Data driven models for fault detection - Combining thermal and indoor air quality grey box models

Gabriel Rojas[*1], Romed Jenewein[1], Klaus Prenninger[2], and Johannes Schnitzer[3]

1 Unit for Energy Efficient Building & Digital Science Center, University of Innsbruck
Technikerstrasse 13
Innsbruck, Austria
**Corresponding author: gabriel.rojas@uibk.ac.at*

2 Smart Building Department
Salzburg University of Applied Sciences
Am Markt 156
Kuchl, Austria

3 Forschung Burgenland
Steinamangerstraße 21
Pinkafeld, Austria

ABSTRACT

The progressive digitalization is providing more and more measurement data from building operation, in particular from heating, cooling and ventilation (HVAC) systems. This work investigates the potential use of data-driven models to simulate indoor environmental conditions, i.e. temperature and CO_2 concentration, for fault detection applications. Herein, a grey-box model, depicting the thermal behaviour of building zones, is coupled with model representing the indoor air quality/ventilation condition in the respective zone allowing the combined use of measurement data from building operation. The models are applied to an office room of a case study building and the model parameters are identified with measurement data for a four-weeks long training period. The identified models are used to predict the timely evolution during a three-day long prediction period. By comparing residual metrics between training phase and prediction phase the model's capabilities to detect simple exemplary faults are evaluated. Herein, preliminary results with rather simple fault cases, like temperature or CO_2 sensor faults or (unintentionally) left-open windows are investigated. Results indicate that indoor temperature anomalies are detected well and that anomalies in CO_2-concentration are also detectable with this modelling approach but depend on the available occupancy estimation (or measurement). Further investigations are underway to test possible adaptions to the presented approach to allow for better occupancy estimation and/or account variable ventilation rates.

KEYWORDS

Fault detection, grey-box models, ventilation model, thermal model, CO_2 model, parameter identification

1 INTRODUCTION

The progressive digitalization is providing more and more measurement data from building operation, in particular from heating, cooling and ventilation (HVAC) systems (Kim and Katipamula 2018). The data is most often used for control and/or monitoring on component level. However, the combined use of measurement data from HVAC components, weather station and building volume (e.g. zone temperature, CO_2 concentration) opens new possibilities. This work investigates the potential use of data-driven models to simulate indoor environmental conditions, i.e. temperature and CO_2 concentration, for fault detection applications. Herein, grey-box (RC network) models, depicting the thermal behavior of building zones, are coupled with simplified models representing indoor air quality condition in the respective zone allowing the combined use of measurement data from building operation. The aim is to use such models to detect anomalous conditions stemming from faulty or misadjusted HVAC components or (unintentional) misused of building, e.g. window left open, misadjusted air inlet valve, etc., by comparing model predictions with current

measurement values. This paper presents first results from training such a combined model for a real case study office room and comparing predictions for periods with artificial or real faults.

2 MODEL DESCRIPTION

So-called grey box models were used for depicting the dynamic behaviour of the building (or the room in this case) (Bacher and Madsen 2011; Bauwens, Ritosa, and Roels 2021; Reynders, Erfani, and Saelens 2021). The thermal behaviour of buildings can be described by a series of first-order differential equations. These can be represented as simple resistance-capacity networks. The level of detail of the model can be determined by the number of dependent parameters. In the course of the case study, different models were applied. Simpler models have the advantage of having fewer model parameters which can be identified more reliably. However, oversimplified models might not be able to reproduce the dynamic behaviour sufficiently well. Starting from a simple RC model with only one capacity (1C), the complexity of the thermal model is increased by introducing a second capacity representing the thermal capacity of the external walls (2C). In a third step the 2C thermal model is coupled with a model representing the mass balance equation for CO_2 in the room, i.e. a ventilation model (2C+V). The mass balance model and the thermal model are coupled via the air exchange rate. In the following, these three models and their respective differential equations are documented in more detail.

2.1 1C Thermal model

This model is one of the simplest modelling variants. In this case, only the thermal behavior of the building is considered. The building envelope is represented by one resistance to the outside and has only one capacity to account for thermal mass, see Figure 1. It therefore has two free parameters (resistance R_{ie} and capacity C_i).

