Reflections on alternative modelling approaches regarding occupants' window operation behaviour

Christiane Berger[*1] and Ardeshir Mahdavi[2]

1 Department of Architecture, Design and Media Technology, Aalborg University, Aalborg, Denmark
**chbe@create.aau.dk*

2 Institute of Building Physics, Services, and Construction, Faculty of Civil Engineering Sciences Graz University of Technology, Graz, Austria.

ABSTRACT

Computational predictions of buildings' indoor-environmental conditions and energy performance would presumably benefit from the inclusion of models that could reliably capture occupants' window operation behaviour. Frequently, models derived from empirical data have a black-box character. However, the utility of window operation models could be conceivably improved, if the model derivation process is preceded by specific hypotheses regarding the variables that are assumed to influence the frequency and timing of window operation actions. In the present contribution, we discuss the process of exploring explicit hypotheses regarding factors that could influence occupants' operation of windows prior to the model derivation step. To illustrate the potential of this approach, we utilize a specific window operation data set from an open plan office. This data set was used to test three distinct hypotheses regarding the factors that influence occupants' window operation actions upon arrival. The results suggest that the most plausible conjecture from the intuitive point of view is not supported by the data set. This observation encourages more in-depth reflections on the motivational background of occupants' behaviour. Purely data-driven black-box models arguably do not provide a similarly strong impetus toward an explicit understanding of occupants' behaviour patterns in buildings.

KEYWORDS

Natural ventilation, window operation, occupant behaviour, computational models

1 INTRODUCTION

Natural ventilation in general and manual operation of windows in particular represent the oldest and most primary means of modulating the rate of air change in buildings and hence influencing both indoor air quality and thermal conditions (Etheridge 2011). Besides indoor-environmental conditions, operation of windows can also influence buildings' energy use. Hence, computational predictions of buildings' indoor-environmental conditions and energy performance would presumably benefit from the inclusion of models that would capture occupants' window operation behaviour. These models may range from simple schedules to sophisticated probabilistic routines (Mahdavi et al. 2016, Tahmasebi and Mahdavi 2019, 2016). Either way, development and validation of window operation models must be ultimately based on observational data.

Frequently, models derived from empirical data are of a black-box type: Observed behaviour is statistically correlated with values of variables pertaining to, for instance, thermal conditions inside and outside buildings (D'Oca and Hong 2014). It is not suggested here that black-box models would be either deficient or ineffective. Depending on the application scenario, such models may be very useful. However, the utility and scalability of window operation models could be arguably improved if, at the outset of the model derivation process, specific hypotheses are stated regarding the variables that are assumed to influence the frequency and timing of window operation actions, particularly when such hypotheses include transparent formulation of the reasoning (and respective narratives) regarding the mapping from independent to dependent variables. Instead of recurrent tweaking of coefficients of black-box models to fit ever new sets of data coming from different buildings, explicit hypotheses and their testing could help to: *i)* identify those independent variables that are more likely to constrain occupant behaviour, *ii)* shed light on the potential causal mechanisms involved, and *iii)* augment the scalability of local window operation models toward more generally applicable computational routines.

In the present contribution, we reflect on the process of formulating and testing explicit hypotheses regarding factors that could influence occupants' operation of windows. To illustrate the potential of this approach, we utilize a specific window operation data set from an open plan office (in a university building in Vienna, Austria). This year-long data set includes time series information on indoor and outdoor temperatures, occupants' presence, and their window operation actions (for the present study we focused on the warmer months of the year, i.e., from May to September). At the outset, examples of general hypotheses are formulated. These are basically conjectures about what factors (i.e., independent variables or predictors pertaining to indoor and outdoor conditions) could influence the dependent variables (i.e., the probability of window opening actions upon occupants' arrival in the office) and what reasoning or narrative is behind these conjectures. This approach is suggested to support the derivation of transparent and scalable window operation models that can be integrated in computational tools for the prediction of buildings' energy and indoor-environmental performance.

