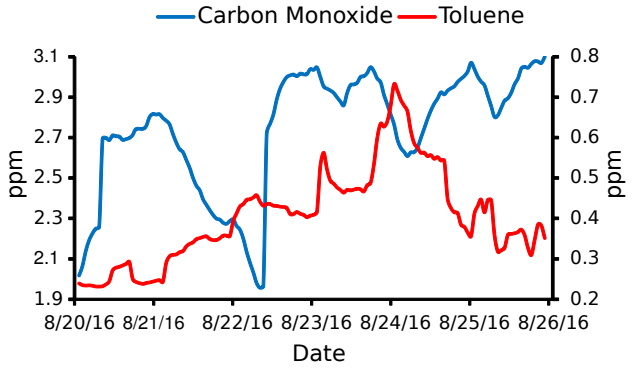

Fig. 8: Carbon Monoxide (ppm) and Toluene (ppm) levels by Waspmote.

outdoor air quality, the AQI is calculated from a ratio introduced by the U.S. EPA in 2006 with the corresponding colour code with six categories ranging from good to hazardous [21], whereby air quality standards are based on common outdoor air pollutants such as ozone, particulate matters $PM_{2.5}$ and PM_{10}, CO, sulfur dioxide (SO_2) and nitrogen dioxide (NO_2).

This research extends the existing AQI for determining the indoor air quality. Based on the AQI breakpoints, which are available online [21], the indoor air quality index can be evaluated with a sensing system [22]. A review of standards and guidelines for the IAQ parameters are given in [23]. Besides the six concentrations for the AQI as mentioned above, additional pollutants are needed [24] to calculate the indoor air quality index (IAQI). These include carbon dioxide (CO_2), volatile organic compounds (VOCs), radon and formaldehyde, which are known to cause concerns of health risk [25]. For example, the hydrogen sulfide breakpoint is set in accordance with the health effects of respiratory exposure [26]. Similarly, toluene, a toxic solvent, together with other contaminants such as formaldehyde can build up in a poorly-ventilated indoor environment. Its effect at different concentrations is explained in [27] with breakpoint details given in [28]. Ethanol vapour may cause irritation of the nose and throat with choking and coughing, depending on the level of concentration in air [29]. Ammonia, of which level breakpoint is defined in [30], may cause more severe problems with eyes, nose, throat and respiratory tract. High concentrations of hydrogen can cause oxygen deficit, which in turns may result in giddiness, mental confusion, loss of judgment, loss of coordination, weakness, nausea, fainting, or even loss of consciousness. Explanation of breakpoints for hydrogen concentration can be found in [31] and for oxygen level in [32]. In summary, Table I lists these gases together with the IAQI in association with their health effects coded in colour.

The air quality index for outdoor or indoor air pollutants can be calculated by using the following linear interpolation formula:

$$I_p = I_{ll} + \Big((C_p - BP_{ll}) \times \frac{I_{ul} - I_{ll}}{BP_{ul} - BP_{ll}}\Big), \qquad (1)$$

where I_p is the index for pollutant p, C_p is its rounded concentration, BP_{ul} (BP_{ll}) is the breakpoint greater (less) than or equal to C_p, and I_{ul} (I_{ll}) is the index value corresponding to BP_{ul} (BP_{ll}).

In the case of oxygen level, the IAQI is calculated using the following linear interpolation formula:

$$I_o = I_{ul} - \Big((BP_{ul} - C_o) \times \frac{I_{ll} - I_{ul}}{BP_{ll} - BP_{ul}}\Big), \qquad (2)$$

where I_o is the index for oxygen, C_o is its rounded concentration in percentage, BP_{ul} (BP_{ll}) is the breakpoint greater (less) than or equal to C_o, correspondingly with the upper (I_{ul}) and lower (I_{ll}) index of oxygen.

For example, the indoor waspmote gave C_p=230.4295 ppm for CO_2. We then obtained from Table I as BP_{ul} = 379, BP_{ll} = 0, I_{ul} = 50, I_{ll} = 0, and the IAQI obtained from (1) is 30.3997, which is in the "good" category. Now if waspmote readings for O_2 is C_o=19.7347 %, the breakpoints found from the table are then BP_{ul}=20.9, BP_{ll}=19, I_{ul}=100, and I_{ll}=51. The IAQI from (2) is therefore 69.9475, which is "moderate" in health concerns.

Data from eight sensors of the waspmote are used to calculate the IAQI correspondingly. To integrate also humidity and temperature for formulating the proposed enhanced indoor air quality index (EIAQI) we consider next the humidity index.

B. Humidex

Since the evaporation process of sweat for cooling down a human body in hot weather usually stops when the relative humidity reaches about 90%, indoor heat may yield a rise in the body temperature, causing illness. To describe the hot or cold feelings of an average person during different seasons, Canadian meteorologists proposed the humidex a dimensionless quantity based on the dew point theory, combining the effect of heat and humidity with breakdowns given in [32]. Accordingly, the humidex is calculated as,

$$h = T + \frac{5}{9} \times \Big(6.112 \times 10^{7.5 \times \frac{T}{237.7+T}} \times \frac{H}{100} - 10\Big), \qquad (3)$$

where T is air temperature in °C and H is relative humidity in %. Humidex ratings can be summarized in Table II.