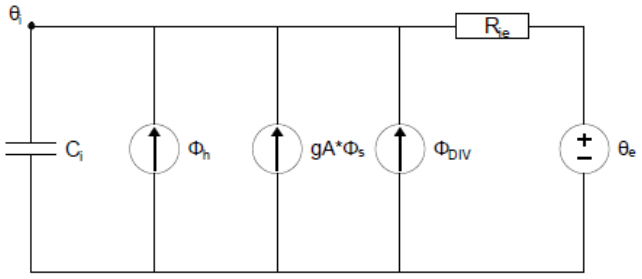

Figure 1: Resistance-Capacity network representing the 1C thermal model.

The model shown in Figure 1 can also be written as a differential equation as follows.

$$\frac{d\theta_i}{dt} = \frac{\theta_e - \theta_i}{C_i * R_{ie}} + \frac{1}{C_i} * \Phi_h + \frac{gA}{C_i} * \Phi_s + \frac{1}{C_i} * \Phi_{DIV} \quad (1)$$

It describes the energy balance of the internal room temperature node θ_i. The other variables are described in section 2.3. This modelling variant has proven to be robust in determining a good initialization value for the parameters.

2.2 2C Thermal model

To increase the level of detail of the model, another differential equation can be added. In this model, as shown in Figure 2, the external resistance is divided. Thus, the thermal capacity of the building can be considered better. As a result, this model has two more free parameters that need to be identified using measurement data.

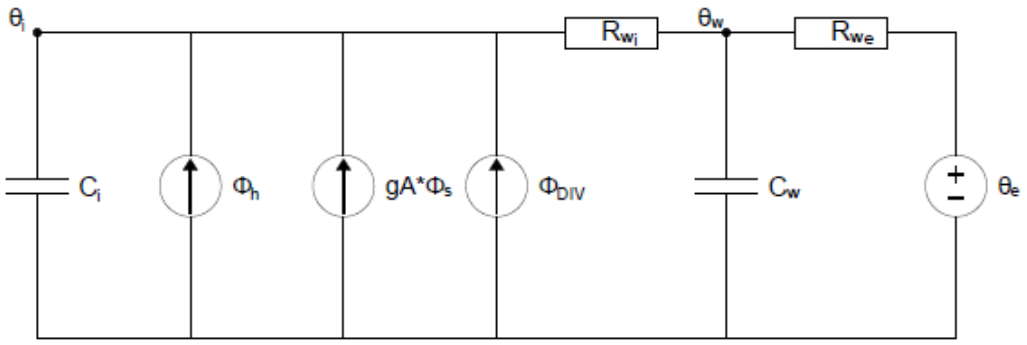

Figure 2: Resistance-Capacity network representing the 2C thermal model.

This model can be described by an additional differential equation representing the energy balance of the external wall node θ_w.

$$\frac{d\theta_i}{dt} = \frac{\theta_w - \theta_i}{C_i * R_{wi}} + \frac{1}{C_i} * \Phi_h + \frac{gA}{C_i} * \Phi_s + \frac{1}{C_i} * \Phi_{DIV} \tag{2}$$

$$\frac{d\theta_w}{dt} = \frac{\theta_i - \theta_w}{C_w * R_{wi}} + \frac{\theta_e - \theta_w}{C_w * R_{we}} \tag{3}$$

See the following section for a variable description.

2.3 2C+V Thermal and ventilation model

The two models above can describe the thermal behaviour of a building or a room. However, the heat flow through massive walls is, compared to heat flow through e.g. ventilation, considerably slower. To be able better reproduce dynamic events, like the influence of a controlled or uncontrolled air exchange with the ambient, the thermal model from section 2.2 can be supplemented with a ventilation model. This model describes the mass balance of CO_2 in the room and is shown Figure 3.