2 METHODOLOGY

As outlined in the introduction, the purpose of the present paper is to explore the process of testing explicitly formulated hypotheses regarding predictors of occupants' window operation against collected data. It is of course possible to confront an observational data set with a set of variables to see if patterns of influence on specific behavioural manifestations (window opening actions in the present case) can be extracted. However, in the absence of explanatory (e.g., causal) stories behind the pattern, the applicability of such a model to other settings (and respective data sets) remains questionable. It thus seems useful if the model development process would start with transparent hypotheses as to what influence mechanisms are postulated and why. To illustrate this possibility, we consider here, as a case in point, the occupant-driven window operation in an office building. Note that the purpose here is not to document a comprehensive instance of a related model development exercise. The illustrative case study involves neither a comprehensive and representative repository of monitored data, nor shall we conduct a detailed statistical analysis to develop or validate a full-fledged window operation model. Rather, the idea is to use a limited data set and simple descriptive statistics to conceptually illustrate the advantages of the proposed approach.

Predicting the timing and frequency of occupants' operation of windows is by all accounts a challenging endeavour, given the extensive set of circumstances that may trigger window operation actions. This set includes, amongst other factors, thermal, air quality, and acoustic conditions inside and outside buildings as well as factors related to individual occupants (activity, clothing, age, sex, health, preferences, habits, cognitive load). Moreover, occupants' position (for instance, in office buildings, the distance of the occupants' workstations from the window units), the quality of window opening interface (ease of opening, potential interference with shading operation, etc.) as well as social settings (e.g., hierarchical relationships amongst occupants in an open plan office) may play a role.

Given the illustrative nature of the present treatment, we focus here on a few factors only, consider a limited data set, and perform a rather simple descriptive statistical analysis. As dependent variable, we focus on the probability of window operation actions upon occupants' arrival in the office. Hence, intermediate window operations are not considered. As to potential influencing factors, we consider only indoor temperature, outdoor temperature, and the so-called comfort temperature (i.e., the ambient air temperature assumed to be preferred by the occupants). Hence, individual occupants' characteristics are not taken into account, nor are workstation configurations and social settings.

Given these choices, we proceed to formulate a number of general (rather qualitative) hypothesis regarding the candidate independent variables and how they influence the dependent variable (window opening probability upon arrival in the office). Note that these hypotheses are in part contradictory, as at this stage (prior to being examined against observation), they entail only conjecture-type narratives as follows:

- H_i The probability of opening a window upon arrival in the office is higher, if the temperature in the office deviates from the comfort temperature. In other words, the larger the perceived discrepancy between the comfort temperature and indoor temperature upon arrival, the higher the probability that occupants would open the windows to change indoor thermal patterns.
- H_{ii} The probability of window opening upon arrival can be influenced by the thermally relevant cooling potential of the outdoor air. In other words, insufficient indoor-outdoor air temperature differences can reduce the thermal (cooling) potential of window opening and thus reduce the respective probability.
- H_{iii} The probability of window opening upon arrival can be influenced by the pre-arrival thermal perception of the outdoor air temperature. In other words, larger deviation of outdoor temperature from occupants' comfort temperature can increase the tendency to open windows upon arrival.

To operationalize the hypothesis H_i above, we designate $\Delta\theta_{ic}$, that is the difference between indoor temperature (θ_i) and comfort temperature (θ_c), as the independent variable ($\Delta\theta_{ic} = \theta_i - \theta_c$). The dependent variable (window opening probability P_w) is operationalized in terms of the number of window opening actions n_{op} within a certain time interval divided by the number of first arrivals in the office n_{ar} within the same time interval ($P_w = n_{op} . n_{ar}^{-1}$).

To operationalize the hypothesis H_{ii} above, we designate, aside from $\Delta\theta_{ic}$ (see H_i description above), the difference between indoor temperature (θ_i) and outdoor temperature (θ_e) as an independent variable ($\Delta\theta_{ie} = \theta_i - \theta_e$). The dependent variable is, as in case of H_i, the window opening probability P_w.

