Detecting night opening of windows in summer: a case study

Valentin Hoyet, Mathieu Bouville, Maxime Robillart, Marie-Lise Pannier

Detecting night opening of windows in summer: a case study

Valentin Hoyet[1,2], Mathieu Bouville[2], Maxime Robillart[2], Marie-Lise Pannier[1]

[1]*Univ Angers, LARIS, SFR MATHSTIC, 49000 Angers, France*
[2]*Kocliko, 74 cours Aristide Briand, 33000 Bordeaux, France*

ABSTRACT: One consequence of climate change is that summer heat waves increase in frequency and intensity, and so do uncomfortably hot summer nights. One effective way to counter this effect is night ventilation. As connected temperature sensors are growing more common, they provide easily accessible data, from which nocturnal window openings can be inferred. Machine or deep learning algorithms, common in the literature, tend to require many labelled data. This article instead proposes an algorithm relying on expert rules and knowledge of the thermal behaviour of the building. It detects night openings or closings based, respectively, on a decrease or increase of indoor temperature. In our case study of an apartment equipped with connected window and temperature sensors, the algorithm is able to estimate the number of hours of night ventilation to within ±15%. It was more reliable in the hot summer of 2022 than in the cooler summer of 2021: the hypothesis of an effect of opening on temperature was then more realistic. This knowledge will eventually be used to quantify the potential temperature decrease from opening windows during summer nights, so as to provide residents with personalised advice on their window management.

Keywords: building energy simulation, night ventilation, summer thermal comfort, temperature sensors, window openings, rule-based algorithm.

Received January 1, 2021; **In final form** January 31, 2021; **Published** June 2021

Corresponding author: Marie-Lise Pannier, e-mail: marie-lise.pannier@univ-angers.fr

1. INTRODUCTION

A comfortable and healthy environment is essential for the well-being and health of humans in buildings. Nevertheless, it can be difficult to make a home thermally comfortable in all seasons. In summer, it can be especially challenging in areas with high temperature and humidity levels. Night ventilation —opening windows and allowing cooler outside air to circulate throughout a building— can contribute to reducing indoor temperatures and to creating a more comfortable sleeping environment during hot summer nights. It is also a free complement to energy-intensive air conditioning.

By automatically detecting when windows are open, building automation systems can adjust heating and cooling, and thus reduce energy waste and cost. Moreover, knowing people's window-opening habits makes it possible to give feedback to inhabitants on good practice toward a better thermal comfort in summer.

1.1. Background

Several methods have been proposed to detect human behaviour in buildings, such as window opening. In their review article, Carlucci *et al.* [1] split these methods into three categories: stochastic, data-driven and rule-based.

1.1.1. Stochastic models

Stochastic methods estimate an outcome by assuming a probabilistic relationship with one or more variables (e.g. probability that an occupant open a window if he enters the room between 8 and 8:15 a.m.). Specifically, they focus on the probability that a window event will occur, using algorithms such as Bernoulli [2] or Markov [3] processes. The Bernoulli approach takes into account only the current situation, thereby ignoring the dynamic process that leads the occupant to open a window. Markov processes correct this shortcoming by making each prediction dependent on the prediction of the previous time steps. While these processes can provide accurate predictions, they only predict the opening of the window and not the closing.

1.1.2. Data-driven models

Data-driven models are a common recent extension of stochastic models. They find a relationship between the input data and the probability of the window state using a statistical model that can learn and adapt without following explicit instructions.

Some of these studies, based on logistic regressions, focus on understanding thermal comfort and

the environmental triggers of window opening. The coefficients of a logistic regression show that the indoor CO_2 level is the most important variable explaining the opening of windows, and outdoor temperature is the main driver for their closing [4]. A statistical analysis based on logistic regressions found that characteristics of the dwelling area of the residence or home insulation have a strong impact on window opening behaviour: large dwellings (> 93 m^2) and those with recent insulation are more likely to have their window open [5]. This study also found that adding variables pertaining to physical characteristics of the dwelling, such as on which floor it is or which direction it faces, improved the prediction.

Another study aimed at improving thermal dynamic simulations, by improving the modelling of human behaviour in residential buildings, in order to obtain a more accurate prediction of the energy performance of a building. The study used XGBoost to reach 80% precision for detecting whether a window was open or shut [6].

Li and coworkers [7] propose a logistic regression that predicts window opening probability, in order to study its impact on the energy consumption of the building.

One limit to data-driven models is that they are specific to the environment they were trained in and will not be able to provide accurate results in a new environment unless they are specifically recalibrated for it [8].

1.1.3. Neural networks increase the variety of input

De Rautlin de la Roy *et al.* [9] use outdoor and indoor air temperatures, outdoor and indoor relative humidity and indoor CO_2 concentration to predict the state of a window. They show that neural networks such as long-short-term-memory (LSTM) and gated recurrent unit (GRU) work slightly better than more traditional machine learning such as linear discriminant analysis (LDA) or support vector machine (SVM) models.

More complex sorts of variables can also be used to detect window openings, but these require more specialized methods. For instance, recent studies used real-time images of the outside façade of buildings and convolutional neural network (CNN) to detect whether a window was open [10, 11]. (This may not work at night, which would be an issue for us.)

Audio signal record has also been passed through LSTM-CNN classification to predict events in a residence, including window opening [12]. A microphone is placed in a non-echoing room. When an audio signal is detected, it is first converted into a grey scale image that is then classified by CNN as an event such as shower audio or window opening.

1.1.4. Rule-based models

Rule-based models use a set of expert rules based on schedules or profiles to predict human behaviour. This category is by far the least common. A study [13] predicts the probability of windows being open in offices using outdoor temperature: the higher the outdoor temperature the higher the probability of windows being open, reaching 80% at an outside temperature of 33°C.

1.2. Aim of the study

1.2.1. Those who open windows and those who do not

Our goal is to be able to know which occupants in a building open their windows at night in the summer. Such information would enable us to let those who do not, know of potential improvement to their comfort in summer by opening windows. And we could do so without disturbing those who already regularly open their windows.