TABLE II: HUMIDEX RATINGS

Humidex Range	Degree of Comfort
16-29	Comfort
30-39	No Comfort
40-45	Some Discomfort
46-54	Great Discomfort
55-60	Dangerous
61-65	Heat Stroke

C. Enhanced Indoor Air Quality Index (EIAQI)

For decades, the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Standard 55 has been using the Fanger's predicted mean vote (PMV) model to evaluate the indoor thermal comfort satisfaction [33]. PMV is based on the average vote of a large group of people on the a seven-point thermal sensation scale, using canonical thermal comfort models. Attempt to extend the IAQI to incorporate a thermal comfort index taking into account humidity can be found in [12]. In this work, we propose to improve the indoor air quality index by complementing it with humidex to formulate the EIAQI as,

$$\begin{aligned} EIAQI =& (W_h \times h) + (W_{IAQI} \times IAQI), \\ W_T =& W_h + W_{IAQI}, \end{aligned} \tag{4}$$

where W_h and W_{IAQI} are respectively the humidex and IAQI weighting factors ranging from -2 to 3, W_T is the overall EIAQI weightage, $IAQI$ is the indoor air quality index, and h is the humidex. The calculation procedure for the EIAQI is shown in Fig. 9. For example, at any given time, the status of IAQI is "Good" (weightage 3) and the Humidex status is "No Comfort' (weightage 2), then the total EIAQI weightage is 5 which refers the overall condition of the room as "Better".

IV. EXTENDED FRACTIONAL KALMAN FILTER

A majority of research work in indoor air quality is to obtain a mathematical model based on a given set of parameters and other information of geometry, shape, size, and contrast, see e.g., [34] to predict the pollutant distribution. On the other hand, inverse modelling generally focuses on the mathematical process of estimating the sources when determining the spatiotemporal distribution via a set of data or observations, see e.g., [35] for an outdoor emissions problem. For indoor applications, here extended fractional Kalman filtering is used to obtain air pollutant profiles in a smart building for accurately predicting the IAQ that the waspmotes installed in the building may overlook.

A. EFKF Estimation Scheme

The EFKF is particularly suitable for accurate and effective state estimation of highly nonlinear systems, where additive uncertainties, initial deviation, noise, disturbance and inevitably missing measurements affect the prediction performance [36]. In outdoor air quality modelling, an EFKF with Matérn function-based covariances has been applied for pollutant prediction [37] to improve accuracy of inventories and to complement missing data taking into account the spatial distribution of the indoor air quality profiles. Here, by adopting a Matérn correlation function for a length scale $l = \sqrt{5}/\lambda$, the EFKF of fractional order α is proposed as

$$\begin{aligned} \frac{d^\alpha f(t_k)}{dt^\alpha} &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\lambda^4 & -4\lambda^3 & -6\lambda^2 & -4\lambda \end{bmatrix} f(t_k) + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} w(t_k), \\ y(t_k) &= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} f(t_k) + d(t_k), \end{aligned} \tag{5}$$

where λ is a positive constant for the system quadruple pole (at $-\lambda$) depending on the correlation length l of the Gaussian process involved [37], $f(t_k)$ represents waspmote data assumed to have initial zero mean and covariance matrix $diag\{0.1\}$ with measurement variance 0.5^2 and spectral density of process noise 10^{-6}.

B. Fractional Order Identification

Fractional-order systems are considered as a generalization of integer-order ones to improve system performance. In this work, our implementation is based on the Fractional-Order Modeling and Control (FOMCON) Toolbox in MATLAB [38] with data collected in the time domain from waspmotes. Air pollutant concentrations, after conversion, are to be processed for prediction of abnormalities by using the EFKF where the fractional order is identified with FOMCON. Here, the black box modelling [39] is applied to infer a dynamic system model based upon experimentally collected data. This filtered model represents a relationship between system inputs and outputs under external stimuli in order to determine and predict the system behavior. Let y_r denote the experimental pollutant profile using eqn. (5) as a plant output, and y_m the identified model output. We consider the single-input and single-output (SISO) case where both y_r and y_m are $N \times 1$ vectors with the model output error:

$$\epsilon = y_r - y_m, \tag{6}$$

where estimation performance can be evaluated via the maximum absolute error:

$$\epsilon_{max} = \max_i \mid \epsilon(i) \mid, \tag{7}$$

or the mean squared error:

$$\epsilon_{MSE} = \frac{1}{N} \sum_{i=0}^{N} \epsilon_i^2 = \frac{\|\epsilon\|_2^2}{N}. \tag{8}$$

To demonstrate the merit and advantage of using EFKF to estimate pollutant profiles in smart buildings, let's take the concentration of CO_2 on the 22^{rd} August 2016 and $N = 78$ from data of the considered building. From conventional system identification, a corrected indoor air quality profile can be obtained from the corresponding rational transfer function

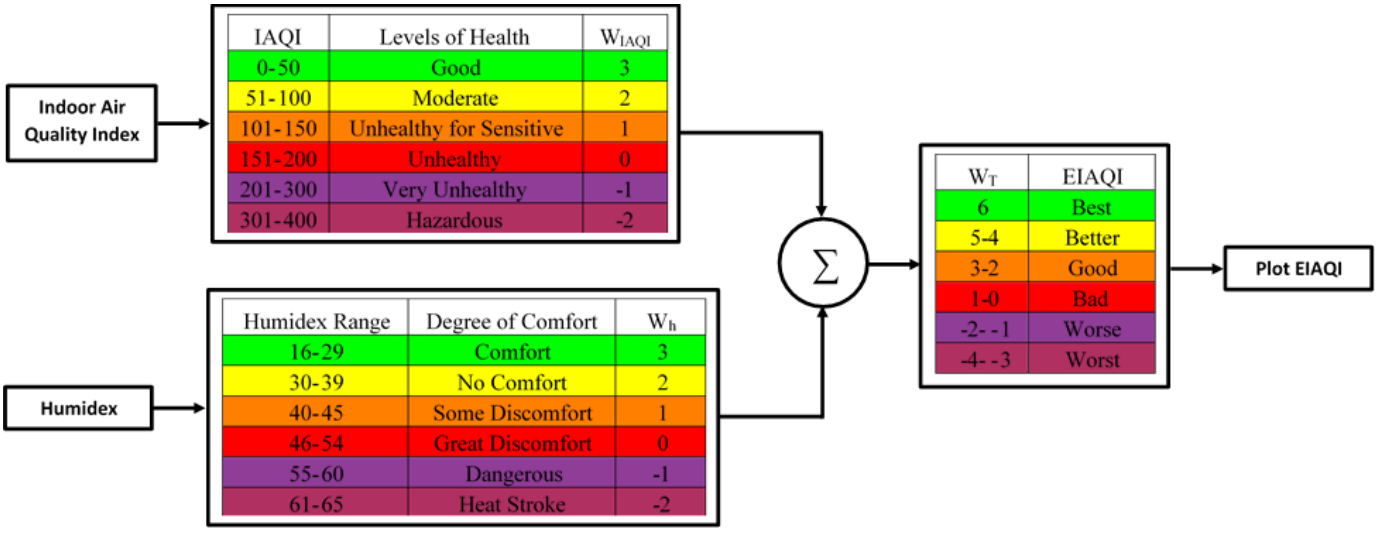

Fig. 9: Calculation procedures for EIAQI.