To operationalize the hypothesis H_{iii} above, we designate, aside from $\Delta\theta_{ic}$ (see H_i description above), the difference between comfort temperature (θ_c) and outdoor temperature (θ_e) as an independent variable ($\Delta\theta_{ce} = \theta_c - \theta_e$). The dependent variable is, as in case of H_i, the window opening probability P_w.

To test the above hypotheses empirically, monitored data from an open plan office area with eight workstations in a university building (in Vienna, Austria) was used (Mahdavi et al. 2019). Figure 1 shows the floor plan of the office area as well as the default position of the eight workstations. This office area was equipped with a comprehensive monitoring infrastructure. Collected data included occupants' presence, state of windows, and a number of indoor environment variables (including air temperature, humidity, and CO_2 concentration). Outdoor parameters (including air temperature, solar radiation, wind velocity, and precipitation) were also monitored using the building's roof-top weather station. For the purpose of the present exercise, 15-minute interval data over the course of five months (May to September) in a calendar year was drawn on. The data set comprised occupants' arrival time, window operation events, as well as indoor and outdoor temperatures. A window opening action is assumed to have occurred upon the occupant's arrival in the office if it is registered either at the same interval in which the arrival is observed or in the immediate next interval. Later actions are considered as intermediate actions and not included in the analysis. As the building is not conditioned (neither heated nor cooled) in the selected time period, the adaptive comfort theory (and a respective comfort temperature equation) was deployed to derive the comfort temperature based on the mean monthly outdoor temperature (Nicol et al. 2012). Table 1 provides detailed information about the sensor types used to monitor indoor temperature and occupancy. Information regarding other indoor and outdoor sensors can be found in Mahdavi et al. (2019). Table 2 and Figure 2 provide an overview of monthly mean temperature for 2013 in Vienna. Note that, given the illustrative nature of the present treatment, potential confounding effects of other variables (e.g., humidity and wind speed) were not taken into consideration.

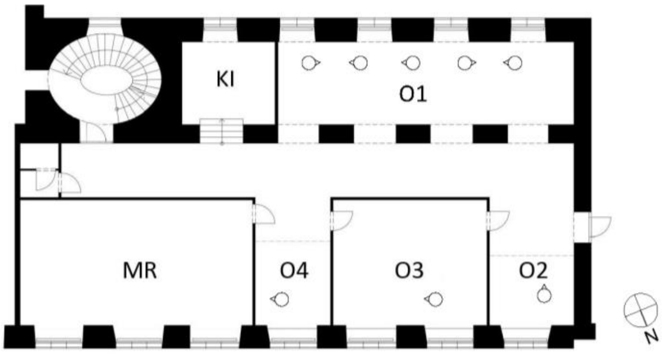

Figure 1: Floor plan of open plan office area (Mahdavi et al. 2019)

Table 1: Description of sensor types for indoor temperature and occupancy (Mahdavi et al. 2019)

Sensor type	Measured variable	Range	Accuracy
Thermokon-SR04 CO2 rH	Indoor air temperature	0 – 51 °C	± 1% of measuring range (typ. at 21°C)
Thermokon – SR -MDS Solar	Motion/ occupancy	0/1	-

Table 2: Monthly mean air temperature in Vienna in °C (GeoSphere Austria 2023)

Month	May	June	July	August	September
Tm	15.0	18.7	22.9	21.3	15.2

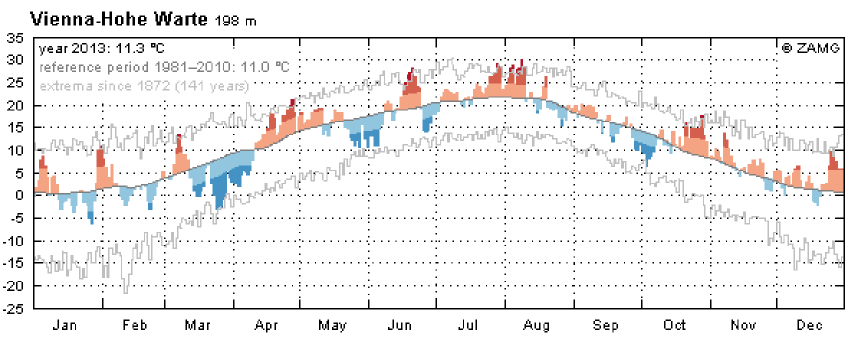

Figure 2: Monthly mean temperature, 2013, Vienna (GeoSphere Austria 2023)

