USING MACHINE-LEARNING ALGORITHMS TO IMPROVE CO_2-BASED DEMAND-CONTROLLED NATURAL VENTILATION

Abstract

Although natural ventilation is applicable to most buildings, architects today struggle to integrate it as an alternative to mechanical ventilation systems. This paper presents the application of regression algorithms and identification methods to single-sided natural ventilation with 1 opening (SS1). The result of this work provides a predictive control-oriented model of CO_2 demand-controlled natural ventilation with SS1 configuration.

Authors

Wei Zhang, Wentao Wu, and Ali Malkawi
Harvard University

Keywords

Single-sided natural ventilation, CO_2 demand control, regression algorithms, system identification

Introduction

Natural ventilation can alleviate sick building syndrome (SBS) and improve the occupants' health in a building. Researchers at Lawrence Berkeley National Laboratory found that "CO_2 is approximately correlated with other indoor pollutants that may cause SBS symptoms" (Erdmann, et al., 2002). Actually, CO_2-based demand control ventilation (DCV) is realized in mechanical ventilation systems (Emmerich & Persily, 2001; Emmerich, et al., 2001). It raises a dilemma: As claimed in the *LEED Handbook* (Kubba, 2016), "ventilation systems themselves can be a source of indoor pollution and contribute to indoor air problems, if they are poorly designed, operated, or maintained."

Most previous studies attempted to find a general expression for single-sided natural ventilation flow rate through computational fluid dynamics (CFD) and experiments (Larsen & Heiselberg, 2008). In CFD, the built environment is simplified into a laboratory setting; while the window is simplified as an opening without shading and thickness. The tracer gas method is an effective experimental approach, but the research is not often in real-life scenarios (Erhart, et al., 2015; Tang, et al., 2016).

This paper considers the inadequacy in previous studies. It implements the machine-learning approach in predictive control-oriented modeling. It evaluates the integration of CO_2 demand control in natural ventilation.

Methodology

EXPERIMENTAL SETTING

The experiment was conducted at a newly retrofitted three-story house in Cambridge, Massachusetts. The south-facing office (Figures 1 and 2) on its third floor was selected for this study.

This office has the following dimensions: floor surface = 10 m^2 (depth: 3.85 m, width: 2.6 m) and room volume = 30 m^3. It has four windows and one skylight. In this study, only the left bottom window is utilized manually in left side-hung mode (Figure 3). Its opening is limited to 61 cm in height and 61 cm in width for ventilation. All other openings are kept closed

The sensors used in the study are:

— The weather station, which provides the outdoor temperature, wind direction, and wind speed measurement.
— Three smart sensors provide the CO_2 concentration and temperature (Figure 4).

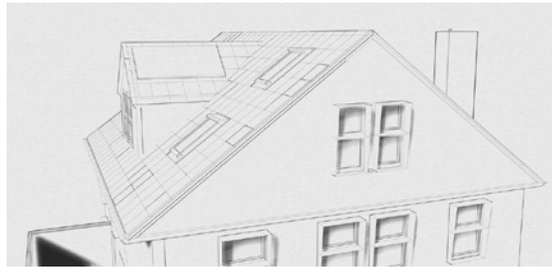

Figure 1: South facade of experiment room.

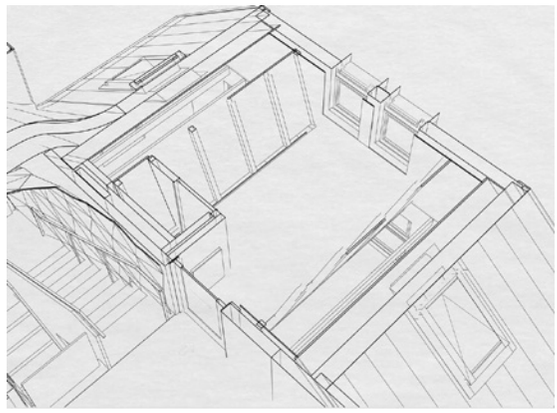

Figure 2: Inside view of the office.

Figure 3: Window configuration. Left bottom window is engaged, with a smart sensor on the window bench.