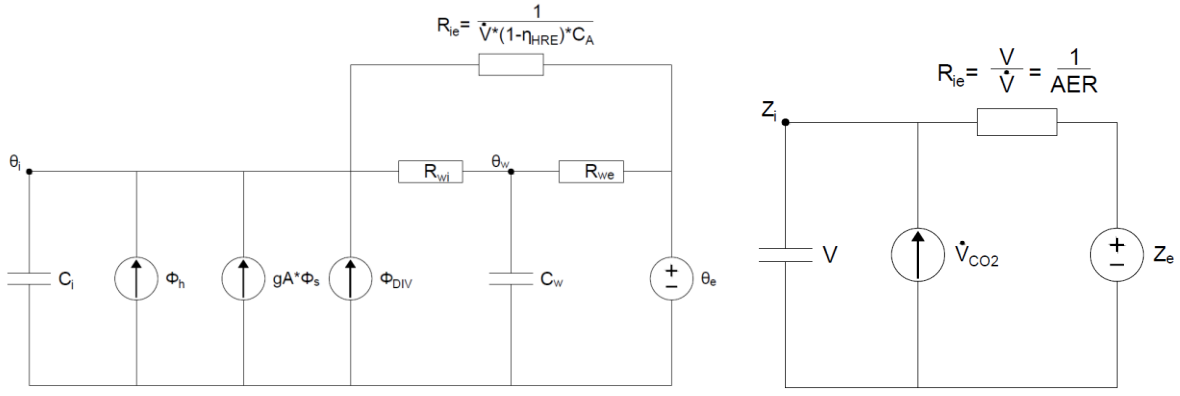

Figure 3: Resistance-Capacity network representing the 2C+V model. Left: Representation of thermal behaviour, i.e. heat flow through building/room envelope. Right: Representation of CO_2 concentration in the building/room.

The models in Figure 3 are coupled via the volume flow. In the thermal model, the volume flow is considered as part of an additional resistance. The idea behind this resistor is to consider the heat flows of the opaque envelope and the ventilation separately. Infiltration losses are estimated (stationary) and included in the effective heat recovery efficiency of the ventilation. The n_{50}-value of the room is estimated (1 h^{-1}) based on the building n_{50}-value.

$$\frac{d\theta_i}{dt} = \frac{\theta_w - \theta_i}{C_i * R_{wi}} + \frac{\theta_e - \theta_i}{C_i} * (1 - \eta_{HRE}) * \dot{V}_{AL} + \frac{1}{C_i} * \Phi_h + \frac{gA}{C_i} * \Phi_s + \frac{1}{C_i} * \Phi_{DIV} \quad (4)$$

$$\frac{d\theta_w}{dt} = \frac{\theta_i - \theta_w}{C_w * R_{wi}} + \frac{\theta_e - \theta_w}{C_w * R_{we}} \quad (5)$$

$$\frac{dz_i}{dt} = \frac{\dot{V}_{CO2}}{V} + (z_e * z_i) * \frac{\dot{V}_{AL}}{V} \quad (6)$$

The differential equations are composed of known input data, unknown parameters, and state variables. The variables for equations 1 through 6, are described here.

State variables:

- θ_i … interior temperature [°C]
- θ_w … wall temperature [°C]
- z_i … CO_2 concentration in the interior [ppm]

Known input data:

- θ_e … external temperature [°C]
- z_e … external CO_2 concentration [ppm]
- η_{HRE} … effective heat recovery efficiency [-]
- Φ_h … heat load [kW]
- Φ_s … solar radiation [kW]
- Φ_{DIV} … combination of diverse heat flow's (internal gains from people; internal gains from electrical devices; the heat flow between the neighbour rooms) [kW]
- $\dot{V}_{CO2}$… CO_2 volume flow caused by people present [m^3/h]
- V ... room volume [m^3]

Unknown(free) parameters:

- C_i … interior capacity [kWh/K]
- C_w … wall capacity [kWh/K]
- R_{wi} … external wall resistance on inner side [K/kW]
- R_{we} … external wall resistance on exterior side [K/kW]
- gA … solar aperture [m^2]
- $\dot{V}_{AL}$ … volume flow between the inside and the outside [m^3/h]

Furthermore, for the training of the model and thus the determination of the unknown parameters, measurement equations are needed. For the 2C+V model the two trivial equations, where the observed state variables are set equal to the measured data are needed as documented in formula (7) and (8).

$$\theta_i = \theta_{i,measured} \quad (7)$$

$$z_i = z_{i,measured} \quad (8)$$

3 METHODS - MODEL TRAINING

3.1 Case study description

The described models were trained using measurement data from on office building in eastern Austria. The building belongs to the research institute Forschung Burgenland and is equipped with a great number of sensors, including room sensors for temperature, relative humidity, CO_2-concentration, supply and extract air flows, supply air temperature as well as an outdoor weather station. The office room number "02" was used for the following analysis. It is usually occupied by one person (occasionally by two) on three to five working days per

week. The nominal ventilation rate in this room is 55 m³/h. Room "02" is located in the upper floor and its glass façade faces south-south-east, see Figure 4. The relevant parameters are summarized in

Table 1.