Our long-term objective is to provide personalized advice to residents, by suggesting to them useful times when to open their windows. This will in addition require to quantify the effect of performing the suggested action and communicate it to the users (e.g., “opening your windows for 2 hours will decrease indoor temperature by 2 °C”).

To the best of our knowledge, no study has so far been dedicated to night ventilation in summer.

1.2.2. A rule-based model

Many models aim at establishing the general probability of some window being open given its environment [2, 3, 6, 7, 13]. The goal of the present article is instead to know whether a given window is open at a given time.

We rely on (indoor and outdoor) temperatures as input, because connected sensors are growing in number anyway: their data can be used for our purpose without extra cost (unlike for instance installing opening sensors for this sole aim).

Although less common in the literature, a rule-based approach is a good option since detecting the window opening itself becomes a deterministic

problem. It requires smaller amounts of hard-to-obtain labelled data than data-driven methods. The latter also have as a drawback that they are trained in a specific environment and may not be portable to other buildings. To overcome this difficulty, we use building-specific physical information (such as insulation and inertia) to root our work in the specifics of the dwelling we study, as do [5, 14].

2. METHODS

2.1. Detecting whether a window is open

Since the focus is on summer night ventilation, the method seeks to detect window openings based on a significant drop in indoor temperature. Conversely, the window is deemed shut again when the indoor temperature rises back. This relies on two variables, indoor and outdoor temperatures, measured every 30 minutes, room by room.

Assuming that there is no internal heat gain, it is possible to approximate the evolution of the room temperature T_{room} at night by:

$$\mathrm{C}\frac{\mathrm{d}T_{room}}{\mathrm{d}t} = \mathrm{U} \times S \times (T_{out} - T_{room}), (1)$$

with:

- C: heat capacity, in W.h/K, of the air and light walls;
- U: heat transfer coefficient per surface area of wall in contact with the outside, in W/(m².K);
- S: area, in m², of the surface in contact with the outside.

A window opening is deemed to take place when the temperature decreases more than would be expected from Eq. (1):

$$\left|\frac{\Delta T_{room}}{\Delta t}\right| > \left|\frac{1}{\tau} \times (T_{out} - T_{room})\right|, (2)$$

with $\tau = \frac{C}{\mathrm{U}\times S}$ the characteristic time in hours of the apartment.

A deep knowledge of the building is needed to compute this constant as it depends on the heat capacity (C) and heat transfer coefficient (U). The time constant τ for the apartment of the case study (§ 3) has been estimated at 21 hours using a model of dynamic thermal simulation [15].

The algorithm takes as input the indoor and outdoor temperatures (as well as the aforementioned constants) and returns a classification: the window is open (output of 1) or closed (0).

2.2. Algorithm

2.2.1. Logic for the algorithm

The algorithm contains two parts. First, it detects an opening event using the criterion of equation (2). To be taken into account, the temperature drop must be greater than 0.1 K (the sensor resolution).

Then, for each opening found, the algorithm infers the closing of the window according to the condition: indoor temperature has risen by 0.1 K (again, the sensor resolution) either in the last 30 minutes or compared to the lowest temperature reached during the opening.

2.2.2. Algorithm

Here is the algorithm used for the detection, a description is proposed in 2.2.1.

```
FOR EACH i IN temp_in
    derivative_in[i] = temp_in[i] − temp_in[i−1]
    delta_temp = temp_out[i] − temp_in[i]
    IF derivative_in[i] < min_drop
    AND abs(derivative_in[i]) > abs(delta_temp) / tau
        open_estimated[i] = True
FOR EACH i IN open_estimated
    IF open_estimated[i] == True
    AND open_estimated[i−1] == False
        j = i + 1
        temp_min = temp_in[i]
        delta_temp = derivative_in[j]
    WHILE delta_temp < min_rise
        IF temp_in[j] >= temp_min + min_rise
            BREAK
        ELSE IF temp_in[j] < temp_min
            temp_min = temp_in[j]
            open_estimated[j] = True
            j = j + 1
            delta_temp = derivative_in[j]
```

Algorithm 1: Window opening detection

2.2.3. Variables

All variables are floats (or 1D arrays thereof), except *open_estimated* which is a 1D array of Booleans. The input takes the form of two 1D arrays:

- *temp_in*: indoor temperature.
- *temp_out*: outdoor temperature.

These and other 1D arrays hold one value per timestep (every 30 minutes).

To ensure that the data used are accurate and do not include any errors (e.g. outliers or gaps in the

data) arising either from the sensor itself or from a problem during data collection, both temperature time series are run through multiple layers of data analysis detailed in annex.

Two local variables are 1D arrays:

- *derivative_in*: first derivative of *temp_in* over 30 minutes.
- *open_estimated*: is the window open? (Boolean).

There is also a float local variable:

- *temp_min*: lowest temperature reached during a given opening, initialized to the first temperature of the opening.

The algorithm also uses three float constants:

- *tau*: characteristic time (see above).
- *min_rise*: minimum temperature increase to consider a closing of the window; set to 0.1 K.
- *min_drop*: minimum drop of temperature to consider an opening; set to 0.1 K.

3. CASE STUDY

This algorithm was tested on a 4-room apartment occupied by 2 people and located on the fourth floor of a building near Bordeaux, France (Figure 1) during the summers of 2021 and 2022. The apartment comprises a living room, a kitchen, a study, and a bedroom, as shown in Figure 2. It does not have air conditioning.

Figure 1: Photograph of the apartment, highlighted in red, from the street.

The living room and kitchen face south; bedroom and study, west. The bedroom and the study have a window each; the living room and the kitchen each have a French window, which can be opened to give access to the balcony. The windows are equipped with opening sensors (green dots in Figure 2): they provide a list of the times when their state changes.

The apartment is also equipped with temperature sensors (red dots). They were positioned carefully following technical recommendations: 1.5 m above the floor and away from direct radiation. In this instance, they were installed on furniture or doors, lest the tenants drill holes in the walls.

