

Notes paper - A data-driven and scalable approach for window operation detection in multi-family residential buildings

Juliet Nwagwu Ume-Ezeoke, Kopal Nihar, Catehrine Gorle, Rishee Jain

June 13, 2023

1 Abstract

Natural cooling, utilizing non-mechanical cooling, presents a low-carbon and low-cost way to provide thermal comfort in residential buildings. However, naturally cooled residential buildings are difficult to design without a clear understanding of the effect opening and closing windows will have on occupant thermal comfort. Predicting when and understanding why occupants open windows is a difficult task in itself. Although work has been done on creating predictive window opening and closing models, they traditionally rely on large datasets, specialized sensing, and/or require training data for each building. This makes them ineffective tools for scaling natural cooling to a large number of buildings. Here, we propose a novel unsupervised window opening and closing detection method that utilizes easy-to-deploy off-the-shelf temperature and humidity sensors. We assess our proposed models' effectiveness on an empirical dataset. We also compare the method's performance with a support vector machine (SVM) model that represents the state-of-the-art in window detection. We find that the proposed method outperformed the SVM on a number of key indicators, but had subpar performance on data in which the indoor and outdoor temperature had small differences. While the SVM's performance was sensitive to the characteristics of the time series features, our proposed method, which relies on indoor temperature alone, demonstrated robust performance in our pilot studies, making it a potential candidate for developing a highly scalable and generalizable model for window operation detection. Overall, this work aims to demonstrate the viability of using unsupervised data-driven methods for understanding window operations in residential buildings. In doing so, we aim to enable more accurate modeling of naturally cooled buildings and further catalyze widespread adoption of this low-cost and low-carbon technology.

2 Introduction

3 Literature Review

4 Methods

TODO brief overview here...

4.1 Data Collection

We collected data for short periods of time over three months in the summer. We placed HOBO sensors ***(sensor name)** temperature/ relative humidity sensors in two adjacent rooms in a multi-family residential building. The opening and closing schedules were not determined beforehand but were allowed to proceed naturally. This created three distinct scenarios for testing the efficacy of the model we developed.

Descriptions of the data are provided in **?@tbl-data-collected**. The experiments are labeled as experiments A, B, and C, corresponding to data recorded at the beginning on July 20, July 27, and September 8 respectively. Experiment B was much longer than the other experiments, about 14 days vs experiments A and C were collected over 3 and 4 days each. Experiment B also had a much shorter period of the window being open, with the window open around 5 percent of the time, vs experiments A and C which had the window open for at least 20 percent of the time. ¹ ²

	A	B	C
Starting Day	2022-07-20 07:15:00	2022-07-27 09:00:00	2022-09-08 08:00:00
Data Length	4 days 00:00:00	14 days 10:45:00	3 days 00:00:00
Room	1.0	0.0	0.0
Opening Percentage	0.412987	0.051873	0.221453
Average Open Time	0 days 09:24:59.999999999	3 days 10:15:00	0 days 18:00:00

Figure 1: Data Collected

To develop an understanding of how temperature and relative humidity, the quantities that were measured, changed across the different experiments, we examined the distributions of the data in Figure 2. We observe that experiment B had the widest distribution of internal temperatures. Experiment C had internal temperatures that were higher and closer to the

¹TODO not sure if really need to discuss the average opening time...

²TODO: check this => i think opening percentage for b should be much higher?? Or maybe include plots of the real data

mean/median** of the external temperatures. This was due to observed higher temperatures during this time period, which makes for a particularly interesting edge case. The internal relative humidity was higher than the ambient relative humidity across experiments.