as:

$$F(s) = \frac{1}{a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0}, \tag{9}$$

where $a_4 = 1$, $a_3 = 1.058 \times 10^{-1}$, $a_2 = 4.2 \times 10^{-3}$, $a_1 = 7.408 \times 10^{-5}$, and $a_0 = 4.9 \times 10^{-7}$ for the pollutant level data collected at the testbed building. In fractional order modelling, the identification problem is included in estimating a set of parameters $a_n = [a_4 \quad a_3 \quad a_2 \quad a_1 \quad a_0]$ and $\alpha_n = [\alpha_{a_4} \quad \alpha_{a_3} \quad \alpha_{a_2} \quad \alpha_{a_1} \quad \alpha_{a_0}]$ for the transfer function of the model (5),

$$F^{\alpha}(s) = \frac{1}{a_4 s^{\alpha_{a_4}} + a_3 s^{\alpha_{a_3}} + a_2 s^{\alpha_{a_2}} + a_1 s^{\alpha_{a_1}} + a_0 s^{\alpha_{a_0}}}. \tag{10}$$

Table III shows the values of fractional orders obtained by using the FOMCON toolbox with the initial transfer function from equation (9) for all the indoor air pollutants, oxygen, temperature and humidity as collected by the building's waspmotes during the week from the 20th to 26th of August 2016. Here, contaminant gases include CO_2, CO, H_2, NH_3, C_2H_6O, H_2S, and C_7H_8 with the corresponding mean squared error ϵ_{MSE} ranging between 0.3 and 0.9 for $N = 78$.

C. Indoor Air Pollutant Profiles with EFKF

To illustrate the improvements in determining indoor air quality profiles by using the proposed EFKF, we compare the time series of the air pollutant as well as oxygen levels over the period of interest from the 23rd to 26th of August. Figure 10 shows the carbon dioxide concentration, distributed within a permissible limit from 400 to 1000 ppm, and rather consistent as obtained by waspmotes, EKF or EFKF. Similarly, the concentration distributions of gaseous contaminants such as hydrogen, ammonia, ethanol, hydrogen sulfide, toluene as well as temperature and humidity profiles are shown respectively in Figs. 11-17. They also display a general coincidence between the ground truth, EKF and EFKF. However, the carbon monoxide and oxygen levels, depicted respectively in Figs. 18 and 19, exhibit a difference on the 24th August with an increase of around 0.13 ppm in CO concentration and 0.4% in O_2 concentration by using EFKF as compared to the measured ground truth.

On one hand, while the levels of hydrogen, ammonia, ethanol and hydrogen sulfide lie in the moderate ranges as referred to Table I, the peak of these profiles all rests with the 24th of August, which may become unhealthy to highly sensitive people. On the other hand, the concentration of toluene C_7H_8 shows clearly a rise on the same day of over 0.7 ppm which is unhealthy for a sensitive person. Moreover, it is interesting to note that from the correction of EFKF, the level of oxygen on the incident date was moderate indeed with over 20%, while the concentration of carbon monoxide was found rather higher than the waspmote measurements and unhealthy for sensitive people. These filtered profiles explain that the cause for the student's fainting was an exposure to not of a low oxygen concentration but of a poor indoor air quality environment with unhealthy levels of gaseous pollutants such as CO and C_7H_8, particularly in association with a substantial rise in humidity on the incident date.

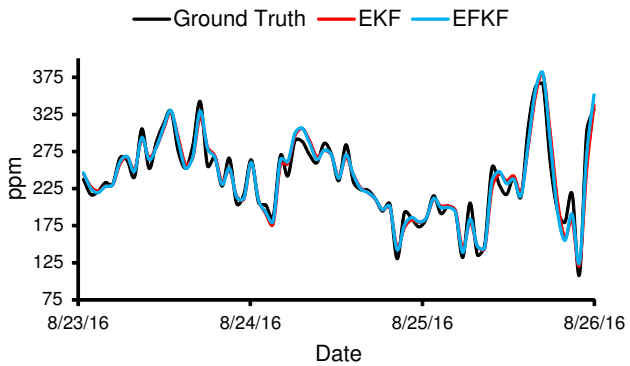

Fig. 10: Carbon dioxide concentration (ppm).

D. Statistical Analysis

In order to evaluate the performance of prediction, we introduce several model performance measures including MAPE

