Dynamic modeling of airflow rate through window openings based on CO2 data

Cédric Schreck, Simon Rouchier, Aurélie Foucquier, Étienne Wurtz

Dynamic modeling of airflow rate through window openings based on CO_2 data

Cédric Schreck[1,2], Simon Rouchier[1], Aurélie Foucquier[2] and Etienne Wurtz[2]

[1] Univ. Savoie Mont-Blanc, Campus du Bourget, 73375 Le Bourget du Lac, France
[2] Univ. Grenoble Alpes, CEA, Liten, Campus Ines, 73375 Le Bourget du Lac, France

E-mail: cedric.schreck@univ-smb.fr

Abstract.
Better estimation of air change rate in naturally ventilated buildings is a key for supporting passive summer thermal comfort strategies. In window aired configurations, this estimation is challenging either with measurements or building simulation methods. In this study, we describe and apply the state space modeling methodology based on the carbon dioxide concentration (CO_2) mass balance equation to a simple test case with numerically generated data. We show that a two-state CO_2 / *Air Change Rate* model is suitable. As a modeling novelty, we demonstrate the benefit of improving the formulation with a variable diffusion term for the air change rate state equation. From the test case study findings, we emphasize that a lower interior-exterior CO_2 difference results in weaker performance and we list some prospects for future work.

1. Introduction

The motivation for this work is to support the development of naturally ventilated strategies for summer thermal comfort in buildings, such as free cooling. Getting a better estimation of air change rate (ACR) through window openings is a key point for estimating cooling potential and would give some insights to study and design complementary retrofit ventilation strategies. The estimation can be performed either with building simulation models [1] or data-driven statistical models using discrete air speed measurement through windows [2] or tracer gas measurement [3]. However, because of its dependency on weather conditions (local wind speed and turbulence, temperatures), occupancy patterns, and real building properties, the ACR estimation suffers from weak reliability with both modeling and measurement methods [3].

Algorithms based on carbon dioxide concentration (CO_2) seem particularly promising for in-situ measurement [4, 3]. Moreover, as opposed to deterministic formulations, stochastic formulations allow considering both uncertainties in modeling and measurement discrepancies. State-space modeling with Kalman filtering has been showed to be particularly suitable for different cases [5, 6, 7, 8]. In order to allow its dynamical estimation, the ACR parameter is moved to a state parameter in the joined state parameter methodology [5, 9, 10, 11].

The purpose of this article is to describe the state space modeling methodology applied to the CO_2 mass balance equation, including an extended novel formulation to adapt to situations where there are sudden changes in airflow rate due to changes in window state.

Section 2 describes data generation methodology used as input and validation for the two models described in section 3. Section 4 will go through model validation and result analysis. Some critical analysis and perspective are provided in section 5.

2. Methodology and test case data generation

This work is motivated by the estimation of dynamical air change rate (ACR) through various window configurations in in-use residential buildings. As outlined in [5, 10], in order to evaluate the model formulation and parameters tuning, before using real case data, some preliminary work with a simple test case should be performed. In this work, we simulate a single zone representative of a living room with volume $V = 70$ m^3, as shown in figure 1. Note that the same methodology could be applied to a bedroom.