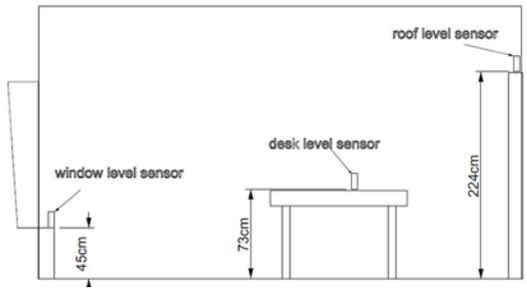

Figure 4: Location of smart sensors.

One of two persons work inside the room. The presence of occupancy is verified through the occupancy sensor in the room. Table 1 shows the sensor summary.

Sensor	Location	Range and Accuracy
Outdoor temperature sensor	On the roof	Range: from -30 °C to +55 °C
Wind Direction sensor	On the roof	Range: from 0° to 359°; Accuracy: < ± 2 °C
Wind Speed sensor	On the roof	Range: from 0 m/s to 75 m/s
Indoor CO_2/ temperature sensors	Inside the room	Temperature ranges from: 0 °C to 50 °C Accuracy: ± 0.3 °C
		CO_2 ranges from: 0 to 5,000 ppm Accuracy: ± 50 ppm (from 0 to 1,000 ppm) or ± 5% (from 1,000 to 5,000 ppm)

Table 1: Sensor summary.

The experiment combines the CO_2 tracer gas decay method and CO_2 demand-controlled ventilation in two types:

Type 1 (see Figure 9):

— Occupant sits inside with windows closed.
— When the CO_2 concentration increases to 1000 or 1500 ppm, the occupant opens the window completely, as shown in Figure 3. After 15 or 20 minutes, the occupant closes the window.
— The actions of opening and closing the window are repeated until the end of the experiment.

Type 2 (see Figure 8):

— Occupant sits inside with windows closed.
— When the CO_2 concentration increases to 1000 or 1500 ppm, the occupant opens the window completely, as shown in Figure 3.
— The CO_2 concentration decays until the end of the experiment.

FLOW RATE ESTIMATION

The change of CO_2 concentration due to natural ventilation is expressed:

$$v\frac{dC_z}{dt} = nG - V_{NV}(C_z - C_{oa})$$

v The volume of room space (m^3)
C_z The CO_2 concentration in the experiment room (ppm)
n Number of persons inside the room
G The CO_2 generation by the one occupant inside the room: approximately 11 (ppm/min)
V_{NV} The natural ventilation flow rate (m^3/min)
C_{oa} The CO_2 concentration in local atmosphere: (400 ppm)

The time interval is 1 minute; the derivative of CO_2 concentration is approximately replaced by the mean value of the CO_2 concentration change over the time interval:

$$\frac{dC_z}{dt} = \frac{C_z^{k+1} - C_z^k}{\Delta t}$$

C_z^{k+1} The CO_2 concentration (ppm) at k+1 moment
C_z^k The CO_2 concentration (ppm) at k moment
Δt The time interval between k+1 moment and k moment (1 min)

The mean natural ventilation rate, V_{nv}, can be calculated mathematically through the following equation:

$$v\frac{C_z^{k+1} - C_z^k}{\Delta t} = nG - \overline{V_{NV}}(C_z - C_{oa})$$

The CO_2 concentration, C_z, is calculated as the sample mean of the 3 smart sensors.

According to previous researches (Larsen & Heiselberg, 2008; Tang, et al., 2016), the empirical formula of single-sided ventilation can be simplified as:

$$V_{nv} = A\sqrt{C_1 \cdot U_{ref}^2 + C_2 \cdot \Delta T + C_3 \cdot \frac{\Delta T}{U_{ref}^2}}$$

A The total area of window
Δt The difference of room air temperature and outside air temperature
U_{ref} The reference wind speed from weather station
C_1, C_2, C_3 The coefficients including the details of wind incidence angle, geometry of window, etc.

This equation provides a group of features, namely, V_{nv}, V^2_{nv}, U^2_{ref}, U_{ref}, and ΔT, for regression.

REGRESSION ALGORITHMS

The regression algorithms employed in this paper are linear regression and polynomial regression (Montgomery, et al., 2012). The outputs of the regression algorithms contain the flow-rate expression and a group of selected features; the details of the output are in the results section.

CONTROL-ORIENTED MODELING

A linear time-invariant system dynamics is the popular prerequisite condition for implementation of control design, for example, model predictive control (MPC) (Borrelli, et al., 2017).