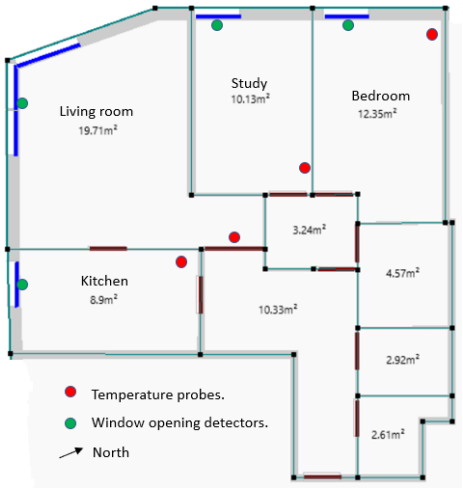

Figure 2: Floorplan of the apartment

These measure indoor temperature every 30 minutes. Since the measurements may not be in phase, the temperature is resampled. For example, the temperature given as 10:30 represents the measure that happened between 10:01 and 10:30. Although the sensors allow a higher frequency, (i) the focus of the study is specifically on night ventilation, so the risk of missing short openings is not a costly one, and (ii) the endgame being to apply the algorithm on a larger scale, it is important to reduce the frequency, and thus the cost, of battery replacement (including manpower and the trouble of entering peoples' homes) through a higher timestep.

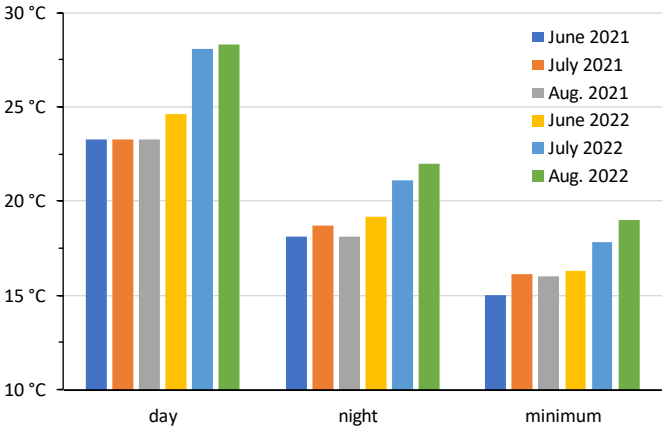

Figure 3: Outdoor temperature, by month, in the summers of 2021 and 2022. Day: 10 a.m. to 8 p.m.

Table 1: Window openings room by room, and month by month in the summers of 2021 and 2022. For each room, left column: total number of openings, centre: total duration [hours], right: median duration per opening [hours]. The last row is the tally over the six months.

room →	living room			kitchen			bedroom			study		
↓ month	nb.	hours	med.	nb.	hours	med.	nb.	hours	med.	nb.	hours	med.
June 2021	44	125	0.6	24	87	1.0	35	38	0.8	28	48	0.8
July 2021	32	98	1.3	12	64	9.8	13	18	1.2	17	38	0.6
Aug. 2021	6	18	3.8	4	17	8.5	5	8	0.3	5	8	0.6
June 2022	40	104	1.2	19	111	3.8	34	45	1.0	20	29	1.2
July 2022	28	199	5.3	23	188	6.7	25	47	1.6	18	36	1.1
Aug. 2022	30	176	4.6	27	149	5.3	32	29	0.6	27	57	0.8
6 months	180	720	1.5	109	616	5.1	144	185	0.8	115	216	1.0

For the case study, no outdoor sensors were used. (They should face north to avoid direct sunlight, which is easier to achieve at the scale of a building than a single apartment, especially one facing south and west.) The outdoor temperature is thus retrieved from a weather station with a frequency of one hour. This is interpolated linearly down to 30 minutes to be in phase with indoor temperatures. The station being outside the city, by the airport, it is likely to have a lower average temperature and less inertia, thus underestimating temperature (especially in the evening). In future, both indoor and outdoor temperature sensors should be available, and thus the local temperature.

Figure 3 shows, month by month in the summers of 2021 and 2022, the average day and night outdoor temperatures. It also shows the lowest temperature in the morning (specifically the median of these 30 or 31 values for the month). These latter two numbers hint respectively at the possibility to open the windows overnight and at the efficiency of opening them in the morning. Clearly, on all three criteria, July and August of 2022 are hotter than the other four months. Table 1 quantifies window openings, room by room and month by month. It shows that there are far more openings in July and August of 2022 (by number and total duration), which are also the hottest months of the study (Figure 3). August 2021 will not be used hereafter because the residents, being away on holidays, did not open windows.

Living room and especially kitchen tend to have their windows open all night long, whereas bedroom and study have mostly hour-long openings. Residents explain this difference (especially for the bedroom) by the fact that the bedroom and study overlook a noisy street.

When a window is open, it takes on average one to two hours for the temperature to decrease by 1 K. The bedroom and study cool down a little faster because (i) they are smaller and (ii) their openings are shorter (in an exponential decrease, the time derivative is initially higher, so the speed of the temperature drop is on average greater for shorter openings, even though the total decrease is smaller).

4. RESULTS

The algorithm checks, at each time step (every 30 minutes), if the window in the room is open (‘positive’) or closed (‘negative’). Comparing to actual state gives rise to four possible outcomes. For example, if the algorithm deems the window closed, and the window is closed, then the classification is a true positive, TP. A false positive, FP, sees open a window which is in fact closed. A false negative, FN, conversely calls shut an open window.

Figures 4 to 7 show night openings over the same week in August 2022 for each of the four rooms. The openings predicted by the algorithm are compared to those identified by the sensors using coloured strips: green for true positive, yellow for false positive and red for false negative. The white background is the default: the window is closed and the algorithm predicts so (true negative). As stated before, the algorithm tries to detect an opening for every time step.

All times are given in local time (GMT + 2 hours). Since the present study focusses on night ventilation, openings during the day (10 a.m. to 8 p.m.) are ignored; these would not be about cooling but for instance simply to go out onto the balcony.