TABLE III: FRACTIONAL ORDER SYSTEM ESTIMATED BY USING FOMCON

System input	Fractional order system	ϵ_{MSE}
Carbon Dioxide	$\frac{1}{10.297s^{3.0213}+10.463s^{1.3718}-81.103s^{1.2455}+74.212s^{1.2287}+1.0541s^{0.0046851}}$	0.8612
Carbon Monoxide	$\frac{1}{40.309s^{4.1246}-53.264s^{3.9793}+20.59s^{3.5975}+3.178s^{1.2639}+1.0523s^{0.0087201}}$	0.3993
Oxygen	$\frac{1}{8.209s^{3.7867}+2.0644s^{1.9433}+2.7378s^{1.5326}+0.215s^{1.5043}+1.142s^{0.026486}}$	0.7082
Hydrogen	$\frac{1}{11.996s^{3.2658}-21.778s^{2.0137}+19.827s^{1.9881}+3.6513s^{1.2145}+1.6149s^{0.010517}}$	0.5122
Ammonia	$\frac{1}{6.9802s^{2.2789}+10.622s^{2.2469}-1.8461s^{2.2268}-1.0216s^{1.915}+1.0213s^{0.00045903}}$	0.6103
Ethanol	$\frac{1}{-5.1094s^{4.2713}+321.5s^{2.6011}-311.47s^{2.5897}+5.5133s^{1.687}+1.0893s^{0.0057901}}$	0.6761
Hydrogen Sulfide	$\frac{1}{51.293s^{2.7929}-44.849s^{2.7567}+6.4413s^{1.6742}-0.47948s^{1.1401}+1.0917s^{0.0089142}}$	0.4738
Toluene	$\frac{1}{13.304s^{2.7828}-8.8256s^{2.5007}+8.3516s^{1.7996}-0.39703s^{1.2433}+1.0903s^{0.007752}}$	0.4262
Temperature	$\frac{1}{12.003s^{3.2165}+18.0826s^{2.6112}+0.0518s^{1.2501}+1.63903s^{1.4201}+0.0190s^{0.0239728}}$	0.3601
Humidity	$\frac{1}{-10.0314s^{3.5015}+1.8006s^{2.9507}+1.3676s^{1.8001}+1.51203s^{1.0052}+10.1093s^{0.80061078}}$	0.3007

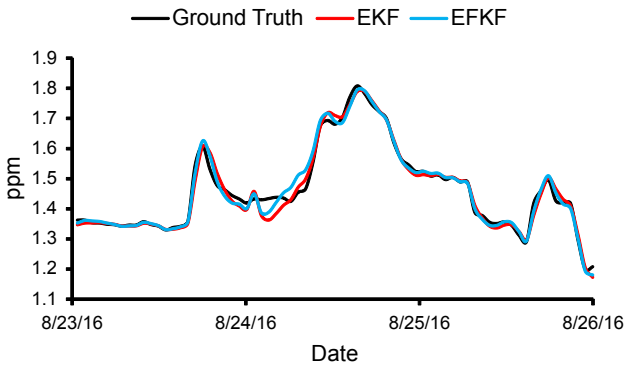

Fig. 11: Hydrogen concentration (ppm).

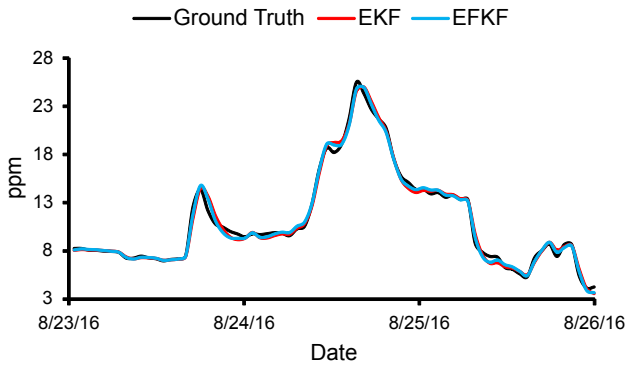

Fig. 12: Ammonia concentration (ppm).

(mean absolute percentage error), RMSE (root mean square error) and R^2 (the coefficient of determination), defined respectively as follows:

$$MAPE = \frac{100}{n}\sum_{j=1}^{n}\left(\frac{|a_j - b_j|}{|a_j|}\right), \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(a_j - b_j)^2}, \tag{12}$$

$$R^2 = 1 - \left(\frac{\sum_{j=1}^{n}(b_j - a_j)^2}{\sum_{j=1}^{n}(b_j)^2}\right), \tag{13}$$

where a_j and b_j are the forecast and observed values, and n is the number of samples. MAPE and RMSE are applied as

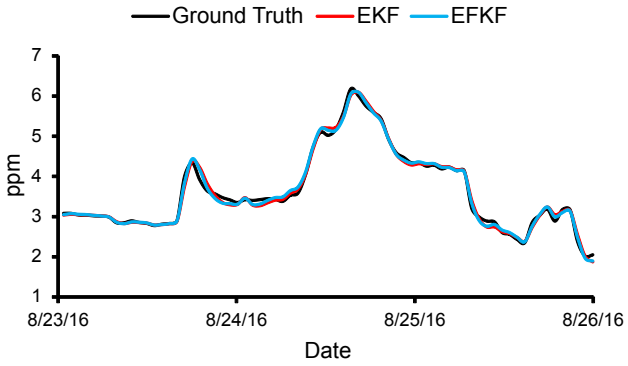

Fig. 13: Ethanol concentration (ppm).

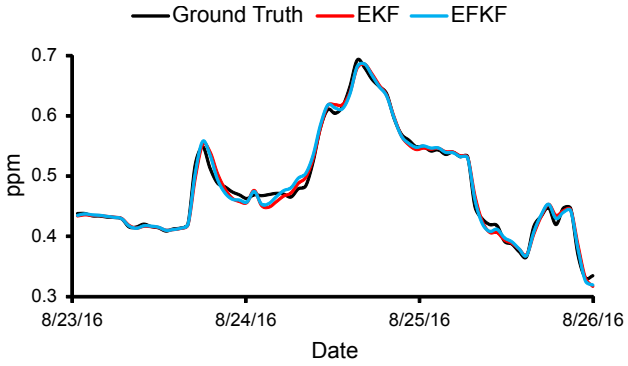

Fig. 14: Hydrogen sulfide concentration (ppm).

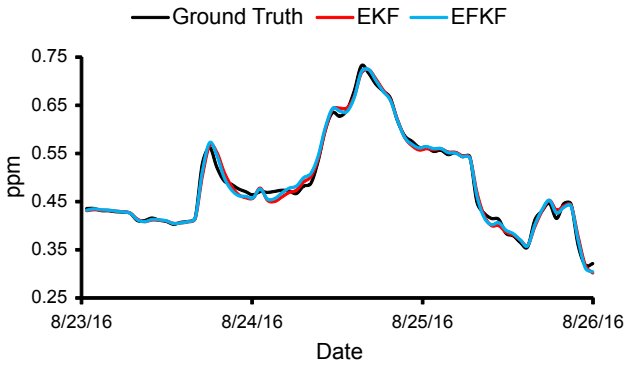

Fig. 15: Toluene concentration (ppm).