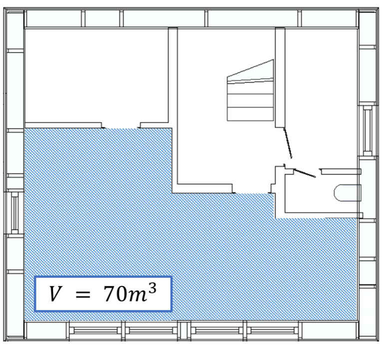

Figure 1. Sketch of the considered zone for synthetic data generation

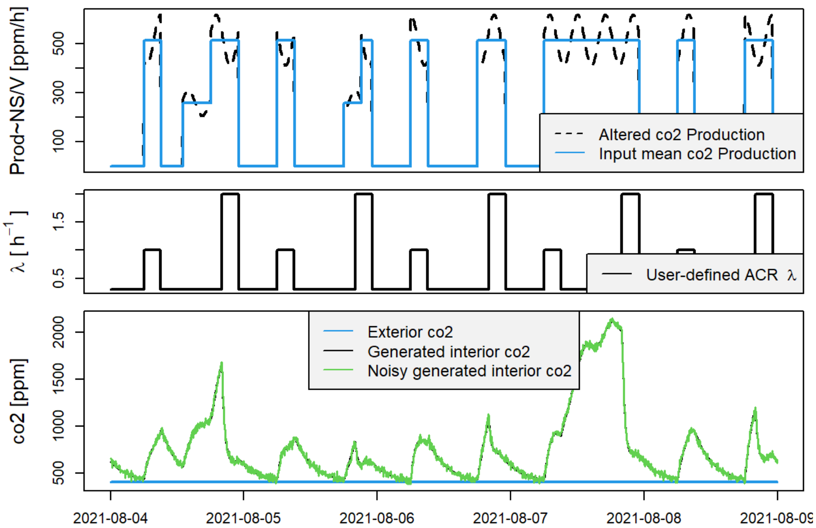

Figure 2. Schedules and synthetic generated data.

In order to compute synthetic data for indoor CO_2 concentration c [ppm], we use the CO_2 mass balance equation in the volumetric form in equation (1).

$$\frac{dc}{dt} = -\lambda\,(c - c_{ext}) + P, \text{ with } P = \frac{NS}{V} \times 3.6\text{e}6 \tag{1}$$

$c_{ext} = 400$ ppm is the exterior CO_2 concentration. $\lambda \in [0.3; 1; 2]$ h^{-1} is the zone ACR taking three typical values for respectively mechanical ventilation, weak, and middle force natural ventilation. $N \in [0; 1; 2]$ is the room number of occupants. S [l/s] is the metabolic rate of CO_2 produced per person. We choose an average value of $\bar{s} = 0.005$ l/s corresponding to adult medium activity [12] and apply harmonic oscillation $S(t) = \bar{s} + 0.001\cos\left(\frac{2\pi t}{5}\right)$. This simulates an occupant metabolic rate fluctuation of ± 0.001 l/s for a typical time duration of 5 h. As an advantage of synthetic generated data, the user is free to choose any schedule for N and λ. They are set to be representative of typical night cooling strategies in summer conditions. The resulting production term P [ppm/h] is multiplied by 3.6e6 for unit homogeneity.

Equation (1) is discretized and data are generated with time step $dt = 5$min. In order to simulate some CO_2 measurement discrepancies for c, we add some theoretical gaussian white noise with standard deviation $\sigma_{meas} = 25$ ppm, to get 95% of data in the sensor measurement accuracy interval ± 50 ppm. The production term P with and without deviation in S, the ACR λ, the exterior CO_2 (c_{ext}) and the resulting noisy indoor CO_2 (c), for a period of 5 days, are plotted in figure 2.

3. State Space modeling

In this study, we apply the state space methodology with Kalman filtering because it allows considering both uncertainties in physical modeling and discrepancies in measurement. Based on the work in [5, 11], we implement the two-state CO_2 / *Air Change Rate* model M1 written in the continuous form in equation (2).

The first equation for the state variable c is a process equation based on the CO_2 mass balance equation (1). The second equation for the state parameter λ is a fictitious process equation in order to model its dynamic [9, 5]. The right term of these two state equations are diffusion terms modeling the process noise. This stochastic white noise parts are composed of Wiener processes $d\omega_c$ and $d\omega_\lambda$ with associated standard deviation σ_{co2} and σ_λ. The third equation for measured time series $Yco2$ is the measurement equation modeling the fact that $Yco2$ differs from the noise-free state c by a gaussian white noise $d\omega_{Yco2}$ with standard deviation σ_{Yco2}. N, S, V and c_{ext} are input of the model. Note that the fluctuation in metabolic rate production per person $S(t) = \bar{s} + 0.001\cos\left(\frac{2\pi t}{5}\right)$ used for CO_2 synthetic data generation in section 2 is unknown in the inverse problem-solving. Only the average value is given as input $S(t) = \bar{s}$.

$$M1 = \begin{cases} dc = \left[-\lambda\left(c - c_{ext}\right) + \left(\frac{NS}{V} \times 3.6\mathrm{e}6\right)\right] dt + \sigma_{co2}\, d\omega_c \\ d\lambda = \sigma_\lambda\, d\omega_\lambda \\ Yco2 = c + \sigma_{Yco2}\, d\omega_{Yco2} \end{cases} . \quad (2)$$