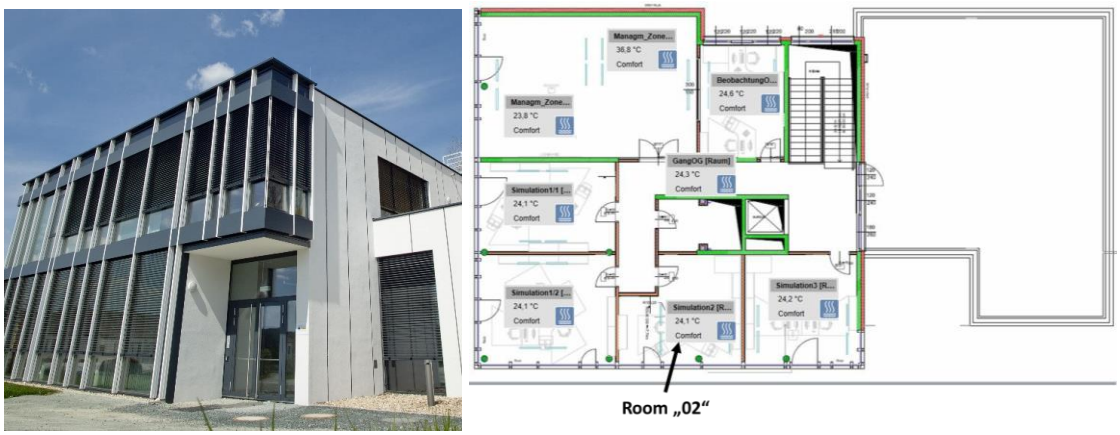

Figure 4: Left: Picture of the south-south-east façade of the case study building "Energetikum". Right: Upstairs floorplan with investigated room "02". Source: Forschung Burgenland.

Table 1: Building and room parameters and assumptions

Parameters & assumptions	Value	Unit
Area external wall: Room "02"	18.8	[m²]
U-value external wall: Room "02"	1.10	[W/(m²K)]
Floor area: Room "02"	26.04	[m²]
Height: Room "02"	3.17	[m]
Average U-value: Entire building	0.39	[W/(m²K)]
Avg. thermal cap. estimate: Entire building	204	[Wh/(m²K)]
Occupancy assumption (7:00-16:00)	5	[d/week]
Occupant sensible heat rate	76	[W]
Occupant CO_2 emission	18	[L/h]

Table 2: Initial values and limiting bounds for free model parameters

Parameter		Source for initial value	Initial-value	Units	MIN	MAX
R	Resistance	Energy Performance Cert.	48.4	[K/kW]	1	500
Ci(+Cw)	Capacity	Estimate [PHPP]	5.3	[kWh/K]	0.01	500
L	Position of the wall-capacity	Estimate	0.50	[-]	0.1	0.9
gA	solar aperture	Estimate	5.00	[m²]	0.1	50
$\dot{V}_{AL}$	airflow	Measurement data	mean	[m³/h]	50	300
HLC	Heat-Loss-Coefficient	=1/(R)	20.7	[W/K]		

3.2 Parameter identification

In order to obtain a suitable model that can be used for simulation, the unknown/free parameters described in section 2 had to be identified. For this purpose, the System Identification Toolbox from the software MATLAB was used. Therein the grey-box estimation algorithm was used to identify the free parameters that minimise the cost function,

i.e. the sum of squares of the difference between simulated and measured values from the training period. Initial values and limiting bounds were defined for the unknown parameters according to Results from the model parameter estimation are documented in Table 3.
. These should lie within a physically reasonable range. For the 2C models the resistance R is split using the factor L, representing the position of the effective wall capacity C_w. Results from the model parameter estimation are documented in Table 3.

Table 3: Estimated model parameters for 1C, 2C and 2C+V model. Initial value and design value (if known) are also listed for comparison.