The identification algorithms used for this study are subspace method and Prediction Error Method (PEM). The subspace algorithm is a black-box parameter estimation method (Overschee & Moor, 1996). The PEM optimizes the parameters through minimizing the error between the measured output and predicted output (Ljung, 1999).

$$V_N(G,H) = \sum_{t=1}^{N} e^2(t)$$

$$e(t) = H^{-1}(q)[y(t) - G(q)u(t)]$$

$V_N(G,H)$ The cost function to be optimized
$G(q)$ The system transfer function
$H(q)$ Feedback transfer function
$y(t)$ System output at moment t
$u(t)$ System input at moment t

Both methods can estimate the model independently. In this study, both subspace method and PEM are employed, respectively, as the initial estimate in stage 1, and as improvement of initial estimate in stage 2 (Table 6).

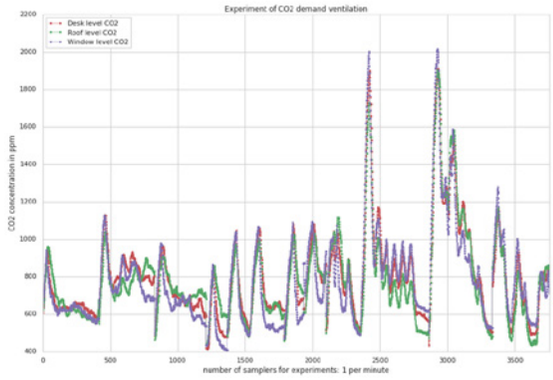

Figure 5: The total CO_2 records.

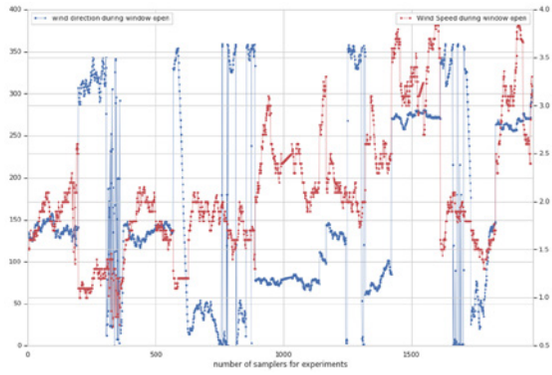

Figure 6: Wind speed and direction during periods of open window.

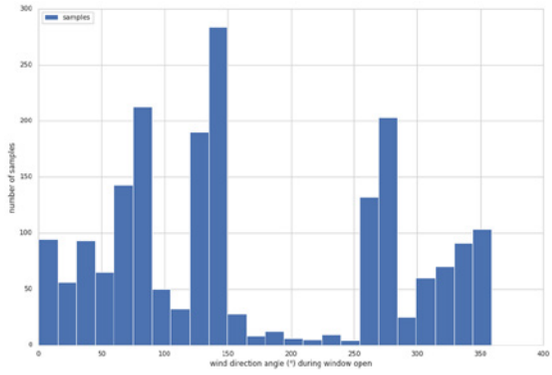

Figure 7: Histogram of the wind direction.

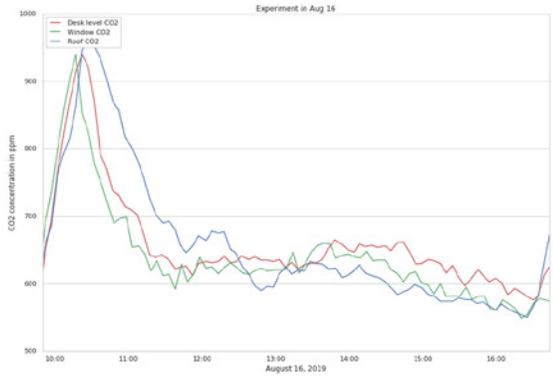

Figure 8: Experiment on August 16, 2019.

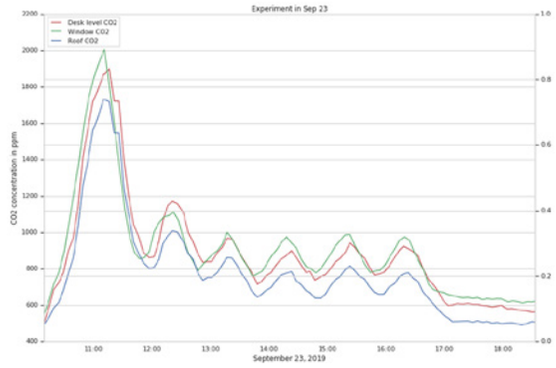

Figure 9: Experiment on September 23, 2019.