In order to give estimates $[\hat{c}(t), \hat{\lambda}(t)]$ of unknown state parameters $[c, \lambda]$, at each time step t, the Kalman filter method consists in two main distinct stages: the predictor stage and the updating stage. From previous step estimates $[\hat{c}_{t-1}, \hat{\lambda}_{t-1}]$ the predictor stage uses the process equations to predict the 1-step ahead prediction $[\hat{c}_{t|t-1}, \hat{\lambda}_{t|t-1}]$. By comparing $\hat{c}_{t|t-1}$ to the known measured value $\mathrm{Yco2}_t$, the updating stage corrects the predicted states to get the current step estimation $[\hat{c}_t, \hat{\lambda}_t]$. The correction relies on the confidence in the process and measurement equations through the values of the stochastic parameters σ_{co2}, σ_λ, and σ_{Yco2}. At this stage, the Kalman filter also computes the likelihood value, which is an indicator representing the likelihood of observing the data given the model parameters. Estimation of constant parameters $\hat{\sigma}_{co2}$, $\hat{\sigma}_\lambda$, and $\hat{\sigma}_{Yco2}$ is computed through minimization of the total negative log-likelihood. Note that because λ is a dynamical parameter, the system is non-linear and an improved version of the Kalman filter must be implemented. See [5] for more explanations of the method.

The diffusion term standard deviation σ_λ in the λ process equation is a key value for modeling the dynamic of state λ. Bigger values for σ_λ will lead to better tracking ability while lower values will smooth the estimation [5, 11, 13]. We thus implemented an improved model M2 allowing some switching in the diffusion term as written in equation (3):

$$M2 = \begin{cases} dc = \left[-\lambda\left(c - c_{ext}\right) + \left(\frac{NS}{V} \times 3.6\mathrm{e}6\right)\right] dt + \sigma_{co2}\, d\omega_c \\ d\lambda = \left[\delta_\lambda K_\lambda + (1 - \delta_\lambda)\right] \sigma_\lambda\, d\omega_\lambda \\ Yco2 = c + \sigma_{Yco2}\, d\omega_{Yco2} \end{cases} . \quad (3)$$

In this formulation, $\delta_\lambda \in [0; 1]$ is a binary input with value 1 if a sudden change in λ is expected (such as window opening or closing) allowing the diffusion term to be multiplied by the constant parameter K_λ. The value of K_λ is unknown and estimated through optimization. In this test case, as we know the user-defined ACR λ, we set $\delta_\lambda = 1$ at each variation in λ and for duration 3 time steps (= 15 min).

In order to implement these models and compute the solving, we use the computer software CTSM [14], which is available on R programming with the package CTSMR. Mathematical description and guidelines about use of the package and results analysis are provided in [15]. The package allows modeling and solving non-linear state space modeling using the Extended

Kalman Filter (EKF) and maximum likelihood optimization. CTSMR has already been used in the field of building energy for heat performances of building [16] and for tracer gas equation solving [6].

4. Model validation and discussion

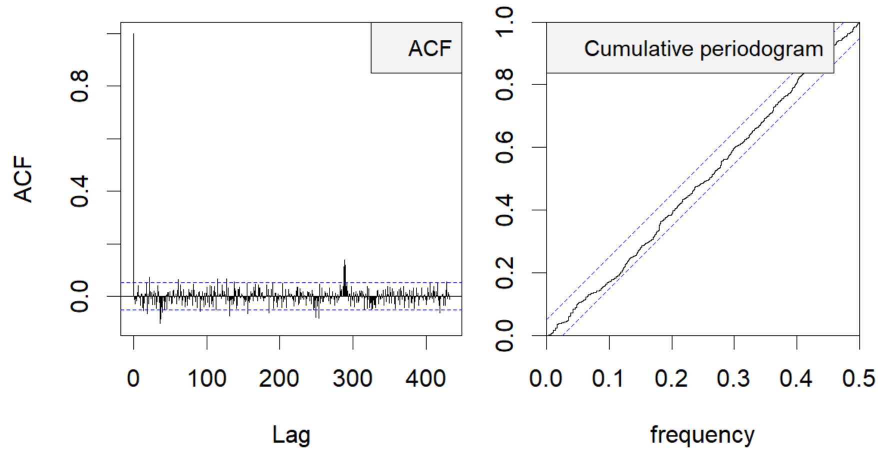

Figure 3. Fit of model M1: Autocorrelation Function (ACF) and cumulative periodogram for $r_{co2} = (Yco2_t - \hat{c}_{t|t-1})$.

