Predicting Air Infiltration and Window State in Residential Dorms using Deep Neural Networks Coupled with EnergyPlus

Pratik Pandey1, Nina Sharifi2, Bing Dong1
1Built Environment Science & Technology (BEST) Lab, Department of Mechanical and Aerospace Engineering, Syracuse University, Syracuse, New York, USA
2Applied Sciences and Technology Research in Architecture (ASTRA) Lab, Syracuse University School of Architecture, Syracuse, New York, USA

Abstract
Buildings use about one-third of all the energy produced. Due to inaccurate modeling from building simulation engines, the total annual energy requirement of the building is mostly underestimated. This discrepancy in energy prediction affects the preliminary HVAC sizing, increasing the overall building energy consumption. We measured the energy usage of sixteen residential dorms over a year. We measured energy consumed by every building element like lighting, HVAC, plug-load, refrigerator, stove, exhaust hood, and water heater and applied these as a fraction schedule inside EnergyPlus. Using contact sensors, we measured the occupant behavior of using 48 windows and 64. We developed a window opening model for estimating the window’s state using three layered Deep Neural Network (DNN). We constructed the mass balance model to estimate the Air Exchanges per Hour (ACH) using the CO2 decay data during window opening. Using these ACH data, we trained another DNN model that could predict the window operation ACH of individual dorms. We trained another DNN model to predict dynamic infiltration ACH. All trained DNN models were embedded into EnergyPlus using the newly released EnergyPlus Application Programming Interface (API). Finally, the actual HVAC energy consumption was compared with the traditional and DNN-embedded EnergyPlus simulations.

Highlights
- Replacing static air exchange rates (ACH) with dynamic results using EnergyPlus Python API
- Modelling infiltration and window operation ACH while using DNN.
- Using Window state prediction DNN model to determine the window operation timeframe.
- Coupling window state, infiltration ACH and window operation ACH DNN models with EnergyPlus Model using appropriate control logic.

Introduction
Buildings consume about 40% of the total energy generated (Becerik-Gerber et al., 2014). Proper accounting of the energy usage profile is critical in determining the appropriate HVAC system sizing. If HVAC is undersized, the system must work overtime to maintain the demand, and hence there will be a significant energy loss. Previous studies have reiterated that there is a significant difference between the simulation results from Building Energy Simulation (BES) engines and actual energy usage. Some studies have found that the simulated energy usage is off by almost 50% (Muroni et al., 2019). There is a unanimous consensus among building scientists that this discrepancy results from not accounting for Occupant Behavior (OB).

Currently, there is a wide-ranging practice of using constant schedules for accounting building systems usage by the occupant. These generalized schedules are based on the survey from ASHRAE, which completely disregards the stochastic nature of OB (ASHRAE 90.1-2022 (I-P)). In another rule-based approach, the window operation schedules might be based on a probability curve with the general rule of assigning the window state as ‘OPEN’ as the outdoor temperature increases (Li et al., 2015). This assignment is based on the observation that the probability of window opening increases as outdoor temperature increases. However, assigning a single cutoff temperature for all occupants in a building cannot be proper validation. To rectify these issues, various studies have been conducted to model the occupants’ window-opening behavior using machine learning methods. Machine learning models can decipher the complex nonlinear relationship between various variables and train the model to minimize loss. If the Machine Learning (ML) model is trained correctly, it can become highly scalable. The ML models can be trained with additional data in the future using the concept of transfer learning. These attributes make ML models more applicable for the prediction of building elements. Due to lower sensor costs and an abundance of data, ML methods are becoming more popular. Due to the popularity of ML modelling, ASHRAE has just released the ‘ASHRAE Global Occupant Database,’ which contains a building system using data from all corners of the world (Dong et al., 2022). The data from this database can be used to make any ML model for different types of buildings across the globe.

Some studies have already used ML methods to predict building system usage. Using Deep Neural Network
(DNN), Romana et al. (2017) was able to predict the state of the windows in office buildings (Markovic et al., 2017). The author trained the data using five hidden layers and 25 independent variables. The model was trained in the data from the buildings in Germany and tested in the office building in Philadelphia, PA, USA. The accuracy of this model was 86%, and the True Positive Rate (TPR) was 0.37. Another study tried to predict the window behavior modeling approaches during the transition season in China. The study obtained the highest accuracy of 75% using two layered neural networks with 25 nodes. A similar study has been conducted in the United Kingdom and has obtained good results (Han et al., 2012). A recent study tried to predict the number of occupants in the indoor space using various machine learning strategies. The occupant number count was used to control the outdoor air flow rate so that the risk of COVID-19 infection could be curtailed (Jiang et al., 2023). Another study tried to predict the thermal comfort of the indoor space using artificial neural networks with occupant behaviors and thermal sensations as inputs (Deng & Chen, 2018).