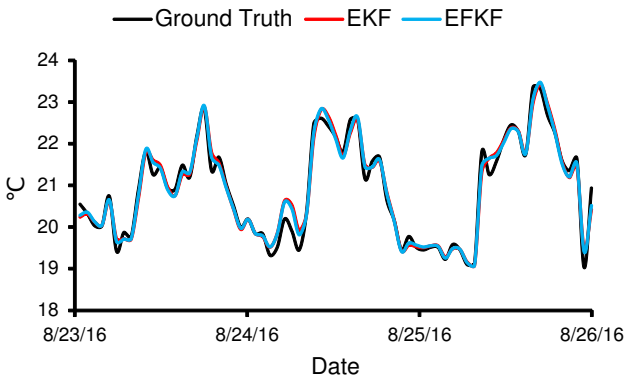

Fig. 16: Temperature (°C).

performance criteria of the prediction model to quantify the errors of forecasting values. The coefficient of determination

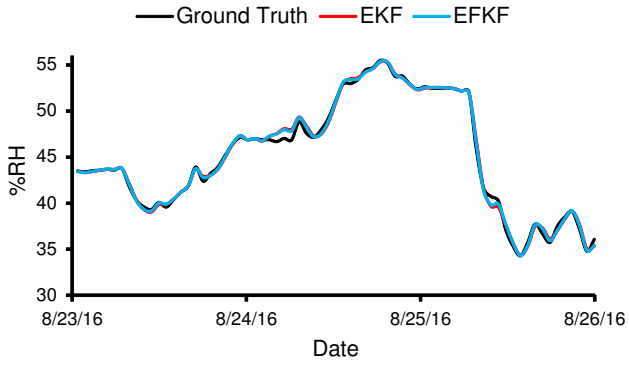

Fig. 17: Humidity (%RH).

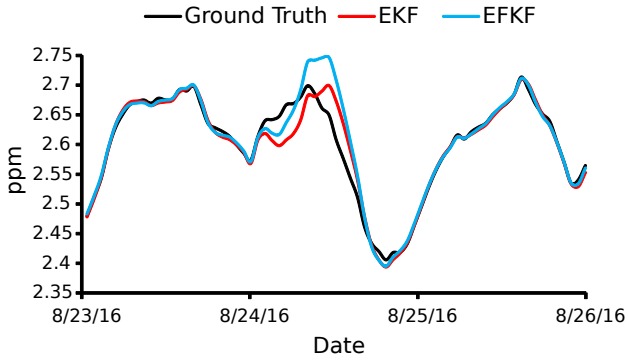

Fig. 18: Carbon monoxide concentration (ppm).

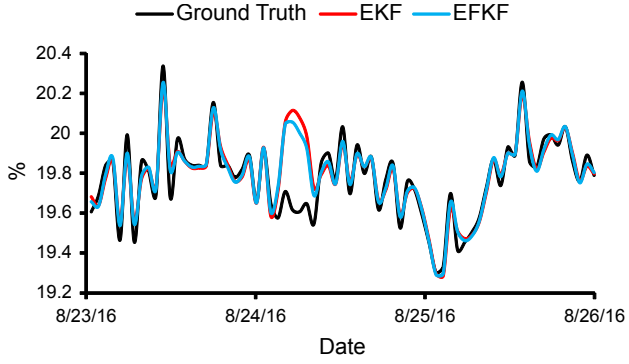

Fig. 19: Oxygen concentration (%).

R^2 is used to assess the strength of the relationship of the estimation to the accurate observation. Table IV provides descriptive statistics of EKF and EFKF prediction data on the 23rd and 24th of August, 2016. It justifies for the improvement of EFKF over EKF in precise estimation of indoor air quality. This is accounted by the merit of the Matérn covariance function associated with the Kalman filters used to allow for better correlation at a suitable length scale between the waspmotes and a location inside the building, here $\lambda = \frac{\sqrt{(5)}}{l}$ and $l = 5m$, the distance from the corridor (waspmote) to the lab room (incident location, in this study). As can be seen, the RMSE value is higher in the case of CO_2 concentration as compared to other IAQ levels. This is explained by a biological factor whereby human beings also produce CO_2 due to the natural process of respiration with a wide range of permissible limits (400-1000 ppm).

TABLE IV: PERFORMANCE STATISTICS OF EKF AND EFKF IN DATASETS OF DIFFERENT DAYS

Pollutant	**8/23/2016**						**8/24/2016**					
	EKF			EFKF			EKF			EFKF		
	MAPE	RMSE	R^2	MAPE	RMSE	R^2	MAPE	RMSE	R^2	MAPE	RMSE	R^2
Carbon Dioxide	2.94%	1.0498	0.9187	2.45%	0.8729	0.9437	3.31%	2.5957	0.8476	2.69%	2.2503	0.8982
Carbon Monoxide	0.71%	0.0076	0.9177	0.70%	0.0073	0.9213	0.94%	0.0078	0.7998	0.88%	0.0078	0.8202
Oxygen	0.70%	0.0818	0.8285	0.59%	0.0692	0.8778	0.52%	0.0631	0.8978	0.45%	0.0551	0.9229
Hydrogen	3.40%	1.7091	0.9631	2.42%	1.3752	0.9761	3.77%	1.8922	0.9892	2.56%	1.4871	0.9928
Ammonia	2.20%	0.2681	0.9918	1.94%	0.2166	0.9946	3.90%	0.4113	0.9948	3.28%	0.3394	0.9964
Ethanol	1.48%	0.0597	0.9942	1.34%	0.0491	0.9960	2.41%	0.0911	0.9937	2.09%	0.0779	0.9954
Hydrogen Sulfide	1.08%	0.0059	0.9946	1.01%	0.0051	0.9959	1.68%	0.0092	0.9918	1.51%	0.0082	0.9934
Toluene	1.18%	0.0065	0.9946	1.10%	0.0055	0.9961	1.86%	0.0100	0.9925	1.65%	0.0089	0.9941
Temperature	0.65%	0.1840	0.9613	0.56%	0.1579	0.9717	0.82%	0.2349	0.9852	0.71%	0.1983	0.9896
Humidity	0.54%	0.2593	0.9987	0.46%	0.2176	0.9992	0.8%	0.4416	0.9984	0.67%	0.3757	0.9988