Parameter		Initial-value	1C estimation	2C estimation	2C+V estimation	Design value	Units
HLC	Heat-Loss-Coefficient	20.7	14.5	16.6	9.6	20.7	[W/K]
Ci(+Cw)	Capacity	5.3	4.8	8.7	19.2	n.a.	[kWh/K]
L	Position of the wall-capacity	0.50	n.a.	0.67	0.39	n.a.	[-]
gA	solar aperture	5.00	1.45	1.9	2.6	unknown	[m²]
$\dot{V}_{AL}$	airflow	4	n.a.	n.a.	134	55	[m³/h]

3.3 Exemplary tests for fault detection (FD)

In order to test the usability of the proposed combined model (2C+V) for FD the trained models were used to predict the temperature and CO_2 concentration for a three-day long period immediately after a four-weeks long training period. Note that the presented prediction is a simulation for the entire three-day prediction period and not a 1-step ahead prediction as often used in this context. As concluded in previous studies, evaluating the multi-step prediction (in this case even for three days), might provide a better model assessment (Reynders et al. 2021). The simulation was initialized with measured data at the beginning of the prediction period.
This paper presents the first results for testing with rather “simple” fault cases, see Table 4. They were produced artificially, by altering the measurement data, simulating e.g. a faulty sensor, or, they occurred in reality, as for the open window “fault”. Note that these simple faults may also be detected with simpler FD-schemes, e.g. rule-based FD techniques. However, the proposed method with physically based grey-box models has the potential to provide further capabilities like fault diagnosis, e.g. by continuously retraining the models and interpreting the identified parameters. Further, more complex and more realistic faults will be tested in this ongoing work.

Table 4: Preliminary fault detecation test cases

Fault		Training start	Prediction start	Fault duration
Temperature sensor faulty	Sensor outputs +2 K (artificial)	2022-11-03 11:15	2022-12-01 11:15	60 hrs
CO_2 sensor faulty	Sensor outputs +200 ppm (artificial)	2022-01-27 11:15	2022-02-24 11:15	60 hrs
Window left open	Window was left open unintentionally (real occurrence)	2022-09-01 7:40	2022-09-29 7:40 2022-09-30 7:40	8 hrs 6 hrs

4 RESULTS

Using the thermal models one can detect only anomalies that alter the thermal behaviour, i.e. the temperature sensor fault and the open window in herein presented cases. Looking at evolution of the measured and simulated temperature during the training month one can see that the simple model with one thermal capacity (1C) follows the general trend fairly well, but

cannot follow the daily temperature variation fully, see Figure 5 (left). The 2C model is able to reduce the root-mean-square of the error/residuals (RMSE) by reproducing the daily variations slightly better, see Figure 5 (right). The temperature simulation of the 2C+V model is very similar to the 2C model, see Figure 6 (left). In all three cases the temperature sensor offset of +2 K can easily be identified by comparing the RMSE between prediction and training period. The RMSE increases by around 200%, i.e. triples, with all three models, see Table 5.

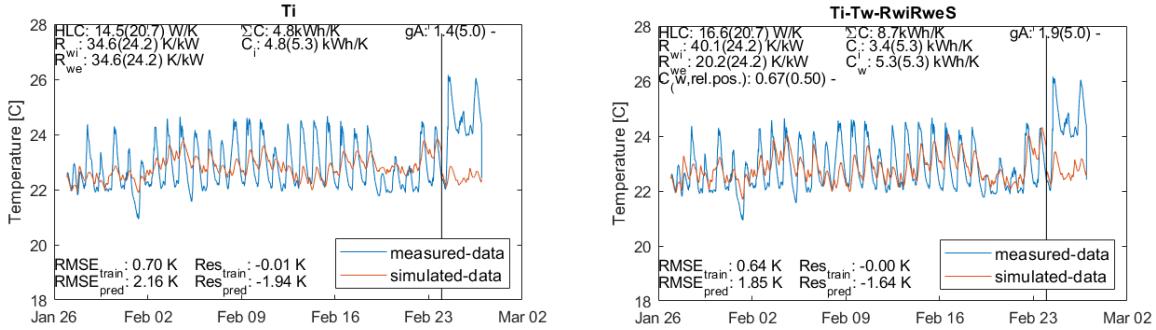

Figure 5: Result for 1C model (left) and 2C model (right) for temperature sensor fault case: Measured and simulated indoor temperature evolution during training (4 weeks) and prediction (3 days) phase. The end of training phase is marked with a vertical line.