Air Changes per Hour, commonly called ACH, is the number of times the air inside a zone is completely changed in one hour. The stale air in an indoor space should be changed to 2.5 L/s/person per ASHRAE guidelines for residential ventilation (“ANSI/ASHRAE Addendum p to ANSI/ASHRAE Standard 62.1-2013,” n.d.). Three mediums can achieve this: Mechanical Ventilation, Natural Ventilation, or Infiltration. Infiltration refers to the air leakage through the cracks and orifice near windows and doors due to indoor-outdoor pressure differences. Generally, the pressure is higher indoors as mechanical ventilation blasts the air in the ventilated zone. Thus, the air from inside tends to be pushed outside, resulting in the loss of conditioned air. Traditionally, the infiltration ACH is manually input inside the EnergyPlus. This paper aims to dynamically model the change in ACH of the zone by embedding a trained DNN model inside the EnergyPlus Python API. The DNN model will read the indoor and outdoor thermal and wind variables and decide the optimal ACH for infiltration. Also, another DNN model will find the optimal ACH during window operation. The window state is also determined by the DNN, which decides whether to use infiltration ACH prediction DNN or window ACH prediction DNN. Finally, the energy consumption during the heating season is calculated using EnergyPlus and compared with the results from the baseline conditions. The detailed methodology and results are presented next.

Methodology

Overall Approach

First, we collect CO₂ and window operation data from the residential dorms. Then we will determine the CO₂ decay timeframes with windows closed and the CO₂ decay timeframes with windows open. Then we will calculate the infiltration and window operation ACH using the mass balance equation. The infiltration ACH thus obtained will be compared using the ACH values obtained from the blower door test. The ACH₅₀ values obtained from the blower door test should be converted to the ACH values using appropriate conversion factors before the comparison, as ACH₅₀ is obtained while artificially maintaining the 50 Pa indoor-outdoor pressure difference. We used conversion coefficients derived from wind and indoor/outdoor temperature along with fan pressurization index approximation. A final dataset is selected if ACH obtained from mass balance is equivalent to the ACH obtained from the blower door test.

The dataset is used to train and test the DNN models to predict the infiltration ACH and window operation ACH. The window state prediction model is also created using DNN. These models can now independently decide whether the occupant will open the window. If the DNN predicts that the occupant opens the window, then the DNN model for predicting the window operation ACH will be used. If the DNN predicts that the occupant will not open the window, the DNN model for predicting the infiltration ACH will be used. These models are integrated into the EnergyPlus, where the DNNs will decide the results based on the values input by the engine. The schematics of the overall methodology are shown in Figure 1.

Data Collection

For this project, we have used data from 16 residential dorms inside Syracuse University, Syracuse, NY. We monitor three broad categories of data: IAQ, window/door operation, and power usage. The location of the sensors in one dorm, along with their power source, is

Figure 1: Overall Methodology of ACH and window state prediction models
shown in Figure 2. Every dorm is divided into five zones: North Facing Bedroom, South Facing Bedroom, Living Room, Hallway, and Bathroom.

![Figure 2: Location of Sensors inside a dorm](image)

### Table 1: Sensor Information

<table>
<thead>
<tr>
<th>Measure</th>
<th>Sensor Model</th>
<th>Accuracy</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window/ Door Status</td>
<td>Netvox (R311A)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Temperature</td>
<td>Milesight (AM107)</td>
<td>0°C to +70°C (-/+ 0.3°C), -20°C to 0°C (-/+ 0.6°C)</td>
<td>0.1°C</td>
</tr>
<tr>
<td>Humidity</td>
<td>Milesight (AM107)</td>
<td>10% to 90% RH (+/- 3%), below 10% and above 90% RH (+/- 5%)</td>
<td>0.5% RH</td>
</tr>
<tr>
<td>Light</td>
<td>Milesight (AM107)</td>
<td>±30%</td>
<td>1 lux</td>
</tr>
<tr>
<td>CO₂</td>
<td>Milesight (AM107)</td>
<td>±30 ppm or ±3 % of reading</td>
<td>1 ppm</td>
</tr>
<tr>
<td>TVOC</td>
<td>Milesight (AM107)</td>
<td>±15%</td>
<td>1 ppb</td>
</tr>
<tr>
<td>Power Usage</td>
<td>ONSET (EG1430 Pro)</td>
<td>NA</td>
<td>1 sec</td>
</tr>
</tbody>
</table>