V. INDOOR AIR QUALITY ASSESSMENT

The above findings indicate the importance of accurate, comprehensive and continuous monitoring with a prediction system for the IAQ, taking into account also human comfort. Such a system should be integrated into a building management for better monitoring the IAQ and, more importantly, prevention of any incidents via, e.g., ventilation control.

A. Time Series Plot of IAQI and Humidex Using Real Data and Estimated Data

To consider the overall indoor air quality index for calculation of the IAQI, Eqn. (1) is used to interpolate data of all pollutants except the oxygen level which is obtained from Eqn. (2). Fig. 20 shows the time series plot of IAQI using real data from waspmotes and processed data from EFKF, where it can be seen that the value of IAQI is rather high and appears to be very unhealthy on the 24^{th} August, 2016, particullarly affecting a sensitive person. This could explain for the student's incident but does not account for the effect of humidity.

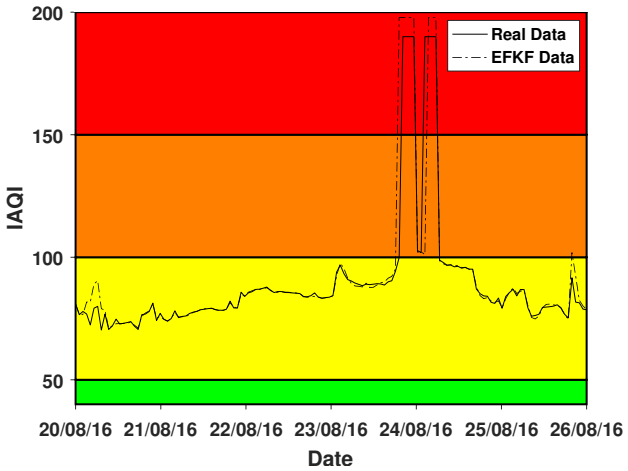

Fig. 20: IAQI plot using real data and EFKF data.

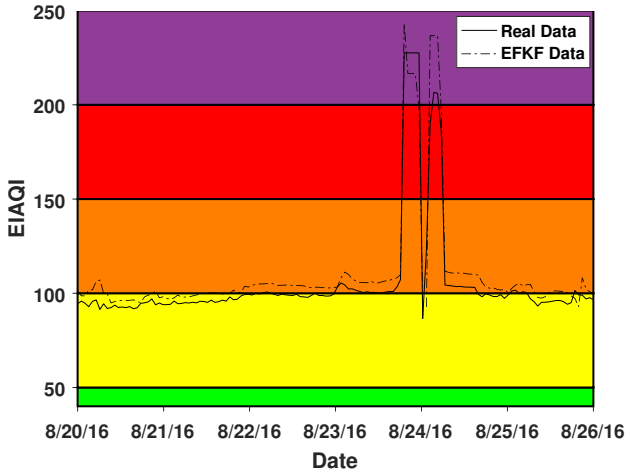

Fig. 21: Overall index for indoor air using real data and EFKF data .

B. Time Series Plot of EIAQI Using Real Data and Estimated Data

To incorporate also humidity, the enhanced indoor air quality index is calculated by using Eqn. 4. For a better illustration of the improvement obtained from the use of EFKF, the EIAQI plot is shown in Fig. 21 for both real data and estimated data according to weightage ($W_h = 1.0$ and $W_{IAQI} = 1.0$). Using the same colour codes for health effects presented in Fig. 20, it can be seen that with EFKF, the obtained EIAQI clearly indicates an increase in the indoor air quality index within a short period of time. This is reflected in the soon recovery of the sensitive student whereas the majority of the class could tolerate. Although the proposed EIAQI with EFKF estimation is not much different from with real data from waspmotes for most of the time, during the episode day, the enhanced indoor air quality index appears to be more accurately reflect the indoor air quality with EFKF data owing to the advantages in handling missing data as well as nonlinear and uncertain spatio-temporal distributions.

Figs. 22-23 show time series plot of oxygen concentration from 23^{rd} to 25^{th} August, 2015 and 23^{rd} to 25^{th} August, 2017 respectively. Similarly carbon monoxide time series are plotted

in figs. 24-25 respectively for 2015 and 2017. These figures show EFKF estimation is more accurate than EKF estimation. The concentration level of carbon monoxide level is lower in year 2015 and 2017 than 2016.

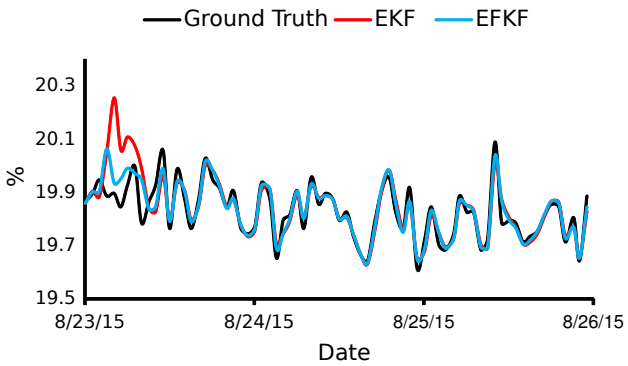

Fig. 22: Oxygen concentration (%).