4.1. Analysis of the fit of model 1

The model result must be evaluated in order to verify if it meets the modeling assumption and if the estimations are physically reasonable [6, 16]. Figure 3 shows the 1-step ahead prediction error $r_{co2} = (Yco2_t - \hat{c}_{t|t-1})$ statistics. The residuals r_{co2} are not correlated and the cumulative periodogram is close to the diagonal line. Along with the time serie plot of r_{co2} in figure 4, this indicates white noise distribution for residuals r_{co2} and a good enough physical dynamic description in the model.

Table 1. Constant parameters estimated values and performance indicators of fit M1 and M2

	$\hat{\sigma}_{co2}$ [ppm]	$\hat{\sigma}_{\lambda}$ $[h^{-1}]$	$\hat{K}_{\lambda}$	$\hat{\sigma}_{Yco2}$ [ppm]	log-likelihood	RMSE [ppm]
M1	0.25	0.79	-	24.7	-7048	0.328
M2	0.25	0.39	4.11	24.6	-6883	0.252

In this study, the estimated diffusion term for the CO_2 balance state equation $\hat{\sigma}_{co2} = 0.25$ ppm models the white noise uncertainty in both the state equation physical formulation and in the inputs variables (such as in c_{ext} and S). The small value of $\hat{\sigma}_{co2}$ (table 1) emphasizes the fact that the modeling equation is close to the true physic. Nevertheless, this term is expected to take bigger values in more complex driving physics cases or with uncertain measured input data.

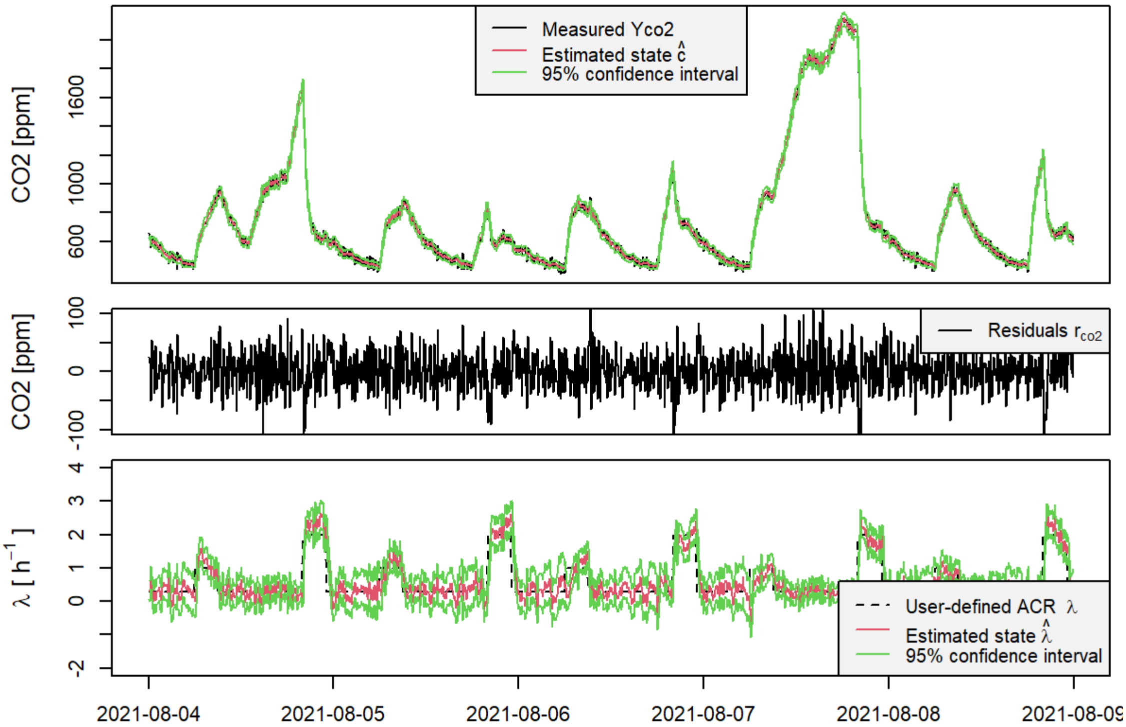

Figure 4. Fit of model M1: time series of state parameters estimation and residuals.