The apartment was unoccupied on the 14th and 15th of August, which is why there are no openings for these two days.

4.1. Living room

Figure 4 shows results for the living room. Five night-long openings are correctly detected (green strips). There is often a false negative (in red) in the evening or a false positive (orange) early in the morning. Evening false negatives indicate that the algorithm tends to lag actual window openings. This is sometimes due to the fact that openings do not trigger an immediate drop of temperature in the room and thus cannot be detected right away by the algorithm (for instance, on the evenings of the 7th, 8th, 9th and 10th).

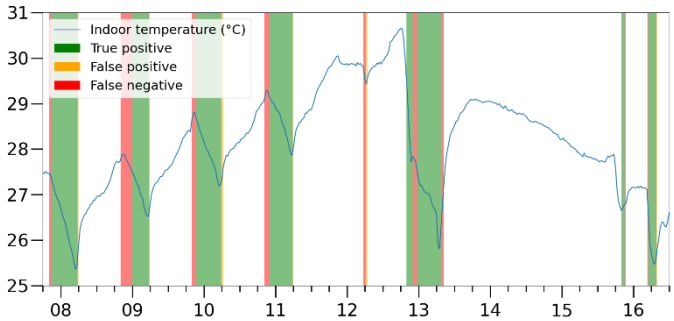

Figure 4: Indoor temperature (line) and predictions (colour strips) for the living room for one week in August 2022.

During the last night in the figure, there is a double opening: the window is open in the evening of the 15th and in the morning of the 16th, but closed during the night itself. The algorithm detects this pattern correctly.

In the morning of the 12th a false positive follows a false negative. It thus seems that the algorithm misses the opening completely (no true positive). But it is in fact a timing error rather than an outright miss: there is a short opening (less than 30 minutes) of the window which is detected a little late and then the opening lingers in the prediction after the window has been actually closed: false positive follows false negative, but the opening has nonetheless been spotted.

4.2. Kitchen

Figure 5 indicates that, as in the living room, the algorithm tends to predict openings later than they actually happen in the kitchen when these openings start with a rise of temperature in the room.

The algorithm detects two unlabelled openings in the kitchen, on the 10th and 13th. These drops in temperature are quite visible, but arise from openings in the living room (Figure 4), not the kitchen.

In the morning of the 12th an opening is missed by the algorithm. This opening lasting less than 30 minutes (the period between temperature measurements) makes it hard for the algorithm to spot.

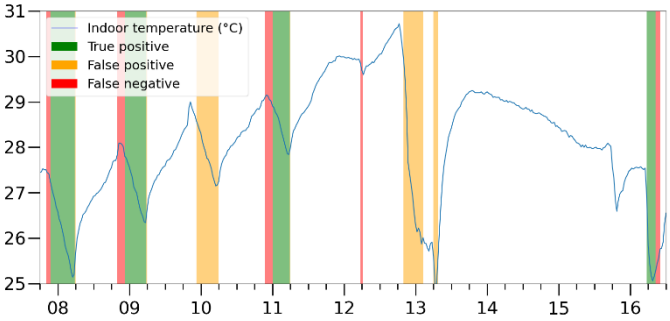

Figure 5: Indoor temperature (line) and predictions (colour strips) for the kitchen for one week in August 2022.

There appears to be a temperature drop in the evening of the 15th. But since it took place unusually early (between 5:30 and 7:30 p.m) it was not deemed *night* opening.

4.3. Bedroom and study

Unlike the living room and the kitchen, the bedroom and the study rarely have night-long openings. Figure 6 highlights openings in the bedroom missed by the algorithm on the first five mornings.

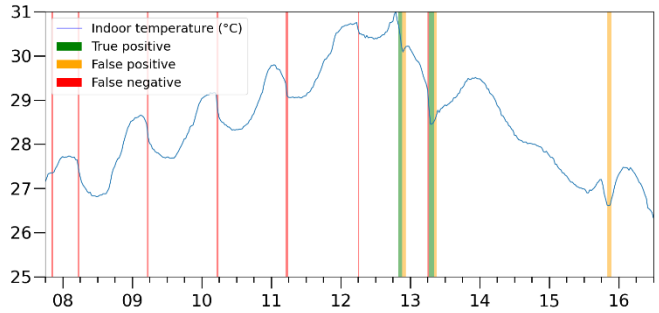

Figure 6: Indoor temperature (line) and predictions (colour strips) for the bedroom for one week in August 2022.

This comes from the details of the criteria used by the algorithm to call an opening, equation (2). On the 10th, for instance, the largest temperature drop occurs between 5:00 and 5:30 a.m.: −0.34 K in half an hour. Calling this an opening would have required a decrease by −0.35 K: the false positive is due to missing the threshold by a mere 0.01 K.

One must bear in mind that openings lasting far less than 30 minutes (the period of the measurements) are bound to result in seemingly small effects on the temperature. A temperature drop of 0.34 K in 10 minutes (2 K/hr) is thus treated as a drop of less than 0.7 K/hr. It makes it difficult to measure their impact on the temperature of the apartment.

In the study, Figure 7, the openings are generally short. However, they are deep enough for the algorithm to identify them. The problem here is the false positive which follows in the morning. The condition for the algorithm to consider a closing of the window is not sensitive enough.

The algorithm misses a big opening on the 16th. This is because after an initial drop the temperature bounces back, which the algorithm interprets as a closing. The temperature does start dropping again later, but not enough for the algorithm to consider an opening.

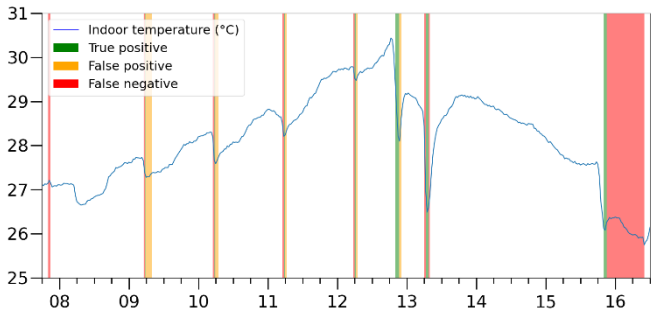

Figure 7: Indoor temperature (line) and predictions (colour strips) for the study for one week in August 2022.