3 RESULTS AND DISCUSSION

As alluded to before, the purpose of the present study is not to either develop or validate a window operation model. Limited monitored data from a single open-plan office and the small number of occupants are certainly insufficient for such a purpose. Rather, the intention is to illustrate the basic features of a process that, given sufficient data and advanced analysis, could lead to more informative and scalable models of occupants' interactions with buildings' control devices and systems. Accordingly, the emphasis is on the prior formulation of relevant hypotheses regarding the background and circumstances that allow for the estimation of the patterns and frequency of occupants' control actions. Accordingly, the observational data is not subjected to a detailed mathematical analysis for explicit model development purposes but applied toward a simple descriptive statistical inquiry. As such, the main findings can be summarized in terms of the information provided in Figure 3. In this Figure, the left-side y-axis marks the window opening probability (P_w) and the x-axis marks the value ranges (bins) of the term $\Delta\theta_{ic}$ (i.e., the difference between indoor temperature θ_i and comfort temperature θ_c). Also shown are the tendencies of $\Delta\theta_{ce}$ (dashed line) and $\Delta\theta_{ie}$ (continuous line) corresponding to the $\Delta\theta_{ic}$ bins of the x-axis. The respective values of these two variables can be obtained from the right-side y-axis.

Consideration of the aforementioned hypotheses H_i to H_{iii} in the light of the data depicted in Figure 3 warrants certain inferences. First, H_i is obviously not supported by observations. Indeed, a trend can be observed, but it is contrary to the one postulated by H_i: It seems the window opening probability is lower for larger deviations of occupants' comfort temperature from indoor temperature arrival times in the office. However, hypotheses H_{ii} and H_{iii} appear to be supported by the observations. The decreasing tendency in window opening probability seems to be consistent with both the outdoor air's diminishing thermal cooling potential as implied by lower $\Delta\theta_{ie}$ values (continuous line in Figure 3) and the lower level of thermal discomfort prior to entering the office space as implied by lower $\Delta\theta_{ce}$ values (dashed line in Figure 3).