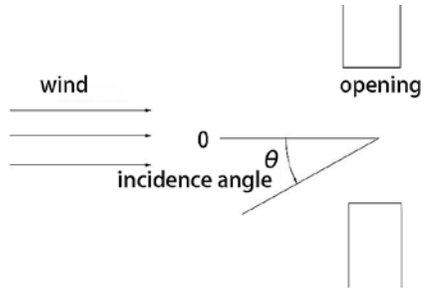

Figure 10: Schematic of the wind angle of incidence.

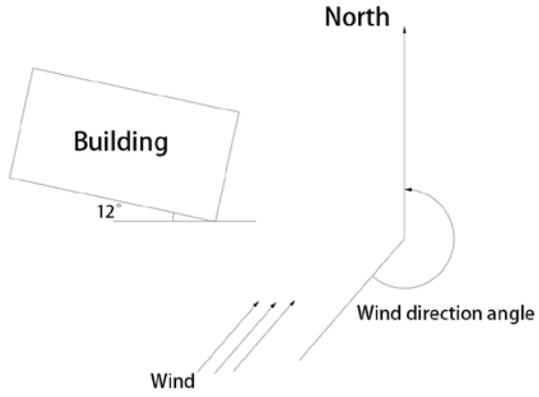

Figure 11: Schematic of the angle of the wind direction.

Results and Discussion

EXPERIMENT DATA

The experiments were realized over 11 days during August and September 2019. All the records of CO_2 concentration are concatenated as indicated in Figure 5.

The wind speed and wind direction information is indicated in Figure 6 and Figure 7. The wind speed data is in a reasonable range, but fewer wind direction data in the range of 150°–250° (wind direction angle Figure 11).

DATASET CLASSIFIED BY WIND SPEED AND INCIDENT ANGLE

The wind incidence angle is defined as 0° for wind directed toward the window opening and measured counterclockwise as in Figure 10. The building orientation angle is indicated in Figure 11.

The data is classified into three sets by wind speed (v), and wind incidence angle (θ):

— Low wind speed
— High wind speed windward
— High wind speed leeward

LOW WIND SPEED DATA (V ≤ 1.5 M/S)

Linear regression equation:

$$V_{nv} = \alpha\Delta T + \beta T_i + \gamma V$$

Polynomial regression equation:

$$V_{nv} = f(\Delta T, T_i, V, \Delta T^n, T_i^n, V^n)$$

HIGH WIND SPEED AND WINDWARD DATA (V > 1.5 M/S, AND θ ∈, (0°, 90°))

Linear regression equation:

$$V_{NV}^2 = \alpha\Delta T + \beta V\, cos(\theta) + \gamma V$$

Polynomial regression equation:

$$V_{NV}^2 = f(\Delta T, V\, cos(\theta), V, \Delta T^n, (V\, cos(\theta))^n, V^n)$$

HIGH WIND SPEED AND LEEWARD DATA (V > 1.5 M/S, AND θ ∈ (90°, 270°))

Linear regression equation:

$$V_{NV}^2 = \alpha\Delta T + \beta V\, cos(\theta)$$

Polynomial regression equation:

$$V_{NV}^2 = f(\Delta T, V\, cos(\theta), \Delta T^n, (V\, cos(\theta))^n)$$

THE REGRESSION RESULTS FOR THE THREE DATASETS

The regression results for the three datasets have been presented in Figures 12 to 15 and Tables 2 to 5.

In linear regression, the R^2 score is ranked as follows: leeward high-speed data > low-speed data > windward high-speed data. This order confirms previous study (Larsen & Heiselberg, 2008).

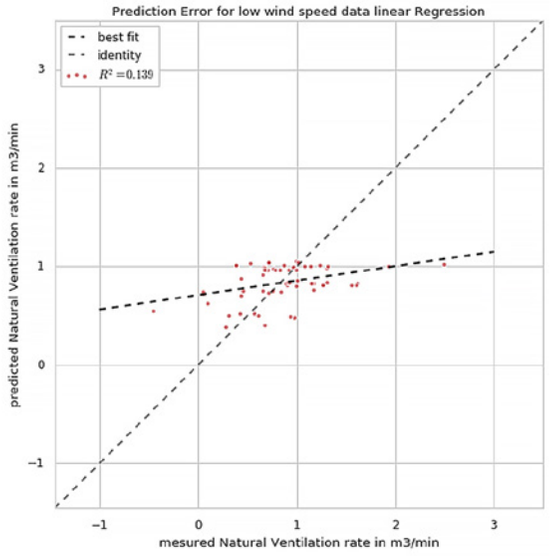

Figure 12: Linear regression of low wind speed data.