4.4. Overall results

Table 2: Results for the four rooms, by month, in the summers of 2021 and 2022. For each room, the columns are: precision (%), recall (%). The last row is the average over the five months.

room →	living room		kitchen		bedroom		study	
↓ month	Pr.	Rec.	Pr.	Rec.	Pr.	Rec.	Pr.	Rec.
June ’21	56	66	64	70	44	35	34	46
July ’21	51	57	49	58	22	31	48	47
June ’22	61	74	72	54	37	26	56	56
July ’22	97	79	87	79	57	41	65	57
Aug. ’22	86	83	85	84	29	34	49	57
average	70	72	71	69	38	33	50	53

Figures 4 to 7 (and Table 1) showed that the windows of the living room and kitchen often remained open all night long, unlike in the bedroom and the study (even though the number of times the windows were open is similar).

Table 2 measures the capacity of the algorithm, for each of the four rooms, to detect an opening at a given time step thanks to two metrics, precision and recall. In the definitions, ‘TP’ is the number of true positives, ‘FP’ that of false positives and ‘FN’ false negatives.

Precision, $\frac{TP}{TP+FP}$, is the probability that the algorithm be right when it makes a positive prediction (predicts that a window is open). A high precision corresponds to few false positives: predicted openings are likely to be genuine.

Conversely, *recall* normalizes based on windows which are actually open rather than on windows predicted open: $\frac{TP}{TP+FN}$. A high recall means few false negatives: we miss few openings.

The living room and kitchen have better results in July and August 2022 with an average precision of 89% and recall of 81% than in June (respectively 66% and 64%) and 2021 (55% and 63%).

The overall performance of the other two rooms is poor. This is probably due to the openings being short (possibly shorter than the period of the temperature measurement) and thus hard to identify.

Precision is more sensitive to temperature than recall. The former doubles between July 2021 and July 2022 while the latter increases by only a third. Figure 8 shows that, on a hot month, precision is better than recall (fewer false positives than false negatives) — the converse is true for the cooler month.

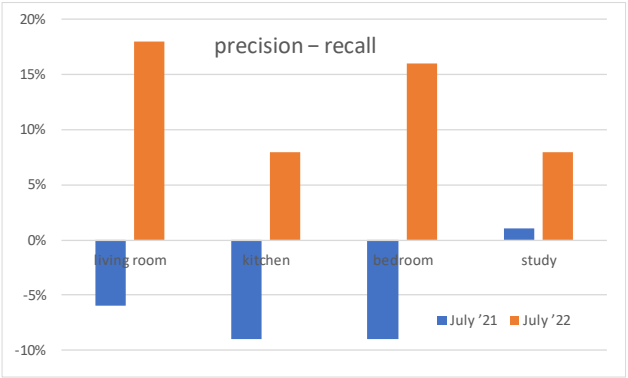

Figure 8 : Precision minus recall for the four rooms in Julies of 2021 and 2022.

Table 3: Relative error of the predictions. Left column for each room: the number of times the windows are predicted open divided by the number of times the windows are actually open, minus 1. Right column: same based on the total duration of the openings, rather than on their number.

room →	living room		kitchen		bedroom		study		average	
↓ month	number	duration	number	duration	number	duration	number	duration	number	duration
June 2021	− 22%	+ 19%	+ 12%	+ 9%	− 60%	− 21%	− 24%	+ 36%	− 23%	+ 13%
July 2021	− 5%	+ 11%	+ 40%	+ 18 %	− 34%	+ 37%	− 8%	− 3%	− 2%	+ 13%
June 2022	− 33%	+ 22%	0%	− 26%	− 71%	− 28%	− 40%	0%	− 36%	− 7%
July 2022	− 20%	− 18%	+ 9%	− 8%	− 56%	− 28%	− 34%	− 12%	− 25%	− 15%
Aug. 2022	0%	− 3%	+ 8%	− 1%	− 67%	+ 14%	− 18%	+ 16%	+ 8%	+ 1%

5. DISCUSSION

Table 3 shows the capacity of the algorithm to estimate the number of times or of hours the windows have been open for each month, in each room. For the summers of 2021 and 2022 the algorithm could estimate the duration of night ventilation with errors better than ±15% at the scale of the entire apartment. The number of openings is less reliable, especially the short openings typical of the bedroom and study.

5.1. Criteria for success

Success depends on the goal sought — especially its temporal and spatial scales. If it is crucial to know exactly in what room the window is open then the contagion between rooms is a genuine problem. On the other hand, if the question is just whether some ventilation is taking place, getting the window wrong is unimportant so long as the algorithm can tell that some window is open (if not which).

Likewise, if the aim is knowing when windows are open (for instance in real time), a delay in detection is troublesome. But if the purpose is more statistical —how often (and/or how long) are windows open—, a pure offset (delay of the same length to detect when the window is opened and when it is closed again) has no impact.

5.2. Wrong room

Figure 5 shows that two night-long openings are detected in the kitchen, when the window is in fact closed. This is a form of contagion: when one window is open (here, in the living room) the temperature also drops elsewhere in the apartment. This leads the algorithm to believe that windows in other rooms are also open, hence the false positives (orange strips). This explains why some rooms (especially the bedroom) perform poorly in Table 3.

To factor out such an effect, Figure 9 shows the total number of windows open (between 0 and 4), over the same period as in Figures 4 to 7. The algorithm (purple curve) may get the exact number (in blue) wrong — for instance believing that windows are open in all four rooms when some are in fact closed. Yet, it is far less likely to confuse the case of all windows closed with that of (at least) some windows open (sum ≥ 1).