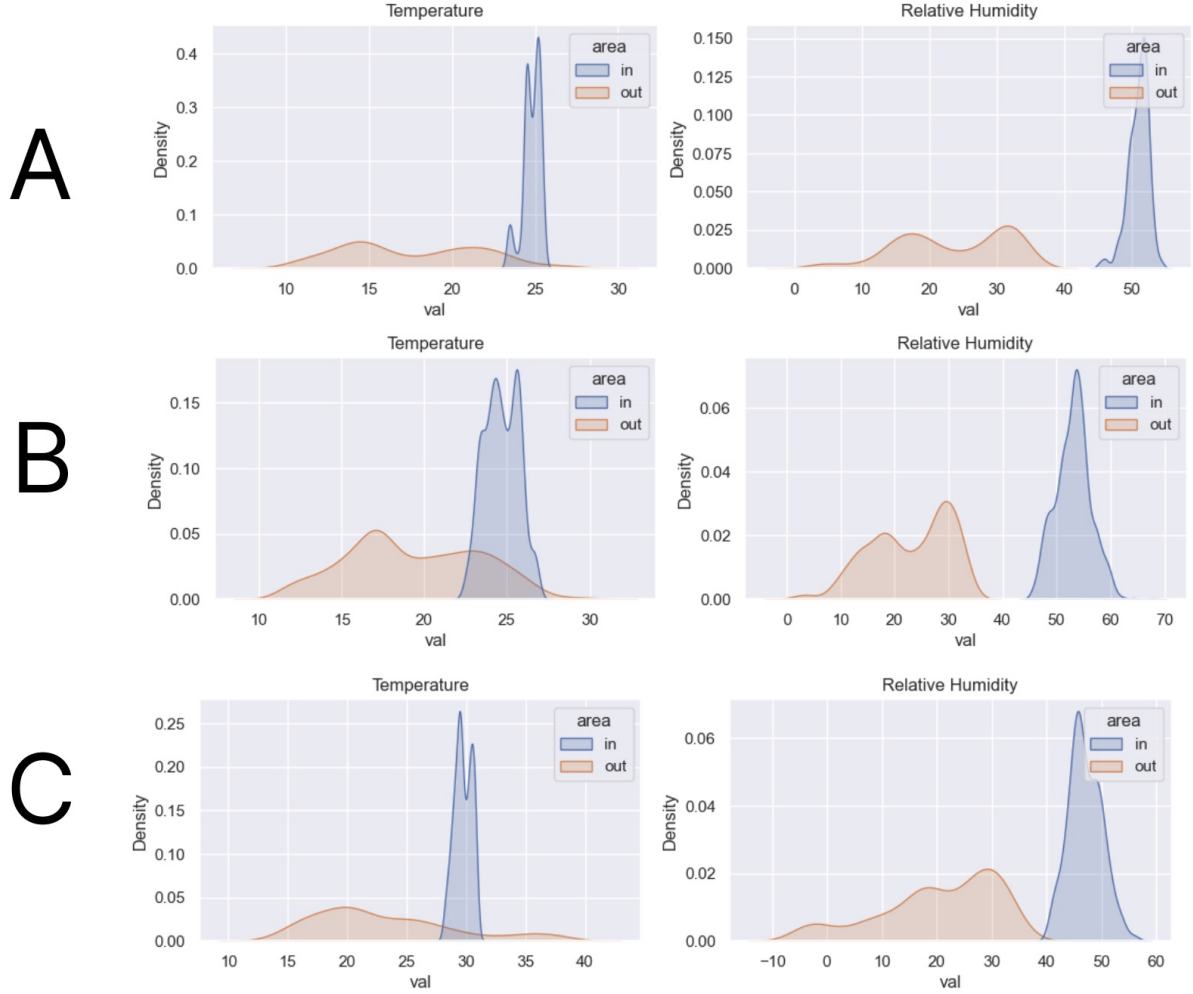


Figure 2: Distributions of Temperature and Relative Humidity Across Experiments

4.2 Detection Methods

some intro here ³

³TODO: include this part: We anticipate that creating a physical model to more explicitly determine the response of the indoor temperature based on the ambient temperature and the building makeup could be an interesting approach to capture the “truth”, or at least approximate it. Also need some discussion about the emphasis identifying where the window state changes...

4.2.1 Smoothing Technique

Our approach relies on an intuitive understanding of window operation detection. We expect a typical time series recording of a quantity of interest, such as temperature or relative humidity, to contain information that reflects the seasonality of ambient quantities, noise due to occurrences within the space where measurements are being taken, and the desired signal of due to a change in the window state. The way in which these three components of a measurement are combined is unknown, and, the noise component in particular cannot be known based on the measurements we have collected. Therefore, our task is to attempt to extricate the seasonality that is due to ambient processes from the measured data, which will ideally reveal the window detection and a bit of noise.

To remove the seasonality from the measured data, we examined three methods of de-seasonalizing: using seasonal-trend-decomposition, using a sinusoidal fit, and using an exponentially weighted mean function. These all reflect a trade-off between capturing pure seasonality and capturing a bit of noise. The sinusoidal method, which is optimized for the indoor temperature time series, purely reflects seasonality whereas the exponentially weighted mean function is simply a moving average of the indoor temperature signal. Therefore, some noise is captured as well. The seasonal component of a seasonal trend decomposition is somewhere in the middle. We found that the exponentially weighted mean function performed the best in isolating window detection, which follows from our intuition that both the seasonal components and noise need to be isolated. ⁴

The smoothing technique proceeds as follows. The goal is to identify $W(t)$, window state as a function of time. This can take on two values: 0, representing window closed, or 1, representing window open. We have an observed variable $T(t)$, which represents the measurement of the indoor temperature. We apply an exponentially weighted mean function to the observed temperature, creating a smoothed time series, $\bar{T}(t)$, which in theory removes strong peaks that would reflect changes in window state, and isolates information concerning the seasonal response and additional noise ⁵. Subtracting $\bar{T}(t)$ from $T(t)$ yields $T'(t)$, which is a time series that reflects changes in the window state and some additional noise. In order to more confidently identify the where the changes in window state occur, we examine the first and second derivatives of $T'(t)$, $\frac{dT'(t)}{dt}$ and $\frac{d^2T'(t)}{dt^2}$. Observing the second derivative in particular, is an effective way to identify change points. In order to predict where window changes occur, we apply a hypothesis test ⁶. We assume that the time series $\frac{d^2T'(t)}{dt^2}$ is normally distributed. Therefore, any value in $\frac{d^2T'(t)}{dt^2}$ that is more than 2 standard deviations away from the mean is unlikely to occur, and could possibly indicate an instance of a change in window state.

⁴TODO: i think the previous 2 paragraphs can be condensed or removed entirely

⁵this assumes that the ewm function is perfectly able to capture seasonality, which in reality it isn't... => need to write down assumptions => that the instantaneous change in indoor temp due to window opening is greater than any other potential source of instantaneous temperature change.

⁶is hypothesis test the right word?, should i use vocabulary like the null hypothesis and the alternative hypothesis??

We will call these initial guesses $G(t)$. They take on positive or negative values depending on whether they are predicting a transition from window open to close, or window close to open. Therefore, we round the values of $G(t)$ to 0 or 1 to reflect this. Finally, we interpolate between the rounded values of $G(t)$, so that we only predict a change in window state when $G(t)$ transitions between 0 and 1. This prediction of the window state is called $I(t)$. We have included [?@fig-smoothing-tech](#) to show how the smoothing technique for window detection works.

[Smoothing Technique in Action]{#fig-smoothing-tech}

4.2.2 Machine Learning Method

4.3 Evaluation Metrics

4.3.1 Standard Metrics

4.3.2 De Rautlin de Roy Metrics

4.3.3 Custom Metrics

5 Results

6 Conclusions

[Windows presentation](#)