The estimated data serie for state $\hat{c}$ (figure 4) shows that the updating stage of the Kalman process successfully filters the noise in CO_2 measurement and the 95% confidence is close to the mean value. Moreover, the estimated value of the measurement error term $\hat{\sigma}_{Yco2} = 24.7$ ppm (table 1) matches the theoretical measurement noise $\sigma_{meas} = 25$ ppm we chose for generating the data in section 2.

The estimated data serie for state $\hat{\lambda}$ in model M1 (figure 4) demonstrates the ability of the model to fit the true value. This dynamical evolution of $\hat{\lambda}$ is allowed by the diffusion parameter $\hat{\sigma}_\lambda = 0.79$ h^{-1}. As the diffusion term is constant in model M1, this leads to unwanted noise in the phases when λ is constant. On the opposite, when there is sudden variation in λ (typically by a window opening), the allowed dynamic is too slow resulting in lags in the estimation [5].

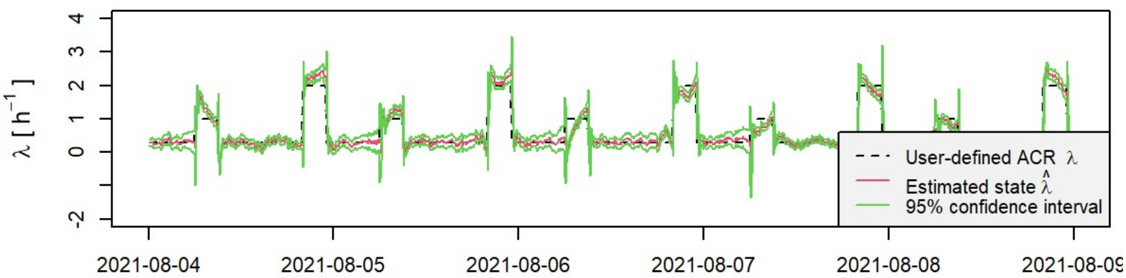

Figure 5. Fit of model M2: time series of state parameter λ estimation.

4.2. Comparison of model M1 and M2

The estimated data serie for state $\hat{\lambda}$ in model M2 (figure 5) shows both better stability during the phases when λ is constant (M1 $\hat{\sigma}_\lambda = 0.79$ h^{-1} > M2 $\hat{\sigma}_\lambda = 0.39$ h^{-1}) and faster dynamical response in the phases when sudden change in λ thanks to the multiplication of the diffusion term by $\hat{K}_\lambda = 4.11$. Parameters $\hat{\sigma}_{co2}$ and $\hat{\sigma}_{Yco2}$ remain similar.

In both models, we observe the effect of the biased input in metabolic rate per person S. We recall that only the mean value $\bar{s}$ is given as input for S as we consider the fluctuating theoretical value unknown. This leads to small errors in ACR λ when there is occupation.

For quantitative performance comparison, CTSMR computes the model global log-likelihood after convergence of the optimization. The log-likelihood of M2 is greater than M1 (table 1). Nevertheless, M2 is a more complex model and thus can not be directly considered a better model. To select the best compromise between complexity and accuracy, we use the likelihood ratio test described in [16]. This statistical tool provides a $p - value$ that quantifies the significance of the model extension compared to the base model. The likelihood ratio test between M1 and M2 reveals a $p - value$ of 0.00 which proves a significant improvement obtained with M2. Moreover, as in this test case we know the user-defined theoretical ACR λ, it is possible to compute the Root Mean Square Error (RMSE) as in equation (4). M2 ($RMSE = 0.252$ h^{-1}) proves to be better than M1 ($RMSE = 0.328$ h^{-1}) (table 1).

$$RMSE = \sqrt{\sum_{t=1}^{N} \frac{\epsilon^2_{\lambda_t}}{N}}, \text{ with } \epsilon_{\lambda_t} = \left(\hat{\lambda}_t - \lambda_t\right) \tag{4}$$

4.3. Maximum error when low CO_2 measurement

After validating and analyzing the performance of both models M1 and M2, we now focus on a specific finding in the results: when indoor CO_2 value c decreases and becomes close to c_{ext}, its 95% estimation confidence interval gets larger (see λ plots figures 4 and 5). This highlights the fact that when $\Delta CO_2 = (c - c_{ext})$ reaches 0, the parameter λ is probably no more identifiable in the CO_2 balance equation (1). This limitation already pointed out by [5, 11] is further investigated in figure 6. We subset all data points according to their associated input ΔCO_2 value. For each subset, and for both models M1 and M2, we plot the distribution of the error $\epsilon_{\lambda_t} = (\hat{\lambda}_t - \lambda_t)$. The result shows that the distribution is tighter for bigger ΔCO_2 which indicates better estimation performance for bigger ΔCO_2 values. It is also visible that model 2 performs better for all ΔCO_2 values.