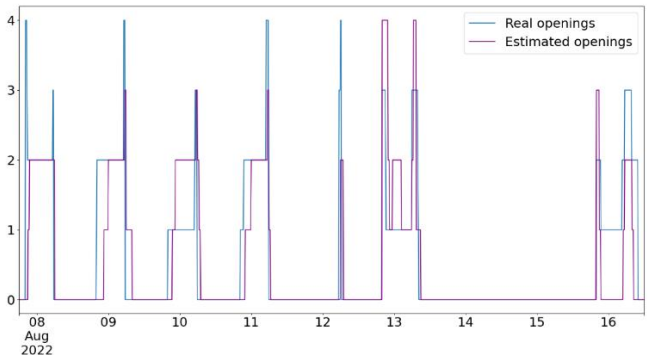

Figure 9: Number of windows open

5.3. Wrong time

Figure 10 shows, for August 2022, statistics at the scale of the apartment as a function of the time of day: the average number of windows which are open (top) and the probability that at least one window be open (bottom). It indicates that openings take place mostly between 8 p.m. and 7 a.m. (The solid lines correspond to the periods when the algorithm was applied.)

Overall, the algorithm detects the correct number of openings. But the bottom of Figure 10, especially, shows a clear delay in the prediction (purple curve). This offset explains the frequent succession of false negatives and false positives in Figures 4–7.

In some cases, like in Figure 4, the delay in detecting the opening comes from the fact that the temperature initially rises after the window gets

opened. This kind of opening is not considered a plausible one for ventilation and thus cannot be detected by the algorithm. However, there is no symmetric reason for the delay in detecting the closing.

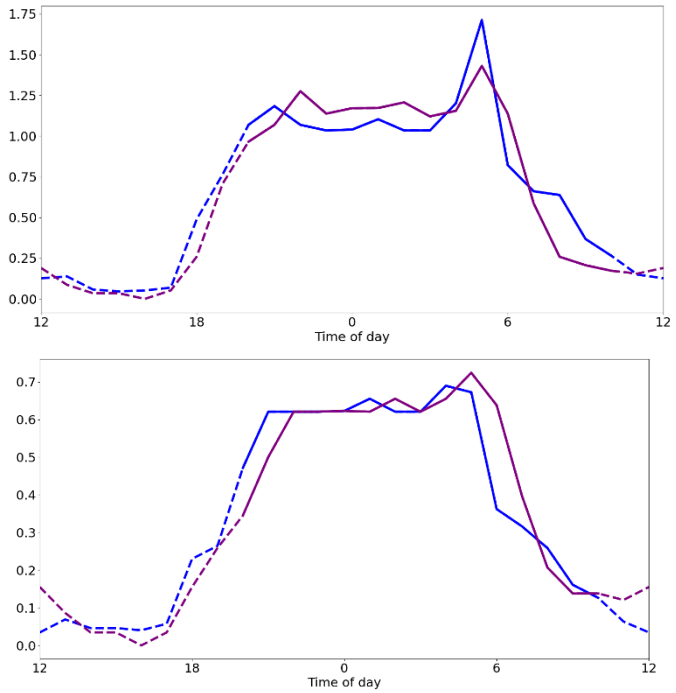

Figure 10: Average number of windows open in the apartment (top) and probability of at least one window open (bottom) by time of the day for August 2022. Blue: real, purple: prediction.

One possibility for fixing this problem could be to have an approximation of the expected temperature evolution while the window is open — the same way equation (2) provides an approximation of the temperature evolution when the window is closed. From this, it would be possible to define more precisely the conditions for the closing of the window. That said, such an approximation would require more knowledge on the environment to be realistic. For example, wind would have a major impact on temperature evolution when the window is open.

A very pragmatic way of improving predictions is: cheating. Since the algorithm consistently delays the detection of events by at least 30 minutes, then one can simply offset all predictions by this amount. When the algorithm says that the opening of the window took place at 9:00, one can improve this raw prediction by turning it into a post-processed prediction of 8:30.

5.4. Comparing cool and hot months

Figure 11 compares the probability of at least one window being open in July 2021 (top) and July 2022 (bottom). (August cannot be used for this purpose because the inhabitants were on holiday in August 2021.)

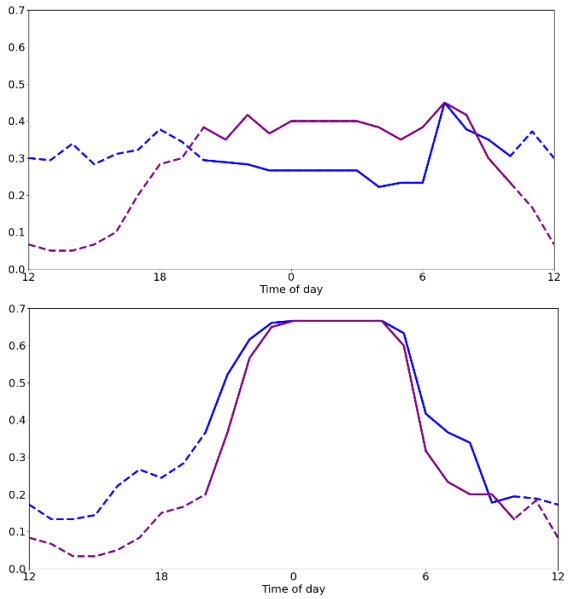

Figure 11: Probability of at least one window open by time of the day. Top: July 2021, bottom: July 2022. Blue: real, purple: prediction.

July 2022 shows a sharp difference between night and day: the windows are opened late in the evening and closed early in the morning. In cool July 2021 the policy is more relaxed than in hot July 2022, with one third fewer openings at night (these are unnecessary) and, conversely, regular openings throughout the day (since it is cool outside, opening a window is not punished by an inflow of hot air).

5.4.1. Useful night ventilation

Clearly, the model (blue lines) is less accurate for the cool summer of 2021 than for the hot summer of 2022. Since regulating the temperature was less crucial in 2021, fewer openings were directly intended to cool the apartment than in 2022. (In particular, the apartment of the case study having a balcony, another possible purpose of opening the living room French window is to access it, regardless of the temperature.)