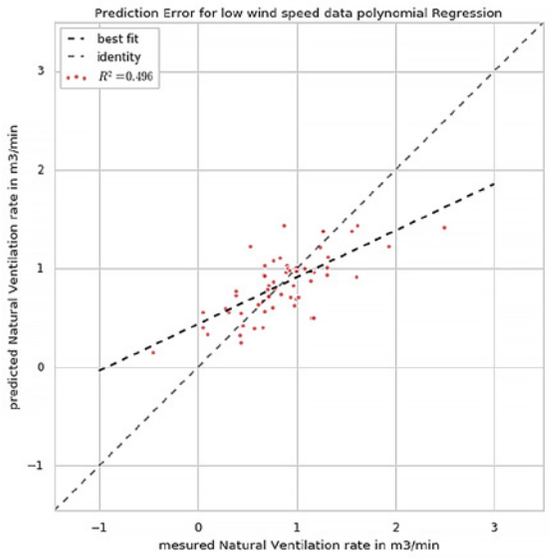

Figure 13: Low wind speed data polynomial regression.

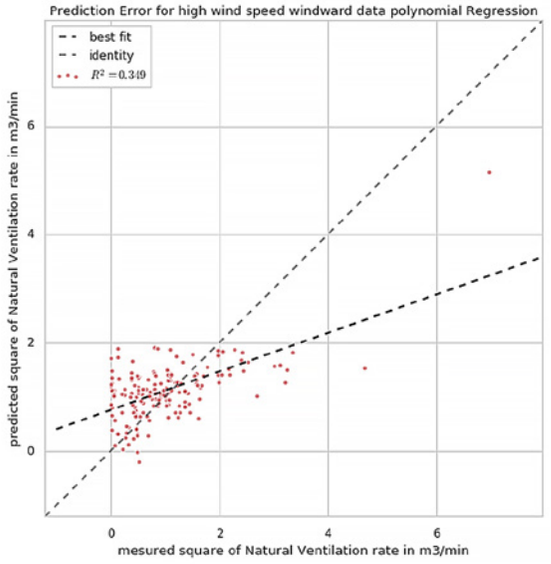

Figure 14: Linear regression of high wind speed and windward data.

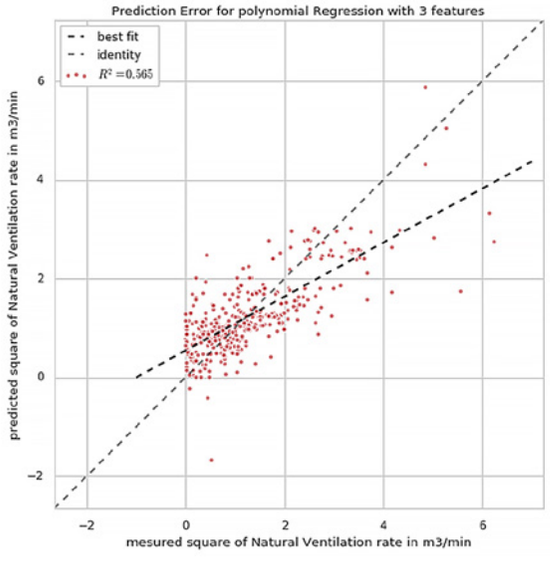

Figure 15: Polynomial regression for the three features.

	Number of coefficients	R^2 score for regression
Linear regression	3	R^2: 0.14
Polynomial regression	20 (3rd order)	R^2: 0.50

Table 2: Regression for low-speed data.

	Number of coefficients	R^2 score for regression
Linear regression	3	R^2: 0.09
Polynomial regression	35 (4th order)	R^2: 0.35

Table 3: Regression for high-speed windward data.

	Number of coefficients	R^2 score for regression
Linear regression	2	R^2: 0.50
Polynomial regression	4 (2nd order)	R^2: 0.50

Table 4: Regression for high-speed leeward data.