The IAQ sensor is located in the living room (adjacent to the kitchen) because the significant sources of indoor pollution, CO₂ and TVOC, are generated during cooking events. This IAQ sensor can also measure indoor temperature and relative humidity. The outdoor temperature, RH, solar radiation, and wind speed are measured using the outdoor weather monitoring system installed at the top of one of the university’s buildings. The detailed list of sensors used for this project is presented in Table 1.

where,

\[ C = \text{Instantaneous concentration at indoor space (ppm)} \]
\[ C_{\text{in}} = \text{Concentration of CO}_2 \text{ coming inside (ppm)} \]
\[ C_{\text{out}} = \text{Concentration of CO}_2 \text{ going out (ppm)} \]
\[ S = \text{Source of CO}_2 \text{ inside (ppm. m}^3/\text{hr)} \]
\[ V = \text{Volume of indoor space (m}^3) \]

### Determining the infiltration rates using the mass balance equation

First, data is required to train the infiltration in the DNN model. For that, the infiltration rates are calculated using the mass-balance model shown by Equation 1.

\[
V \frac{dC}{dt} = Q_{\text{in}}C_{\text{in}} - Q_{\text{out}}C_{\text{out}} + S
\]  (1)

where,

\[ Q = Q_{\text{in}} = Q_{\text{out}} = \text{Airflow rate in (m}^3/\text{hr)} \]

The Ordinary Differential Equation (ODE) 1 is the first-order linear differential equation that can be easily solved by hand using an integrating factor. The limitation of this model is that it does not account for the source of CO₂ inside the chamber during the decay phase. The degradation factor of CO₂ is zero, as it is a very stable compound. Also, we have assumed that the rate of air inflow and air outflow is constant.

Many studies have used CO₂ as the tracer gas as it is non-reactive. A cooking/smoking/gathering event will significantly elevate the CO₂ concentration in the indoor space. After the source of CO₂ generation ceases, the indoor CO₂ concentration starts to plummet in the indoor space and reaches the baseline concentration. This baseline is slightly elevated as compared to the outdoor space. In this instance, we check if any windows or door
has been opened. During this decay phase, if the windows and doors are continuously closed, we can use this instance of decay to find the infiltration rates.

Using these infiltration rates without validating the results from the mass balance equation is not plausible. Hence, we use the results from the blower door test to validate the results from this ODE. During the blower door test, an artificial pressure difference of 50 pascals is applied in the indoor space to identify the leakages. This ACH is commonly referred to as ACH_{50}. Several correlations have been developed for converting ACH_{50} to ACH. For example, we can obtain ACH by dividing ACH_{50} by a Conversion Coefficient (CC). CC is calculated using fan pressurization data, wind, and stack variables as shown by Equations 2 – 4 (Ji et al., 2022).

\[
ACH = \frac{ACH_{50}}{CC} \tag{2}
\]

\[
CC = \left(\frac{1}{s}\left(\frac{8\pi^2}{\rho} \left(\frac{50}{4}\right)^n \right) s = (f_v^2 v^2 + f_T^2 |\Delta T|)^n \tag{3}
\]

where,

\[n = 2/3 \text{ (commonly ranges between 2/3 to 1 depending on fan pressurization data)}\]

\[\rho = \text{density of air}\]

\[v = \text{Wind Speed}\]

\[f_v / f_T = \text{structure coefficients obtained from literatures/ measured in (m/s).K}^{1/2}\]

After this conversion, the real data from the blower door test is compared with the results from the mass balance equation. The DNN model can be trained using this datasheet if the comparison is valid.

The accelerated CO\textsubscript{2} decay can be observed during the window opening phase. The ACH during window operation can be stochastic as it depends on wind and stack effects. The ACH during window operation also depends on the surface area of the window opened, which is not measured. The lack of availability of this data is the biggest impediment to creating a good DNN model for predicting the ACH during window operation. The ACH data table used for training and testing the model is also created using the mass balance model. However, this dataset has much fewer data points (102 decay instances during window operation) than in the infiltration ACH dataset (2300 decay instances during infiltration). Due to the lack of data points and the non-availability of the window surface area opened, the RMSE of the DNN model for predicting the window opening ACH is higher than the infiltration ACH prediction model, as described in the results section.