The overarching philosophy of the algorithm makes it capable of detecting only *useful* night ventilation, i.e. a night window opening associated with a significant drop of temperature in the accommodation. In July 2021, the algorithm (blue curve) enforces a pattern that does not actually exist.

The model still has possible improvements which will be the focus of future work. The natural first

step was to treat the rooms separately, but correlations between rooms (convective contagion) could come next. However, it depends very much on the configuration of the apartment, so this may be hard to automate.

5.4.2. False positives

Figure 8 shows a difference between precision and recall. It may be due to the inertia of the walls. Opening windows quickly lowers the indoor air temperature. When indoor and outdoor air are at equilibrium, the wall temperature then decreases slowly. In practice, one closes windows in the morning because outside temperature has risen above inside air temperature. The walls which are still warm then heat up the indoor air. But in July 2021, this effect must have been weak: the walls were less warm (closer to thermal equilibrium with the air) and the rise in indoor temperature after closing must have been limited. This increases the delay to recognize a closing of the window, and thus the risk of false positive. We plan on adding the effect of the inertia of the walls to future versions of the model.

6. CONCLUSION

Night ventilation can decrease indoor temperature in summer — a need which will only grow in importance as climate change unfolds. This article proposed an algorithm based on expert rules and the thermal behaviour of the building to evaluate night ventilation in summer. It has the advantage of requiring fewer labelled data than the machine or deep learning algorithms which are more common in the literature. It relies on indoor and outdoor temperatures, which are growing more readily available as the number of connected temperature sensors itself grows: no need to install dedicated equipment. The method also uses characteristics of the specific dwelling, such as its insulation and inertia.

In our case study of an apartment equipped with connected window and temperature sensors, the algorithm estimated the number of hours of night ventilation to within ±15% at the scale of the entire apartment for the months of the summers 2021 and 2022.

Overall, the model works better with night-long openings than for short ones (especially shorter than the time increment of half an hour). But these are the ones which really contribute to cooling in any case.

Openings and closings are identified based, respectively, on a drop and increase of indoor temperature between two consecutive time steps — it explicitly assumes that either event will have an impact on temperature and thus the thermal comfort. We necessarily miss windows opened for a few seconds to let the cat out. This means that the model works better on a hot summer day, when no one casually opens windows.

There are two other types of imprecisions in the model: in time and space. The predictions tend to lag the actual events. For a short opening (especially if shorter than half an hour, our time increment), the method may even detect that a window has been opened only after it has already been shut back. There is also a risk of convective contagion: opening a window in one room may decrease the temperature in the entire dwelling, thereby leading the algorithm to wrongly infer the opening of windows in other rooms.

On the bright side, we intend to now use these results to tell, in a building, who seems not to open windows at night. Instead of telling all residents in a blanket message that they should ventilate at night, we will be able to focus on inhabitants most likely to need such advice. And, since this is at the scale of the dwelling rather than the room and averaged over a summer rather than hour-by-hour, our method is certainly reliable enough for our purpose.

A survey is also currently being designed to find patterns in people's behaviour. For instance, is it the same high up in the building and on ground or low floors (where noise and the risk of burglary are greater)? There may also be a policy of opening in the west in the morning and on the east side in the evening. Such information may be included in the model in a Bayesian fashion: what is the probability of the window facing east being opened on the fourth floor in the morning given the temperature?

ACKNOWLEDGEMENT

This work was performed within the frame of the research project CoLoC (Comfort of Connected Dwellings), granted by the "France Relance" plan for the preservation of R&D employment.

REFERENCES

[1] Carlucci, Salvatore, *et al.* 2020. "Modeling Occupant Behavior in Buildings". *Building and Environment* 174 (May): 106768. https://doi.org/10.1016/j.buildenv.2020.106768.

[2] Schweiker, Marcel, Frédéric Haldi, Masanori Shukuya, and Darren Robinson. 2012. "Verification of stochastic models of window opening behaviour for residential buildings". *Journal of Building Performance Simulation* 5 (1): 55-74. https://doi.org/10.1080/19401493.2011.567422.

[3] Andersen, Rune, Valentina Fabi, Jorn Toftum, Stefano P. Corgnati, et Bjarne W. Olesen. 2013. "Window Opening Behaviour Modelled from Measurements in Danish Dwellings". *Building and Environment* 69 (November): 101-13. https://doi.org/10.1016/j.buildenv.2013.07.005.

[4] Haldi, Frédéric, and Darren Robinson. s. d. 2012. "A Comparison of Alternative Approaches for the Modelling of Window Opening and Closing Behaviour". https://www.researchgate.net/publication/242108226.

[5] Shi, Shanshan, Hongjian Li, Xue Ding, and Xin Gao. 2020. "Effects of Household Features on Residential Window Opening Behaviors: A Multilevel Logistic Regression Study". *Building and Environment* 170 (March): 106610. https://doi.org/10.1016/j.buildenv.2019.106610.

[6] Mo, Hao, Hejiang Sun, Junjie Liu, and Shen Wei. 2019. "Developing Window Behavior Models for Residential Buildings Using XGBoost Algorithm". *Energy and Buildings* 205 (December): 109564. https://doi.org/10.1016/j.enbuild.2019.109564.

[7] Nan Li, Juncheng Li, Ruijuan Fan, and Hongyuan Jia. 2015. "Probability of Occupant Operation of Windows during Transition Seasons in Office Buildings". *Renewable Energy*, Sustainable Development in Building and Environment (SuDBE) 2013, 73 (January): 84-91. https://doi.org/10.1016/j.renene.2014.05.065.

[8] Yao Mingyao, and Bin Zhao. 2017. "Window Opening Behavior of Occupants in Residential Buildings in Beijing". *Building and Environment* 124 (november): 441-49. https://doi.org/10.1016/j.buildenv.2017.08.035.