As seen in Figure 6 (right) the simulated CO_2-concentration can reproduce the measured CO_2-concentration roughly, but the parameter identification algorithm seems to overestimate the ventilation rate (134 m³/h). Therefore, the simulated daily peaks fall short compared to the measured ones. Nevertheless, the CO_2-concentration is reproduced in a decent manner, considering that the occupancy (and the corresponding emission rate) is just based on the naïve assumption using a fix working schedule (Mo-Fr, 7:00-16:00) for the presented results. One of the difficulties in this case is, that the ventilation rate is controlled via schedule and CO_2 concentration and is therefore not a fix value. However, a varying ventilation rate cannot be depicted in this model variant. This issue was addressed in a model extension, see below. With that, a ventilation rate of around 90-100 m³/h was estimated during regular working hours. Note that the measurements provided by the building control sensors provided unrealistic values of around 10-15 m³/h.

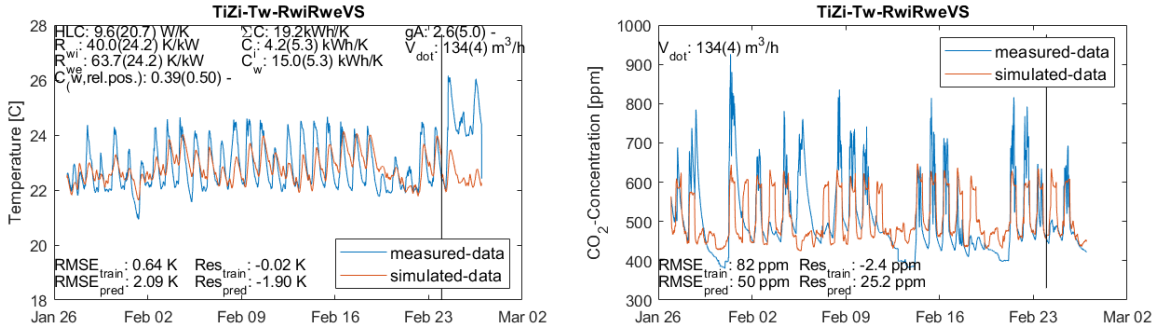

Figure 6: Result for 2C+V model for temperature sensor fault case: measured and simulated indoor temperature (left) and CO_2 concentration (right) evolution during training (4 weeks) and prediction (3 days) phase. The end of training phase is marked with a vertical line.

Figure 7 shows the prediction results for the case where a faulty CO_2 sensor is simulated. While the RMSE of the prediction period shows no increase, the RMSE of the predicted CO_2 concentration increases by almost 300%, allowing a fault detection if an occupancy estimate (or measurement) is available, see Table 5. When only rough occupancy estimates are available, then a prolonged prediction period would be advisable to average out the rough estimate. See “open window” test case, where inaccurate occupancy estimate hinders the fault

detection. Note that this model can be altered to allow for a prediction of CO_2 emission rate instead of indoor concentration, see below.

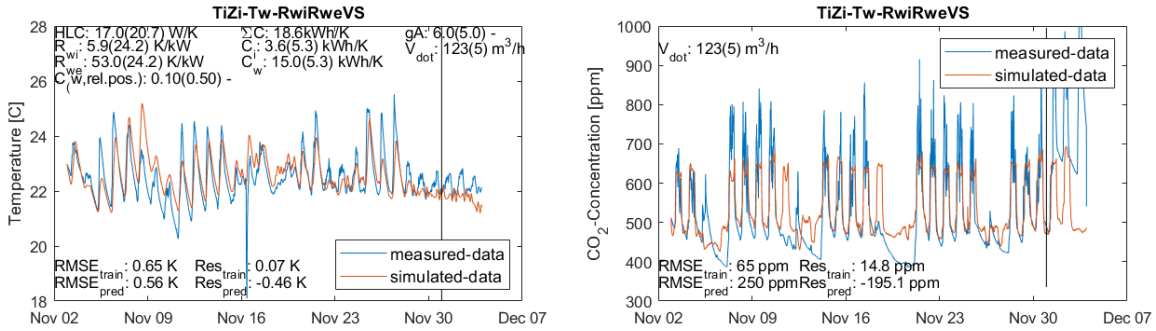

Figure 7: Result for 2C+V model for CO_2 sensor fault case: measured and simulated indoor temperature (left) and CO_2 concentration (right) evolution during training (1 month) and prediction (3 days) phase. The end of training phase is marked with a vertical line.