	Number of coefficients	R^2 score for regression
Linear regression	3	R^2: 0.2
Polynomial regression	126 (4th order)	R^2: 0.57

Table 5: Regression for three features.

The implementation of regression on the three features from the previous study (Larsen & Heiselberg, 2008): $^{\Delta T}/_{Ti}$, $^{\Delta T}/_{v^2}$, V^2, doesn't return a better output.

FEATURE SELECTION

Two outputs exist from previous analysis:

— Natural ventilation flow-rate expression
— Corresponding features for natural ventilation

Each lead to a different strategy:

— **MASS BALANCE METHOD:** Based on the flow-rate formula and mass conservation equation, a control-oriented linearized model is proposed.
— **DATA BALANCE METHOD:** Based on the selected features, the corresponding parameters are reorganized through data-driven algorithms to establish a control-oriented model.

The low R^2 score of linear regression reveals the difficulty to gain a reliable flow-rate expression in low order. The data balance method is therefore adopted in this study.

The selected features from regression analysis are: $V \cos\cos(\theta)$, V, T_i, $\sqrt{\Delta T}$, and ΔT.

SYSTEM IDENTIFICATION

Based on the selected features, the system dynamics of CO_2-based demand-controlled ventilation is proposed:

$$C_z^{k+1} = a \cdot C_z^k + \left(b \cdot \Delta T + c \cdot \sqrt{\Delta T} + d \cdot V\right) \cdot U_1 + e \cdot U_2$$

C_z^{k+1}, C_z^k	The CO_2 concentration at moment k and k+1 (ppm)
ΔT	Difference of room air temperature and outside air temperature (°C)
V	Scalar wind speed (m/s)
U_2	Window opening indicator (1: open; 0: closed)
U_1	Number of occupants in room

The coefficients (a,b,c,d) are to be fitted via a data-driven identification algorithm. e: in order to simplify the problem, it is assigned as 11 ppm, based on 0.005 L/s (adult CO_2 generation rate).

The expression is simplified when the window is open:

$$C_z^{k+1} = a \cdot C_z^k + b \cdot \Delta T + c \cdot \sqrt{\Delta T} + d \cdot V + e \cdot G,$$

in a linear time-invariant system form:

$$X_{k+1} = AX_k + BU,$$

where

A = a
B = $[b,c,d,e]$ (e is assigned to be 11 ppm)
U = $[\Delta T, \sqrt{\Delta T}, V, G]^T$

The original dataset isn't a single continuous set. Instead, it contains 11 independent datasets. The algorithms in machine learning can't be applied directly. Two different strategies can solve this problem:

— Data item independence
Algorithms for multi-experiments

DATA ITEM INDEPENDENCE

To implement regression algorithms, the data items should be independent. That means each data item contains all the necessary information. In this case, the information C_z^{k+1} and C_z^k connect two continuous moments k+1 and k, but the $\Delta C_z^k = C_z^{k+1} - C_z^k$ can be attached to the moment k.

Under this strategy, the regression problem is transformed as:

$$\Delta C_z^k = a' \cdot C_z^k + b \cdot \Delta T + c \cdot \sqrt{\Delta T} + d \cdot V + e \cdot G.$$

Then, we can randomize the data order and use cross-validation to select the optimized model:

$$\Delta C_z^k = -0.015 \cdot C_z^k - 2.37 \cdot \Delta T + 4.37 \cdot \sqrt{\Delta T} - 0.54 \cdot V + 11 \cdot G$$

Combining with:

$$C_z^{k+1} = C_z^k + \Delta C_z^k$$

the expression becomes:

$$C_z^{k+1} = 0.985 \cdot C_z^k - 2.37 \cdot \Delta T + 4.37 \cdot \sqrt{\Delta T} - 0.54 \cdot V + 11 \cdot G$$

Figure 16 shows the results of regression for system dynamics.

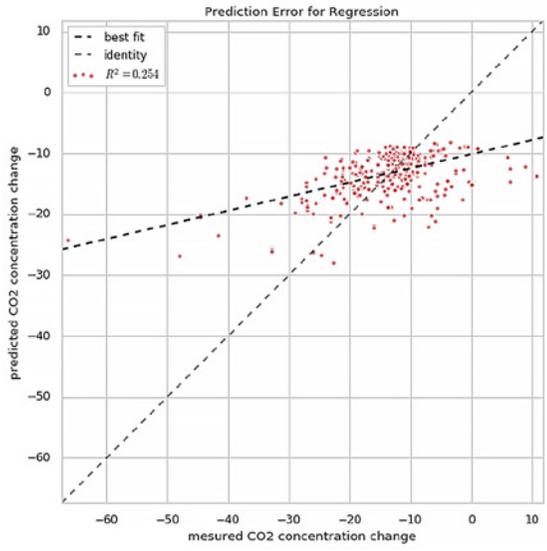

Figure 16: Regression for system dynamics.