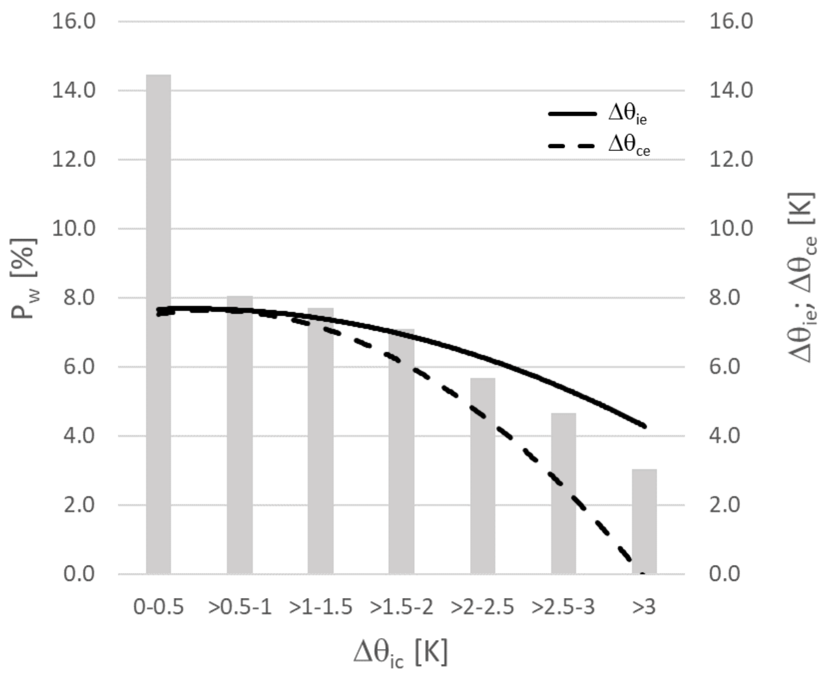

Figure 3: Probability of window opening actions (P_w) as a function of $\Delta\theta_{ic}$. Also shown are the tendencies of $\Delta\theta_{ce}$ and $\Delta\theta_{ie}$ corresponding to the $\Delta\theta_{ic}$ value ranges

These observations underline the initial argument regarding contrasting methodological approaches toward development of behavioural modes. Frequently, such models are based on extraction of patterns from locally and typologically limited data sets. They may be expressed in terms of various types of equations, which map various independent variables onto dependent variables that are meant to capture occupant actions. However, one typically ends up with black-box models, when the model derivation process is not preceded by hypothesised mechanisms mediating between independent and dependent variables. Black-box models can be of course useful in specific scenarios. For instance, a locally calibrated black-box model customised for use in an existing building, with well-documented system details and occupancy patterns, can effectively support building control operations. However, black-box models may be less effective in at least three regards. First, on their own, black-box models do not necessarily shed light on the motivational background and logic of occupants' control-oriented actions. Hence, their role in promoting transparent and knowledge-based design and operation strategies remains limited. Second, they may be less effective as generally deployable prediction models to be integrated, for instance, in building performance simulation applications: Their predictive performance can radically decline without recurrent recalibration to local data. Third, the efforts to further develop and enhance existing black-box models can be hampered by their lack of explicit insights into the intervening mechanisms between independent and dependent variables.

Let us further explain this point. The data underlying Figure 3 could be also analysed in a "theory-free" manner to arrive at a function for the computation of P_w. Equation 1 below provides an illustrative instance of such a function emerging from this data. Using this function, the window opening probability (P_w) can be estimated based on indoor air temperature at arrival time (θ_i) and the preceding mean long-term outdoor temperature ($\theta_{e,l}$), whereby a, b, c, d, and e represent coefficients with values that are adjusted to the data set:

$$P_w = a.\theta_i^2 + b.\theta_i + c.\theta_{e,l}^2 + d.\theta_{e,l} + e.\theta_i.\theta_{e,l} + f \quad (1)$$

Using such a function, one can perhaps obtain reasonably good predictions of the window operation frequency for our case study building and for the same set of occupants. It is also conceivable that others could use this general formalism and adapt the coefficients to match monitoring data from their own buildings. Yet it is not clear to which extent such a formalism could provide any essential insights regarding the logic of occupant behaviour, nor is it clear how the approach can be helpful in cases where local observational data is not available, which de facto includes the bulk of building design scenarios.

To put things in perspective, in our illustrative case study, scanning the data through the filter of the hypotheses does not confirm the conjecture (H_i): The magnitude of perceived thermal discomfort upon arrival cannot, on its own, explain the observed window operation tendency. As to the other conjectures, they are not rejected by the analysis: Both higher cooling potential of the outdoor air (H_{ii}) and the pre-arrival experience of thermal discomfort (H_{iii}) could have encouraged window opening behaviour immediately after entering the office space. However, given the simplistic nature of the analysis, it is important to emphasize that these two conjectures cannot be suggested to have been proven. But the study does suggest that they are worthy of further pursuit based on a richer set of data and more robust statistical methods of analysis.

4 CONCLUSIONS

Occupants' operation of windows can influence indoor-environmental conditions (temperature, humidity, and air quality) and hence occupants' comfort. It can also influence buildings' energy performance. Understanding and predicting window operation behaviour and the availability of related prediction models can thus inform the design and operation of high-performance buildings. In the present study, we contrasted two general approaches to developing such models. Put simply, one approach starts by formulation of prior explicit conjectures and examines those on the basis of available observational data, whereas the other approach starts by a theory-free pattern search in available observational data. We suggested that the former approach has certain advantages in that it can *i)* provide insights into the underlying logic of behavioural patterns, *ii)* support the development of more generally applicable models, *iii)* facilitate the successive improvement of existing models.