The results for the "open window fault case" is shown in Figure 8. As can be seen, the temperature drop due to the open window is clearly visible and increases the residuals (RMSE) between simulation and measurement substantially, see also Table 5. Looking at Figure 8 (right) one can also see that the difference between simulated and measured CO_2-concentration is rather large. However, due to the fact that the occupancy estimate is rough and due to the fact that the volume flow control is not depicted in this model, the RMSE is also rather high for the training period. Therefore, the RMSE increase of the CO_2-simulation cannot be used to detect the open window state for this particular case. Possibilities to increase CO_2 prediction accuracy by allowing better occupancy estimates are currently investigated further.

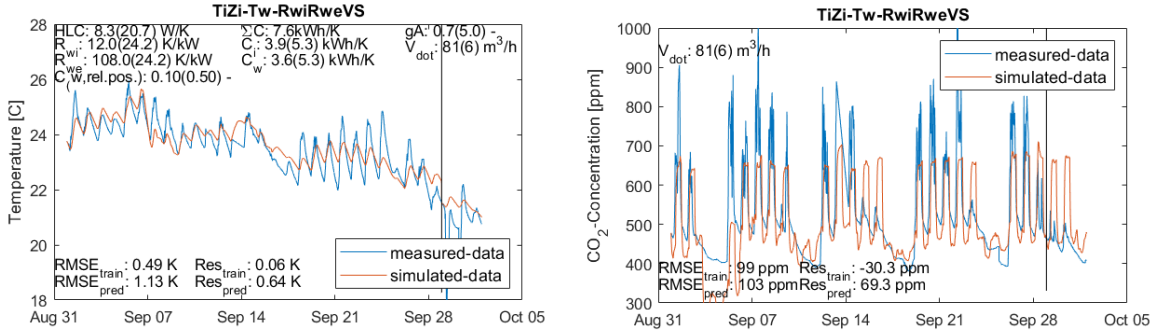

Figure 8: Result for 2C+V model for "open window fault" case: measured and simulated indoor temperature (left) and CO_2 concentration (right) evolution during training (1 month) and prediction (3 days) phase. The end of training phase is marked with a vertical line.

Table 5: Summary table of prediction results. The fields were coloured as follows. Green: fault present and RMS increase >100%; yellow: fault present and RMS increase >33%; red: fault present and RMS increase <33%.

		Failure type					
		T sensor offset (February)		**CO2 sensor offset (November)**		**Window left open (September)**	
		T [K]	CO2 [ppm]	T [K]	CO2 [ppm]	T [K]	CO2 [ppm]
1C	RMS train.	**0.70**	n.a.	0.74	n.a.	**0.51**	n.a.
	RMS pred. w/o fault	0.65	n.a.	0.37	n.a.	n.a.	n.a.
	RMS pred. with fault	**2.16**	n.a.	0.37	n.a.	**1.10**	n.a.
	RMS increase	**209%**	n.a.	-50%	n.a.	**116%**	n.a.
2C	RMS train.	**0.64**	n.a.	0.70	n.a.	**0.77**	n.a.

	RMS pred. w/o fault	0.52	n.a.	0.42	n.a.	n.a.	n.a.
	RMS pred. with fault	**1.85**	n.a.	0.42	n.a.	**1.10**	n.a.
	RMS increase	**189%**	n.a.	-40%	n.a.	**43%**	n.a.
2C-V	RMS train.	**0.64**	82	0.65	**65**	**0.49**	**99**
	RMS pred. w/o fault	0.58	50	0.56	137	n.a.	n.a.
	RMS pred. with fault	**2.09**	50	0.56	**250**	**1.13**	**103**
	RMS increase	**227%**	-39%	-14%	**285%**	**131%**	**4%**

5 OUTLOOK

5.1 Stepwise parameter identification

Previous work has shown that physical parameter identification can be challenging (Rojas et al. 2023). The tighter that certain free parameters, e.g. building thermal mass, can be bounded, the easier it is for the algorithm to find physical sensible values, in particular for more complex models with higher number of free parameters. In an ongoing work, the following stepwise identification procedures are being compared:

- IC1 (as used in for the results presented in section 4): The starting values of the parameters are defined via the energy performance certificate of the case study whenever possible and the bounding limits are left fairly loose.
- IC2: The parameters are first identified as in method IC1, but with training data in the winter period. The value of these parameters then serves as the starting value for the actual parameter identification. The lower and upper limit are defined +/-95% from this determined value in the winter period.
- IC3: With this method, the parameters are first determined with the simple model from section 2.1 (1C). The identified values are used as initial values and for defining boundaries (+/- 95%). Then the same procedure as within method IC2 is followed.