ALGORITHMS FOR MULTI-EXPERIMENTS

The system identification algorithm for multiply datasets is an alternative method. The entire dataset is split into 32 experiments. Each starts with window opening and ends with window closing. The validation strategy adapted is the leave-one-out cross validation (Table 6 and Figure 17).

The system identification results for 5-step prediction (5 minutes) is:

$$x(t + Ts) = Ax(t) + Bu(t) + Ke(t)$$

Ts = Time interval of 60 s
E(t) = Gaussian noise error
A = 0.9852
B = [-5.115, 16.59, -1.823, 11]
K = 0.9007

The discrete system dynamics are based on all the datasets:

$$C_z^{k+1} = 0.985 \cdot C_z^k - 5.115 \cdot \Delta T + 16.59 \cdot \sqrt{\Delta T} - 1.823 \cdot V + 11 \cdot G$$

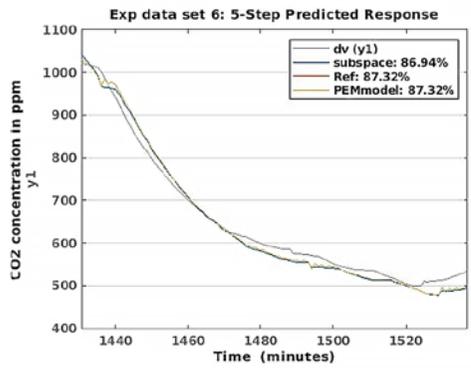

Figure 17: The validation on experiment number 6.

No° Dataset for validation	Accuracy stage1 Subspace	Accuracy stage 2 PEM	A	B	K
1	87.58%	88.32%	0.986	-4.942; 15.97; -1.921	0.890
3	72.10%	75.60%	0.985	-5.002; 15.75; -1.357	0.904
4	84.25%	85.83%	0.985	-5.396; 17.56; -2.068	0.897
5	91.44%	91.17%	0.985	-5.113; 16.73; -1.98	0.899
6	86.94%	87.32%	0.986	-5.406; 17.11; -1.946	0.904
7	87.35%	87.09%	0.986	-4.928; 16.18; -1.855	0.893
8	80.85%	82.18%	0.985	-5.194; 16.74; -1.81	0.907
9	85.50%	86.22%	0.985	-5.112; 16.58; -1.819	0.901
11	86.05%	86.42%	0.985	-5.165; 16.67; -1.796	0.893
12	87.37%	88.15%	0.985	-5.102; 16.75; -1.966	0.901
13	67.71%	67.96%	0.987	-4.293; 13.92; -1.493	0.901
14	62.44%	65.08%	0.985	-5.173; 16.74; -1.836	0.903
15	79.97%	80.23%	0.985	-5.146; 16.65; -1.825	0.901
16	64.06%	66.30%	0.985	-5.204; 16.82; -1.876	0.901
17	67.76%	68.83%	0.985	-5.152; 16.66; -1.825	0.900
18	88.36%	89.36%	0.985	-5.105; 16.52; -1.759	0.900
19	73.34%	74.89%	0.985	-5.13; 16.69; -1.837	0.902
22	75.60%	75.78%	0.985	-5.093; 16.51; -1.831	0.899
24	71.21%	75.36%	0.985	-5.104; 16.6; -1.811	0.901
25	48%	69.80%	0.985	-5.115; 16.61; -1.826	0.901
26	84.50%	87.96%	0.985	-5.36; 17.34; -1.899	0.900
27	52.40%	54.10%	0.985	-5.101; 16.54; -1.813	0.902
28	77.33%	80.19%	0.985	-5.219; 16.75; -1.594	0.908
29	64.80%	66.18%	0.985	-5.142; 16.65; -1.836	0.901
30	80.99%	80.90%	0.985	-5.125; 16.6; -1.854	0.902
32	75.51%	78.63%	0.985	-5.133; 16.66; -1.842	0.901

Table 6: Leave-one-out cross validation table.