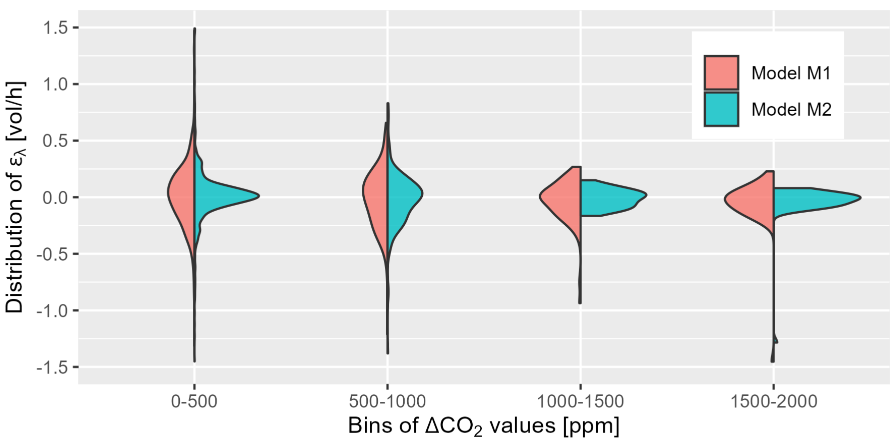

Figure 6. Time series of ϵ_λ distribution against $\Delta CO2$ intervals.

5. Critical analysis and perspectives for future work

5.1. Critical analysis

The investigation of the state space methodology with Kalman filtering for a test case are the result of an exploratory work in methodological approach. The formulations are meant to be robust to uncertainties in physical modeling and measurement discrepancies, as well as sudden changes in ACR λ (see model M2). It is mainly consistent with problems that are close to the CO_2 mass balance assumptions: homogeneous tracer gas concentration in the zone, exclusive exchange with outdoor air, and perfect mixing of outdoor air within the zone [3]. The model robustness needs to be further investigated in real free cooling configurations when:

- Several occupants in different parts of the room, stratification of CO_2 distribution, and non-optimal sensor positions
- Uncertain measurement of occupancy and associated metabolic CO_2 production rate
- Non-negligible physical effects such as air leakage with adjacent rooms or CO_2 buffering from building materials
- ΔCO_2 gets close to 0 (as shown in section 4.3) due to low occupancy or quick removal of CO_2 by a high ACR

5.2. Prospects for future work

Insufficient description of the governing physics or highly biased inputs will lead to a non-gaussian stochastic error and thus wrong estimation in ACR. In order to overcome some of the limitations mentioned in 5.1, several points may be investigated in future work:

- When low ΔCO_2 measurements, variation in exterior concentration c_{ext} may be considered as an input [17]. Yet not covered air change rate weather dependencies may be incorporated through extra state equations including wind speed, wind direction or air temperature measurements [7]. Smaller timescales (for example for wind measurements) may be set [8]
- As reported by several studies [4, 10, 8], the measurement of the window opening position may also be used as an input parameter
- To handle multi-room cases, the formulation may be improved including state equations for each zone, or the neighboring zone being considered as background concentration as in [11]

For exploring these prospects, a more complete test case that fully explores the physical system would be required. Larger data sets could be generated through building simulation software (*EnergyPlus*). We recall that the motivation of this work is in situ diagnostic of real buildings and the state space with Kalman filtering framework is meant to filter physical modeling errors and measurement discrepancies. The method should thus be tested and validated with an experimental case.