[9] Rautlin de la Roy, Enguerrand de, Thomas Recht, Akka Zemmari, Pierre Bourreau, and Laurent Mora. 2023. "Deep Learning Models for Building Window-Openings Detection in Heating Season". *Building and Environment* 231 (March): 110019. https://doi.org/10.1016/j.buildenv.2023.110019.

[10] Luong, David, Russell Richman, and Marianne Touchie. 2022. "Towards Window State Detection Using Image Processing in Residential and Office Building Facades". *Building and Environment* 207 (January): 108486. https://doi.org/10.1016/j.buildenv.2021.108486.

[11] Sun, Cheng, Xumiao Guo, Tianyu Zhao, and Yunsong Han. 2022. "Real-Time Detection Method of Window Opening Behavior Using Deep Learning-Based Image Recognition in Severe Cold Regions". *Energy and Buildings* 268 (August): 112196. https://doi.org/10.1016/j.enbuild.2022.112196.

[12] Pandya, Sharnil, and Hemant Ghayvat. 2021. "Ambient Acoustic Event Assistive Framework for Identification, Detection, and Recognition of Unknown Acoustic Events of a Residence". *Advanced Engineering Informatics* 47 (January): 101238. https://doi.org/10.1016/j.aei.2020.101238.

[13] J Fergus Nicol. 2001. "characterising occupant behaviour in buildings: toward a stochastic model of occupant use of windows, lights, blinds, heaters and fan". *Seventh international IBPSA conference (*August): http://www.ibpsa.org/proceedings/BS2001/BS01_1073_1078.pdf.

[14] D'Oca, Simona, and Tianzhen Hong. "A Data-Mining Approach to Discover Patterns of Window Opening and Closing Behavior in Offices". Building and Environment 82 (1 December 2014): 726-39. https://doi.org/10.1016/j.buildenv.2014.10.021.

[15] Peuportier Bruno, and Isabelle Blanc Sommerreux. "Simulation Tool with Its Expert Interface for the Thermal Design of Multizone Buildings". *International Journal of Solar Energy* 8, n° 2 (1 january 1990): 109-20. https://doi.org/10.1080/01425919008909714.

ANNEXES

Data analysis

1. Outliers

Data points whose distance to the mean of their distribution is considered outliers and thus deleted:

$|T_t - \overline{T}| > 3\sigma$. (4)

with:

- T_t: Probe temperature at time step t.
- $\overline{T}$: Temporal mean of probe temperatures.
- σ : Standard deviation of probe temperatures.

2. Data leaps

As the data are often uploaded, it is possible that some data points are missing or not representative of the environment. For instance, the probe may be poorly located or malfunctioning. This can be visualised easily by plotting the temperature time series.

To correct these "spikes" (Figure a.1), they are first detected using the formula below, then the data point affected is deleted from the time series.

$\min(|T_{t-dt} - T_t|, |T_{t+dt} - T_t|) > \varepsilon_1$ (5)

with:

- ε_1: temperature difference threshold, in K. Here, $\varepsilon_1 = 2$ K.

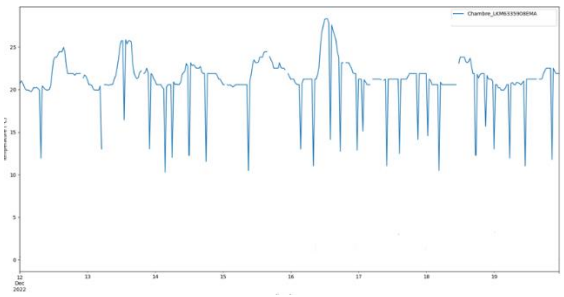

Figure a.1: Temperature time series with data leap over 1 data point

Another kind of spike we can observe in the data spans multiple data points (Figure a.2). In this case, however, it is not possible to identify which part of the signal is physical and which is erroneous. So, if the problem is detected, the probe is considered faulty and therefore entirely removed from the dataset.

$\frac{1}{n}\sum_{t=0}^{t=n}(T_t - T_{t-dt})^2 > \varepsilon_2$ (6)

with:

- n: number of elements in the time series.
- ε_2: temperature difference threshold, in K. Here, $\varepsilon_2 = 5\ K$.

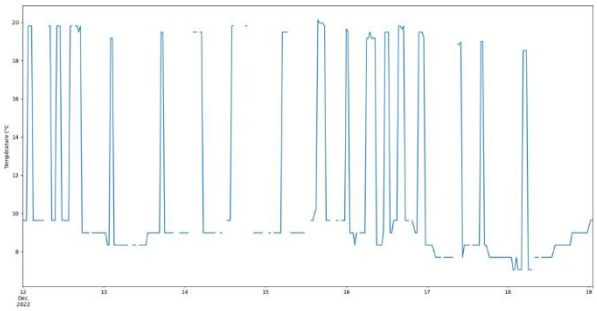

Figure a.2: Temperature time series with data leap over multiple data points

Constant probe

Some probes can stop working over a period without it being notified (figure a.3), leading to a constant signal.

In order to prevent this phenomenon, the formula below checks if the time series is constant over an interval, in which case the interval is deleted.

$\sum_t^{t+n} |T_t - T_{t-dt}| = 0$ (7)

n : length (in number of time steps) of the interval.

In practice, for numerical reasons, a low threshold, rather than zero, will be used.

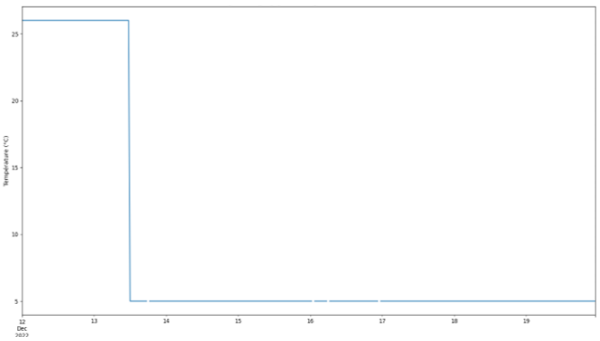

Figure a.3: Temperature time series with a constant probe