Even though the study was of illustrative character and was not meant to rigorously prove anything, it would be prudent to mention various simplifications and limitations it involved. As already mentioned, the observational data, collected over a period of six months, was limited to one space and a small number of occupants. Only thermal factors were considered. Even though both CO_2 concentration and indoor humidity were monitored, they were not used for hypothesis formulation and testing. It is worth mentioning though, that CO_2 concentration, even if it would be considered as a proper proxy of air quality, would have not yielded a proper predictor variable in this study, as its concentration was generally below 600 ppm. Likewise, given the location of the selected office space, which faced a rather quiet internal courtyard, outdoor noise did not represent a potentially relevant influencing parameter. More importantly, the window opening probabilities were aggregated over multiple occupants, thus disregarding inter-individual differences regarding age, gender, health, personal preferences, and habits.

Depending on the relevant application scenarios, the predictive models of occupant behaviour can be of course developed at various level of resolution and detail, that is from simple rule-based formulations to complex stochastic algorithms (Mahdavi and Tahmasebi 2016). Nonetheless, the main contention of the present contribution arguably applies irrespective of the selected level of resolution: More insights can be gained, and more scalable models can be developed, if researchers approach observational data with prior explicit hypotheses, use the verdict emerging out of data analysis to confirm or falsify those, and thus advance the state of their knowledge in this critical domain of inquiry.

5 ACKNOWLEDGEMENTS

In developing the ideas underlying the present paper, the authors benefited from discussions in the framework of IEA EBC Annex 79 – Occupant-Centric Building Design and Operation.
Ardeshir Mahdavi is supported by the FWF (Austrian Science Fund: "Der Wissenschaftsfonds") Project MuDoCo (Project I 5993).

6 REFERENCES

D'Oca, S., & Hong, T. (2014). A data-mining approach to discover patterns of window opening and closing behavior in offices. *Building and Environment*, *82*, 726–739. https://doi.org/10.1016/j.buildenv.2014.10.021

Etheridge, D. (2011). *Natural Ventilation of Buildings: Theory, Measurement and Design*. Wiley.

GeoSphere Austria (2023). *Monthly Climate Report*. https://www.zamg.ac.at/cms/en/climate/climate-overview/current_climate/monthly_climate/klimawerte?monat=09&jahr=2013. Accessed 20.06.2023

Mahdavi, A., Berger, C., Tahmasebi, F., & Schuss, M. (2019). Monitored data on occupants' presence and actions in an office building. *Scientific Data*, *6*(1), 290. https://doi.org/10.1038/s41597-019-0271-7

Mahdavi, A., Del Bolgia, M., Tahmasebi, F., & Schuss, M. (2016). Prediction of user-driven window operation in buildings. *Indoor Air 2016 - The 14TH International Conference of Indoor Air Quality and Climate*, Paper-Nr. 750.

Mahdavi, A., & Tahmasebi, F. (2016). The deployment-dependence of occupancy-related models in building performance simulation. *Energy and Buildings*, *117*, 313–320. https://doi.org/10.1016/j.enbuild.2015.09.065

Nicol, F., Humphreys, M., & Roaf, S. (2012). *Adaptive Thermal Comfort: Principles and Practice*. Routledge.

Tahmasebi, F., & Mahdavi, A. (2016). An inquiry into the reliability of window operation models in building performance simulation. *Building and Environment*, *105*, 343–357. https://doi.org/10.1016/j.buildenv.2016.06.013

Tahmasebi, F., & Mahdavi, A. (2019). Revisiting the benefits of diversity representation in window operation models for building performance simulation. *Bauphysik*, *41*, 30–37. https://doi.org/10.1002/bapi.201800022