5.2 Emission rate prediction

The 2C+V model can be adapted to output CO_2-emission rate instead of the CO_2-room concentration. This could be used to extract room occupancy for improving FD prediction models or other application such as model predictive control. Figure 9 shows preliminary results with such an adapted model for the “open window test case”.

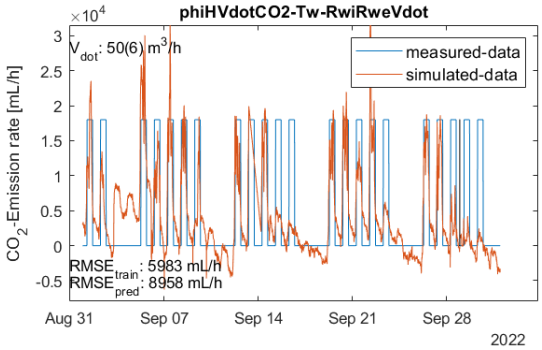

Figure 9: Result for adapted 2C+V model outputting CO_2 emission rate for “open window fault” case. “Measured value” refers to the naïve occupancy assumption. A validation with an actual measured occupancy profile has not been performed yet.

5.3 Accounting variable ventilation rates

As noted above, the 2C+V model has the limitation, that only a fix/static ventilation rate can be modelled and identified. Introducing model states with a modified external temperature

and modified external CO_2 concentration which use the ventilation flow control signal to alter the measured values accordingly, are being tested to account variable ventilation rates.

6 CONCLUSIONS

This paper documents work-in-progress for combining a thermal grey-box model with a simple ventilation mass balance model, i.e. a model that reproduces the CO_2 concentration in a room. Both models are coupled via the ventilation rate. The models are applied to an office room of a case study building. The model parameters are identified with measurement data for a four-weeks long training period. Their values compare well with design values from the energy certificate. The identified models are used to predict the timely evolution during a three-day long prediction period. By comparing residuals (measurement minus simulation) between training phase and prediction phase, this work investigates the model's capabilities to detect simple exemplary faults, like temperature or CO_2 sensor faults or (unintentionally) left-open windows. The preliminary results show that larger indoor temperature anomalies can be detected well with this approach. Anomalies in CO_2-concentration are also detectable but depend on the available occupancy estimation (or measurement). It should be noted that the herein investigated faults are simplified and additional testing is needed. Further investigations are underway to test possible adaptions to the presented approach to allow for better occupancy estimation and/or account variable ventilation rates.

7 ACKNOWLEDGEMENTS

This work was financed through the Austrian Research Promotion Agency (FFG) in the framework of the funding program "Stadt der Zukunft" by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK) (Grant Number 879437).

8 REFERENCES

Bacher, Peder, and Henrik Madsen. 2011. "Identifying Suitable Models for the Heat Dynamics of Buildings." *Energy and Buildings* 43(7):1511–22. doi: 10.1016/j.enbuild.2011.02.005.

Bauwens, Geert, Katia Ritosa, and Staf Roels. 2021. *IEA EBC Annex 71: Building Energy Performance Assessment Based on in-Situ Measurements - Physical Parameter Identification*.

Kim, Woohyun, and Srinivas Katipamula. 2018. "A Review of Fault Detection and Diagnostics Methods for Building Systems." *Science and Technology for the Built Environment* 24(1):3–21. doi: 10.1080/23744731.2017.1318008.

Reynders, Glenn, Arash Erfani, and Dirk Saelens. 2021. *IEA EBC Annex 71: Building Energy Performance Assessment Based on In-Situ Measurements - Building Behaviour Identification*. KU Leuven.

Rojas, G., S. Metzger, M. Blöchle, M. Šipetić, R. Jenewein, and S. Öttl. 2023. *IEA Energie in Gebäuden Und Kommunen (EBC) Annex 71 : Bewertung Der Gebäudeenergie- Effizienz Mit Hilfe Optimierter in Situ Messverfahren*.