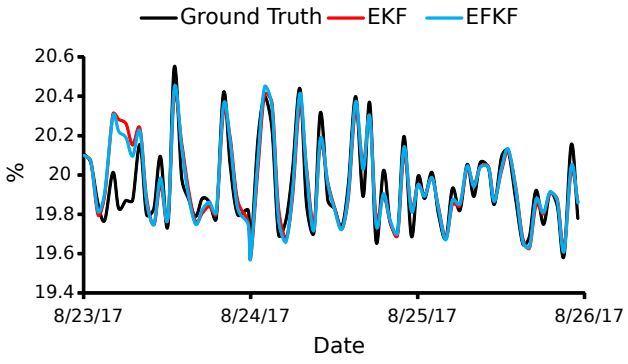

Fig. 23: Oxygen concentration (%).

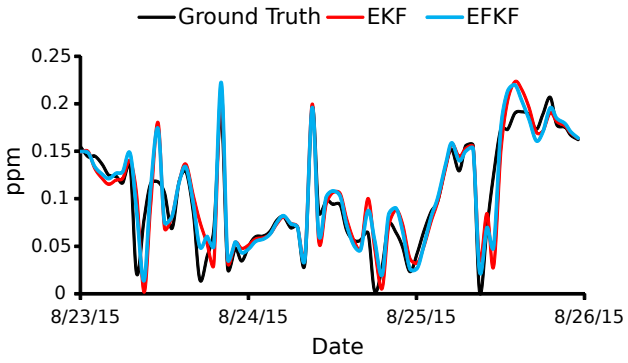

Fig. 24: Carbon monoxide concentration (ppm).

VI. Conclusion

In this paper, we have proposed an effective approach to improve accuracy in predicting indoor air pollutant profiles taking into account their nonlinear and stochastic nature, along with a novel index for indoor air quality considering also humidity. Here, an extended Kalman filter with a fractional order is developed for the indoor air quality model, in dealing with high nonlinearity and missing or inaccurate data collected from the building's sensors. To verify the performance imrovement,

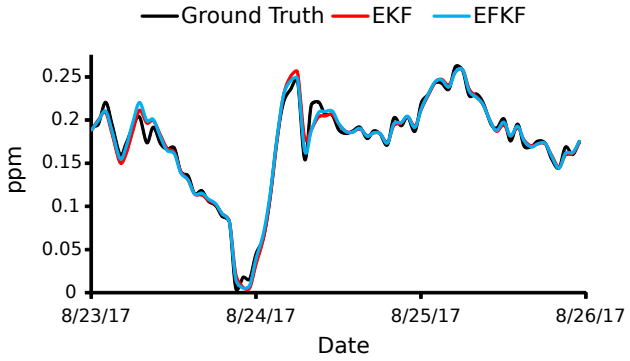

Fig. 25: Carbon monoxide concentration (ppm).

both EKF and EFKF algorithms have been implemented and compared. For illustration, an incident of a student with some slight fainting, is used as a case study to not only evaluate the effectiveness of the proposed estimation framework but also to emphasize the need of integrating accurate IAQ monitoring and prediction into the overall building management system to better maintain the inhabitants' wellbeing. In addition, a combination of IAQI and humidex is proposed to address the effect of humidity on indoor air quality.

References

[1] 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, April 2017.

[2] A. Lay-Ekuakille and A. Trotta, "Predicting voc concentration measurements: Cognitive approach for sensor networks," *IEEE Sensors Journal*, vol. 11, no. 11, pp. 3023–3030, Nov 2011.

[3] T. Tang, J. Hurraß, R. Gminski, and V. Mersch-Sundermann, "Fine and ultrafine particles emitted from laser printers as indoor air contaminants in german offices," *Environmental Science and Pollution Research*, vol. 19, no. 9, pp. 3840–3849, Nov 2012.

[4] O. A. Sahin, N. Kececioglu, M. Serdar, and A. Ozpinar, "The association of residential mold exposure and adenotonsillar hypertrophy in children living in damp environments," *International Journal of Pediatric Otorhinolaryngology*, vol. 88, no. Supplement C, pp. 233 – 238, 2016.

[5] M. Jin, N. Bekiaris-Liberis, K. Weekly, C. J. Spanos, and A. M. Bayen, "Occupancy detection via environmental sensing," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 2, pp. 443–455, April 2018.

[6] L. Zimmermann, R. Weigel, and G. Fischer, "Fusion of nonintrusive environmental sensors for occupancy detection in smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2343–2352, Aug 2018.

[7] M. Kim, B. SankaraRao, O. Kang, J. Kim, and C. Yoo, "Monitoring and prediction of indoor air quality (IAQ) in subway or metro systems using season dependent models," *Energy and Buildings*, vol. 46, pp. 48 – 55, 2012.

[8] Q. Ha and V. Vakiloroaya, "Modeling and optimal control of an energy-efficient hybrid solar air conditioning system," *Automation in Construction*, vol. 49, pp. 262–270, 2015.

[9] B. Sun, P. B. Luh, Q. Jia, Z. Jiang, F. Wang, and C. Song, "Building energy management: Integrated control of active and passive heating, cooling, lighting, shading, and ventilation systems," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 3, pp. 588–602, July 2013.

[10] L. Zhao, J. Liu, and J. Ren, "Impact of various ventilation modes on IAQ and energy consumption in Chinese dwellings: First long-term monitoring study in Tianjin, China," *Building and Environment*, vol. 143, pp. 99 – 106, 2018.

[11] A. Merabtine, C. Maalouf, A. A. W. Hawila, N. Martaj, and G. Polidori, "Building energy audit, thermal comfort, and IAQ assessment of a school building: A case study," *Building and Environment*, vol. 145, pp. 62 – 76, 2018.

[12] R. Rana, B. Kusy, R. Jurdak, J. Wall, and W. Hu, "Feasibility analysis of using humidex as an indoor thermal comfort predictor," *Energy and Buildings*, vol. 64, pp. 17 – 25, 2013.

[13] P. Wolkoff, "Indoor air humidity, air quality, and health - an overview," *International Journal of Hygiene and Environmental Health*, vol. 3, pp. 376–390, Apr 2018.