The detailed methodology for creating and integrating the residential dorms window state prediction DNN model into EnergyPlus has already been published (Pandey et al., 2022). We use Deep Neural Network (DNN) for making the window state prediction model. After collecting the data from all the sensors located at the dormitory, they are compiled into a single data file. The data is cleaned and split randomly into training and testing sets. 80% of the data file is used for training, and 20% is used for testing. We used three layered DNN with Adam optimizer, and the learning rate is set at 3e-4. The inner activation function is ‘ReLU,’ and the outer activation function is ‘Sigmoid’; the final prediction is the window state with binary outputs. The trained model is then dumped into an ‘ANNSequential’ object. This object now contains all the optimized weights and biases. The model is then used to test the data from the testing set. If the testing set results are satisfactory, the model can be dumped into a ‘pickle’ object, which can then be embedded into the EnergyPlus using the newly released python API.

The second model is trained for predicting the Air Exchange Rates due to infiltration. The infiltration rates derived from real data using a mass balance equation are used for training the DNN. The procedure for training and testing is similar to that of the window state prediction model. However, the model has different hyperparameters and loss functions. Similarly, the third model for predicting the ACH during window operation has the same structure but different hyperparameters.

Table 2 details different inputs and hyperparameters used for training and testing these three models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Input Variables</th>
<th>DNN Architecture</th>
<th>Optimizer</th>
<th>Learning Rate</th>
<th>L2 Reg</th>
<th>Loss Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window State ACH</td>
<td>Hour of Day, Day of week, Outdoor/Indoor Temperature, Dew Point Temperature, Outdoor/Indoor RH</td>
<td>7-36-36-36-1</td>
<td>Adam (Inner)/ Sigmoid (Outer)</td>
<td>3e-4</td>
<td>None</td>
<td>BCEWithLogitsLoss()</td>
</tr>
<tr>
<td>Infiltration ACH</td>
<td>Outdoor/Indoor RH, Home ID, Wind Speed and Indoor-Outdoor Temperature Difference</td>
<td>5-32-32-32-32-1</td>
<td>Adam</td>
<td>3e-4</td>
<td>0.0006</td>
<td>MSELoss()</td>
</tr>
<tr>
<td>Window Operation ACH</td>
<td>Outdoor/Indoor RH, Home ID, Wind Speed and Wind Direction</td>
<td>5-64-64-64-64-1</td>
<td>Adam</td>
<td>3e-4</td>
<td>0.00032</td>
<td>MSELoss()</td>
</tr>
</tbody>
</table>
After training all three DNN models, they can be integrated into EnergyPlus using Python API. All independent variables used for training the models can be retrieved using EnergyPlus variable handle. Some of the variables may be derived from Schedule:File Object. The window state prediction DNN is used first to determine if windows are opened or closed. If windows are closed, infiltration prediction DNN is used. If windows are opened, window ACH prediction model is used. The detailed process flow diagram is shown in Figure 3.

The ACH value is in agreement with the blower door test. Hence, the data table consisting of ACH values of all decay events can be used for training and testing the DNN model.

The window operation ACH has higher RMSE as the degree of the window opening is different at every instance of opening. This information cannot be recorded and, thus, cannot be used to train the model. The window operation model’s accuracy and True Positive Rate (TPR) are 94.34% and 0.78, respectively. The blue-shaded region in Figure 4 shows the instances of window operation in a year at this dorm. The plot looks satisfactory as the model predicts that window operation is nonexistent during winter and maximum during transition seasons. It should be noted that this prediction was made inside EnergyPlus using Python API.

EnergyPlus has a dedicated actuator handle to change volumetric flow rates expressed in m$^3$/s. Using the flow logic shown in Figure 3, the three DNNs are embedded inside the EnergyPlus, and the infiltration and window operation ACH is predicted. Since the actuator handles changing ‘per hour’ infiltration values are not available in EnergyPlus; the ACH values were converted to volumetric air flow rate. Figure 4 shows that infiltration air flow rates are lower than window operation air flow rates. This is consistent with findings from various studies before. Also, the indoor-outdoor temperature difference is maximum during the coldest month of February. Hence, the DNN correctly predicts maximum infiltration flow rates during this month. The model correctly predicts window operation flow rates that are four times higher than the infiltration flow rate on average. This can be observed by looking at flow rates in the transition season. During transition season, the window operation peaks as the indoor space lacks a dedicated cooling system.