5.3. What did not work

It seems useful to the authors to briefly report some modeling attempts that could not be further detailed in this article :

- The state space formulation investigated in [5, 10] which consists in moving the metabolic CO_2 production S from simple input to a measured state did not show better estimation performance
- Using bigger time steps ($dt = 10$ min, 15 min, 20 min) with averaged values did not improve the performance as it induces a lost in information and the state space formulation with Kalman filtering can already deal with noisy data

6. Conclusion

The present work proposes a description and analysis of the two-state CO_2 / *Air Change Rate* model applied to naturally ventilated dynamical air change rate (ACR) estimation. A simple test case with numerically generated data allowed to demonstrate the benefit of improving the model with a variable diffusion term for the air change rate state equation. We also emphasized, among others, the difficulty of using the CO_2 mass balance equation for estimating ACR through window openings when the interior-exterior CO_2 difference value is below 500 ppm.

Future work may thus investigate improvement to the present model by considering multi-zone cases or incorporating measurement of weather data or window opening configurations. The framework may be tested and validated with an experimental case.

References

[1] M. A. M. Brandan, F. D. Espinosa, and T. Tomasetti, "Modeling natural ventilation in early and late design stages: Developing the right simulation workflow with the right inputs," *IBPSA-USA*, p. 8, 2018.

[2] G. A. Faggianelli, A. Brun, E. Wurtz, and M. Muselli, "Assessment of different airflow modeling approaches on a naturally ventilated Mediterranean building," *Energy and Buildings*, vol. 107, pp. 345–354, Nov. 2015.

[3] G. Remion, B. Moujalled, and M. El Mankibi, "Review of tracer gas-based methods for the characterization of natural ventilation performance: Comparative analysis of their accuracy," *Building and Environment*, vol. 160, 2019.

[4] D. Calì, "CO2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings," *Building and Environment*, 2015.

[5] R. Duarte, M. Glória Gomes, and A. Moret Rodrigues, "Estimating ventilation rates in a window-aired room using Kalman filtering and considering uncertain measurements of occupancy and CO2 concentration," *Building and Environment*, vol. 143, pp. 691–700, Oct. 2018.

[6] M. Macarulla, M. Casals, N. Forcada, M. Gangolells, and A. Giretti, "Estimation of a room ventilation air change rate using a stochastic grey-box modelling approach," *Measurement*, vol. 124, pp. 539–548, Aug. 2018.

[7] S. Rouchier, "Interacting multiple models for learning the air change rate and occupancy of a room from co2 measurements," *Roomvent*, p. 4, 2020.

[8] S. Wolf, D. Calì, J. Krogstie, and H. Madsen, "Carbon dioxide-based occupancy estimation using stochastic differential equations," *Applied Energy*, vol. 236, pp. 32–41, Feb. 2019.

[9] Y. Li, J. Castiglione, R. Astroza, and Y. Chen, "Real-time thermal dynamic analysis of a house using RC models and joint state-parameter estimation," *Building and Environment*, vol. 188, p. 107184, Jan. 2021.

[10] S. Liu, Y. Jian, J. Liu, R. Guo, and W. Zhu, "Associating occupants' interaction with windows with air change rate –One case study," *Building and Environment*, vol. 222, p. 109387, Aug. 2022.

[11] G. Remion, B. Moujalled, and M. El Mankibi, "Dynamic measurement of the airflow rate in a two-zones dwelling, from the CO2 tracer gas-decay method using the Kalman filter," *Building and Environment*, vol. 188, p. 107493, Jan. 2021.

[12] AFNOR, "Ventilation for buildings - Design and dimensioning of residential ventilation systems," 2006.

[13] M. Brabec and K. Jílek, "State-space dynamic model for estimation of radon entry rate, based on Kalman filtering," *Journal of Environmental Radioactivity*, vol. 98, pp. 285–297, Dec. 2007.

[14] R. Juhl, J. K. Møller, and H. Madsen, "Ctsmr – Continuous Time Stochastic Modeling in R," p. 11, 2016.

[15] N. R. Kristensen, H. Madsen, and S. B. Jørgensen, "Parameter estimation in stochastic grey-box models," *Automatica*, vol. 40, pp. 225–237, Feb. 2004.

[16] P. Bacher and H. Madsen, "Identifying suitable models for the heat dynamics of buildings," *Energy and Buildings*, vol. 43, pp. 1511–1522, July 2011.

[17] S. Park, P. Choi, D. Song, and J. Koo, "Estimation of the real-time infiltration rate using a low carbon dioxide concentration," *Journal of Building Engineering*, vol. 42, p. 102835, Oct. 2021.