Within the 32 experiments, the prediction of identified system dynamics:

— 26 with good accuracy
— 2 with poor performance
— 4 with failure

In No°31, the CO_2 is stagnant. In No°21 experiment, the wind direction changed dramatically. The detail is shown as follows. Table 7 shows the experiments of poor performance.

No°	Time	Wind direction	Wind speed	Delta T (°C)
2	8/16/2019 4:25:00 PM– 4:42:00 PM	140° – 146°	2.4 – 2.6 m/s	0.8 – 1.1
10	9/17/2019 10:40:00 AM– 10:57:00 AM	61° – 66°	2.6 – 2.7 m/s	2.3 – 2.8
20	9/25/2019 12:19:00 PM– 12:41:00 PM	310° – 340°	1.8 – 2 m/s	0.5 – 1.1
21	9/25/2019 12:51:00 PM– 1:03:00 PM	350° – 0° – 180°	1.9 – 2.1 m/s	1.2 – 1.3
23	9/25/2019 2:24:00 PM– 2:48:00 PM	330° – 180°	1.8 – 1.9 m/s	1.2 – 1.5
31	9/26/2019 4:27:00 PM– 4:56:00 PM	270° – 279°	2.4 – 2.8 m/s	2.2 – 2.6

Table 7: Experiments of poor performance.

The identified system dynamics from the two strategies shared a high similarity, as shown in Equation 19 and Equation 21. The second strategy in 3.5.2 provides much more meaningful intuition:

— CO_2 concentration decreased, following the identified system dynamics in 26 experiments.
— Very weak natural ventilation for 5 experiments
— Special case No°21: dramatic change of wind direction

The identified system dynamics has the strong potential to be implemented in a controller for CO_2-demand predictive control.

Conclusion

Based on machine-learning algorithms, this paper proposes a new data-driven approach to study controlled natural ventilation. This study doesn't follow the traditional way to estimate a general flow-rate formula. Instead, this study focuses directly on predictive control-oriented modeling. The final identified model distinguishes the useful natural ventilation scenario from the weak flow-rate scenario. The effort of a controlled natural ventilation will be meaningless for the second scenario. In pursuit of an integrated design in ventilation and window systems, future work will focus on the application of this data-driven approach on other forms of natural ventilation, and the implementation of identified dynamics in real-world control design.

References

Borrelli, F., Bemporad, A., & Morari, M. (2017). *Predictive Control for Linear and Hybrid Systems.* Cambridge University Press.

Emmerich, S. J., & Persily, A. K. (2001). *State-of-the-Art Review of CO_2 Demand Controlled Ventilation Technology and Application.* Building and Fire Research Laboratory. Gaithersburg: National Institute of Standards and Technology.

Emmerich, S. J., Dols, W. S., & Axley, J. W. (2001). *Natural Ventilation Review and Plan for Design and Analysis Tools.* Gaithersburg: National Institute of Standards and Technology.

Erdmann, C., Steiner, K., & Apte, M. (2002). Indoor Carbon Dioxide Concentrations and Sick Building Syndrome Symptoms in the BASE Study Revisited: Analyses of the 100 Building Dataset. *Proceedings of Indoor Air 2002, Vol. III,* 443–448.

Erhart, T., Guerlich, D., Schulze, T., & Eicker, U. (2015). Experimental validation of basic natural ventilation air flow calculations for different flow path and window configurations. *Energy Procedia,* 2838–2843.

Kubba, S. (2016). 7.3.7 Ventilation Systems. In *LEED v4 Practices, Certification, and Accreditation Handbook* (2nd Ed.). Elsevier.

Larsen, T. S., & Heiselberg, P. (2008). Single-sided natural ventilation driven by wind pressure and temperature difference. *Energy and Buildings, 40*(6), 1031–1040.

Ljung, L. (1999). Chapter 7. *System Identification: Theory for the User* (2nd Ed.). Prentice Hall PTR.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (5th Ed.). John Wiley & Sons.

Overschee, P. v., & Moor, B. L. (1996). *Subspace identification for linear systems.* Kluwer Academic Publishers.

Tang, Y., Li, X., Zhu, W., & Cheng, P. L. (2016). Predicting Single-sided Airflow Rates Based on Primary School Experimental Study. *Building and Environment, 98,* 71–79.