The purpose of making neural networks is to accurately model the dynamic changes in ACH values. EnergyPlus

![Figure 3: Flow of DNN Predictions](image)
will take feedback from the model to simulate other variables. As stated in the introduction section, the energy use data from the simulation model vary significantly from actual energy use data. Since we have HVAC energy use data from power meters, we can compare it with traditional EnergyPlus simulations (without DNN) and simulations from EnergyPlus with DNNs embedded in them. For this comparison, the transition months of March, and April data were chosen. The daily energy use from the traditional simulation engine, DNN embedded simulation engine, and actual energy data were compiled in a single data frame for further analysis. It was found that the RMSE and MAPE of actual energy use and energy simulated from a traditional engine was 16.70 kWh and 47.83% respectively.

On the contrary, the RMSE and MAPE of actual energy use and energy simulated from the DNN engine was 9.39 kWh and 20.23%. Hence, the RMSE was reduced by 7.30 kWh and MAPE is reduced by 27.6%. Figure 5 compares daily energy use calculated from these methods.

![Predicted Volumetric Air Flow Rate due to Infiltration and Window Opening](image)

*Figure 4: Annual Infiltration ACH, Window Operation ACH and Window State predicted by DNN within EnergyPlus*
Discussions

Based on the literature review, this method of using machine learning methods for the prediction of ACH has not been done before. The lack of affordable sensors and dedicated communication systems for retrieving and managing significant volumes of data were the biggest impediments in the past that did not allow data collection on this scale. Furthermore, we could establish the validity of loss rates derived from the mass balance equation by comparing it with the blower door test. Using data from August 2021 to May 2022, we found thousands of CO₂ decay instances through infiltration. Because of the volume of this data, the RSME error for the DNN model was lower than the RMSE error for the window operation ACH data.

The cracks and orifices near the windows of every dorm are different. Hence, we included the home id column in the dataset used for training the DNN model. This helped to decrease the RMSE error for the infiltration model. Also, using the home id column helped encapsulate each dorm’s OB when calculating the ACH during window usage. While calculating the window operation ACH for all decay instances in every home, we observed that the ACH is lowest at the end homes compared to the window operation ACH of mid-homes. One of the reasons for this type of behavior is that the end homes have greater surface area exposed to the outdoor surface. Hence, indoor and outdoor temperature convergence will be faster in these homes. Thus, opening a lesser surface area of windows at the end homes might accomplish the same effect as opening the greater surface area of windows at the mid-homes. However, the RMSE error of the window operation ACH is very high. Even though we can distinguish the window operation behavior at the end and mid-homes, the lack of data related to the total surface area opened is the biggest impediment to this experiment, which needs to be addressed by future researchers.

The RMSE error between the energy simulations from DNN-embedded EnergyPlus and actual energy use is around 43% lesser than the RMSE error between traditional simulation and actual energy use. This is undoubtedly a substantial leap towards better calibration of simulation models. Improving the performance of DNN models may help to reduce this RMSE further down and may contribute to reducing the historical gap between actual and simulated energy use. In the future, researchers and energy auditors can use this data-driven methodology of computing dynamic changes in air exchange rates using DNN by properly accounting for window opening OB and outdoor weather fluctuations. Researchers can use other sophisticated ML methods to model ACH and get better results.
Conclusion

The average ACH values obtained using the mass balance equation are in agreement with the ACH values obtained from the blower door test. The Conversion Coefficient (CC) equal to 9.92 was calculated using wind and thermal variables to convert $A_{CH50}$ from the blower door test to regular ACH values. Dynamic values predicted by DNNs successfully changed the static air infiltration and window operation ACH values. These DNNs were embedded into EnergyPlus using Python API. The window state prediction model has an accuracy of 94% and a True Positive Rate (TPR) of 0.78. The infiltration ACH and window operation ACH prediction DNN models have testing accuracy of 0.17 $h^{-1}$ and 1.03 $h^{-1}$, respectively. The latter has higher RMSE as the total surface area of the window opening is different at every instance of window operation. Moreover, this variable is not available as it cannot be measured. This is a significant constraint for this study. The models were used to simulate the HVAC energy use during the winter months from January 2022 to April 2022. The actual HVAC energy use obtained from the power meter installed in the residential dorms was used to compare the simulation results. The DNN-embedded EnergyPlus model has an average daily RMSE of 9.39 kWh with actual energy use. Also, the traditional simulation engine, which uses static ACH values, has an average daily RMSE of 16.70 kWh. Furthermore, DNN simulations reduced the MAPE by 27.60%.

Acknowledgement

We would like to thank New York State Energy Research and Development Authority (NYSERDA) for funding this project.

References

ASHRAE 90.1-2022 (I-P)
Shanghai, China, Sept. 4-6, 2023
https://doi.org/10.26868/25222708.2023.1401