[14] A. Assa and F. Janabi-Sharifi, "A kalman filter-based framework for enhanced sensor fusion," *IEEE Sensors Journal*, vol. 15, no. 6, pp. 3281–3292, June 2015.

[15] M. Carminati, G. Ferrari, R. Grassetti, and M. Sampietro, "Real-time data fusion and mems sensors fault detection in an aircraft emergency attitude unit based on kalman filtering," *IEEE Sensors Journal*, vol. 12, no. 10, pp. 2984–2992, Oct 2012.

[16] X. Li and J. Wen, "System identification and data fusion for on-line adaptive energy forecasting in virtual and real commercial buildings," *Energy and Buildings*, vol. 129, pp. 227 – 237, 2016.

[17] R. Wang, Y. Li, H. Sun, and Z. Chen, "Analyses of integrated aircraft cabin contaminant monitoring network based on kalman consensus filter," *ISA Transactions*, vol. 71, pp. 112 – 120, 2017, special issue on Distributed Coordination Control for Multi-Agent Systems in Engineering Applications.

[18] E. Pardo-Igúzquiza, K. V. Mardia, and M. Chica-Olmo, "MLMATERN: A computer program for maximum likelihood inference with the spatial Matérn covariance model," *Computers & Geosciences*, vol. 35, no. 6, pp. 1139 – 1150, 2009.

[19] M. Jun, "Matérn-based nonstationary cross-covariance models for global processes," *Journal of Multivariate Analysis*, vol. 128, pp. 134 – 146, 2014.

[20] A. Beloconi, N. Chrysoulakis, A. Lyapustin, J. Utzinger, and P. Vounatsou, "Bayesian geostatistical modelling of PM10 and PM2.5 surface level concentrations in Europe using high-resolution satellite-derived products," *Environment International*, vol. 121, pp. 57 – 70, 2018.

[21] (2018) AQI Breakpoints. https://www.epa.gov/aqs. [Online]. Available: https://www.epa.gov/aqs

[22] J. Y. Kim, C. H. Chu, and S. M. Shin, "ISSAQ: An Integrated Sensing Systems for Real-Time Indoor Air Quality Monitoring," *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4230–4244, Dec 2014.

[23] S. A. Abdul-Wahab, S. C. F. En, A. Elkamel, L. Ahmadi, and K. Yetilmezsoy, "A review of standards and guidelines set by international bodies for the parameters of indoor air quality," *Atmospheric Pollution Research*, vol. 6, no. 5, pp. 751 – 767, 2015.

[24] S. Saad, A. Shakaff, A. Saad, A. Yusof, A. Andrew, A. Zakaria, and A. Adom, "Development of indoor environmental index: Air quality index and thermal comfort index," in *AIP Conference Proceedings*, vol. 1808, no. 1. AIP Publishing, 2017, p. 020043.

[25] H. Wang, C. Tseng, and T. Hsieh, " Developing an indoor air quality index system based on the health risk assessment," in *In: Proceedings of indoor air 2008*, Copenhagen, Denmark, 2008, paper 749.

[26] J. W. Snyder, E. F. Safir, G. P. Summerville, and R. A. Middleberg, "Occupational fatality and persistent neurological sequelae after mass exposure to hydrogen sulfide," *The American Journal of Emergency Medicine*, vol. 13, no. 2, pp. 199 – 203, 1995.

[27] J. Bælum, G. R. Lundgvist, L. Mølhave, and N. T. Andersen, "Human response to varying concentrations of toluene," *International Archives of Occupational and Environmental Health*, vol. 62, no. 1, p. 65, Jan 1990.

[28] J. T. Ayers. (2002) Approaches to a Total (or Grouped) VOC Guideline Final Report. https://open.alberta.ca/dataset/8163e248-eed2-41d5-aca7-59d7a7a7b2fc/resource/964e35c4-acb5-4717-a862-adf483cd007b/download/6686approachestoatotalorgroupedvocguideline.pdf.

[29] NationalToxicologyProgram, "Ntp 12th report on carcinogens." *Report on carcinogens: carcinogen profiles*, vol. 12, p. iii, 2011.

[30] (2018) Occupational Safety and Health Administration. https://www.osha.gov. [Online]. Available: https://www.osha.gov

[31] C.-S. Huang, T. Kawamura, Y. Toyoda, and A. Nakao, "Recent advances in hydrogen research as a therapeutic medical gas," *Free Radical Research*, vol. 44, no. 9, pp. 971–982, 2010.

[32] (2018) Canadian Centre for Occupational Health and Safety. https://www.ccohs.ca. [Online]. Available: https://www.ccohs.ca

[33] W. Hu, Y. Wen, K. Guan, G. Jin, and K. J. Tseng, "itcm: Toward learning-based thermal comfort modeling via pervasive sensing for smart buildings," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 4164–4177, Oct 2018.

[34] H. Wahid, Q. Ha, H. Duc, and M. Azzi, "Neural network-based meta-modelling approach for estimating spatial distribution of air pollutant levels," *Applied Soft Computing*, vol. 13, no. 10, pp. 4087 – 4096, 2013.

[35] S. Metia, Q. P. Ha, H. N. Duc, and M. Azzi, "Estimation of power plant emissions with unscented kalman filter," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 8, pp. 2763–2772, Aug 2018.

[36] X. Hu, H. Yuan, C. Zou, Z. Li, and L. Zhang, "Co-estimation of state of charge and state of health for lithium-ion batteries based on fractional-order calculus," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 319–10 329, Nov 2018.

[37] S. Metia, S. D. Oduro, H. N. Duc, and Q. Ha, "Inverse air-pollutant emission and prediction using extended fractional kalman filtering," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 2051–2063, May 2016.

[38] A. Tepljakov, *Fractional-order Modeling and Control of Dynamic Systems (Springer Theses)*. Springer, 2017.

[39] L. Ljung, *System Identification: Theory for the User (2nd Edition)*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1999.