A machine learning approach to enhance indoor thermal comfort in a changing climate

Downloaded from: https://research.chalmers.se, 2026-10-13 00:32 UTC

Citation for the original published paper (version of record):
Kramer, T., Garcia-Hansen, V., Omrani, S. et al (2021). A machine learning approach to enhance indoor thermal comfort in a changing climate. Journal of Physics: Conference Series, 2042(1). http://dx.doi.org/10.1088/1742-6596/2042/1/012070

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

A Machine Learning approach to enhance indoor thermal comfort in a changing climate

Tobias Kramer [1], Veronica Garcia-Hansen [1], Sara Omrani [1], Vahid M. Nik [2,3], Dong Chen [4]

[1] School of Architecture & Built Environment, Queensland University of Technology, Brisbane, Australia
[2] Division of Building Physics, Department of Building and Environmental Technology, Lund University, Lund, Sweden.
[3] Division of Building Technology, Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden
[4] CSIRO – Commonwealth Scientific and Industrial Research Organisation, Melbourne, Australia

t.kramer@qut.edu.au, v.garciahansen@qut.edu.au, s.omrani@qut.edu.au, vahid.nik@byggtek.lth.se, dong.chen@csiro.au

Abstract. This paper presents an alternative workflow for thermal comfort prediction. By using the leverage of Data Science & AI in combination with the power of computational design, the proposed methodology exploits the extensive comfort data provided by the ASHRAE Global Thermal Comfort Database II to generate more customised comfort prediction models. These models consider additional, often significant input parameters like location and specific building characteristics. Results from an early case study indicate that such an approach has the potential for more accurate comfort predictions that eventually lead to more efficient and comfortable buildings.

1. Introduction

Due to the built environment's considerable contribution to the emission of harmful greenhouse gases globally, energy-efficiency is one of the most rigorously pursued goals of climate change mitigation within the sector. As a false estimate of the occupants' thermal comfort perception can be a major driver of excessive energy use in buildings, part of the mitigation strategy is the further refinement of conventional thermal comfort models. Currently, by specifying too narrow indoor temperature ranges, the conventional PMV model [1] is still more likely to be associated with promoting energy use in buildings [2]. Thus, the enhancement of these models is could significantly increase energy savings as well as provide higher thermal comfort for occupants [3] in the future.

Recent studies highlight the inaccuracy of conventional comfort models [4] and propose the inclusion of more, often significant variables. In addition to the well-known thermal comfort parameters like air temperature, operative temperature, or relative humidity, several studies analysing data from the recently published ASHRAE Global Thermal Comfort Database II (ACD) [5] found that additional parameters like building characteristics, geographical factors, and individual factors influence the thermal sensation of building occupants. Wang et al. [6] as well as Zhang and de Dear [7] observed that building characteristics like HVAC operation mode and building typology have a significant influence on the

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

occupant's thermal sensitivity. In another study, where researchers analysed the discrepancy between predicted PMV and actual thermal sensation vote they observed an influence of individual factors like age and gender [8].

The technological advances of recent years have encouraged the improvement of conventional workflows and further insights from already existing datasets [9, 10]. For example, the availability of comfort data in large quantities, like in [5], together with the easier and more comprehensible access to new technologies such as Artificial Intelligence is facilitating the development of more advanced comfort models and workflows. For instance, findings from several studies like [10] highlight the feasibility to use the emerging technologies of AI & Machine Learning algorithms for comfort prediction.

The purpose of this paper is to propose an alternative thermal comfort assessment workflow, that is based on the combination of high-quality real-world data, the use of AI technology and computational design. The aim is to compare the prediction performance of the customised, Machine-Learning-based comfort model to the conventional PMV approach. This is accomplished by investigating simulated indoor climate data from an exemplary air-conditioned office building in Sydney.

2. Methodology

The gathering of comprehensive and open-source datasets like in [5], which contains around 100.000 thermal comfort data samples, enables learning from existing buildings by using accessible occupant feedback data to design more efficient and comfortable new buildings. The objective of the framework introduced in this paper is to propose an alternative workflow for comfort prediction practices that utilises the extensive comfort datasets openly available. The aim is to integrate emerging technologies like Machine Learning (ML) and use their leverage to further develop comfort modelling practices in the early building design process to literally "learn" from existing buildings when designing new ones. In general, the workflow provides an alternative to the common practise of using thermal simulation output as an immediate input for either PMV or the Adaptive Comfort Model.

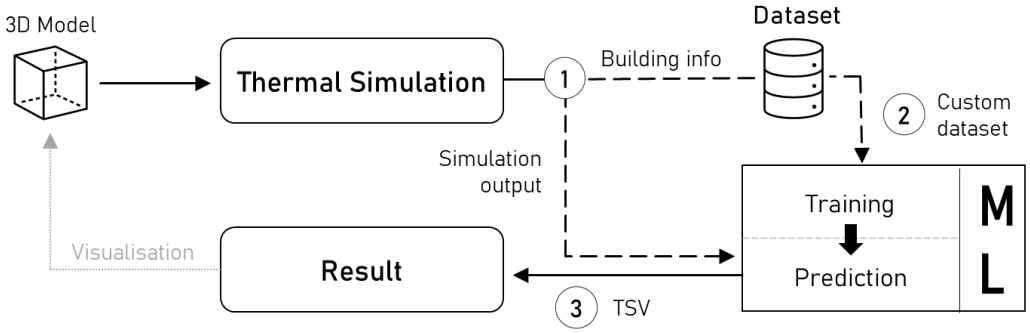


Figure 1. Simplified workflow diagram.

The proposed workflow involves three stages: first, a thermal simulation of the investigated building. Simulated indoor climate output as well as additional building information like location, building use and conditioning strategy are also generated. Secondly, these building information parameters are used to refine the existing dataset of occupant feedback and indoor climate data to only relevant data samples which match certain characteristics of the investigated building. Finally, the customised dataset is used to train a ML algorithm to create an advanced and building-specific comfort prediction model. The previously simulated indoor climate data is used as an input for the customized ML model to make a comfort prediction and visualise the result within the modelling environment. A simplified diagram of the workflow is provided in Figure 1 above.

2.1. Thermal simulation

The first step of the workflow involves the thermal simulation of the building model to generate simulated indoor climate data that is later used as the main input for the ML comfort model. In this

study, the thermal simulation is performed using Ladybug Tools (LT). LT is an open-source computational design tool for Grasshopper that allows to easily customise workflows by Python-scripting and enables to efficiently interface EnergyPlus-based thermal building simulation with up-to-date Machine Learning packages like *scikit-learn* [11]. In addition, to enhance the capabilities of the workflow further building specific information like location, building type and cooling strategy is manually added to the thermal simulation output.

2.2. Model development

For this workflow, the thermal sensation vote (TSV) was selected as the target variable. Concerning the features, theoretically every feature from the ACD can be used to train the ML model. But to attain a dataset of maximum density and to widely exclude samples with missing values, the number of features is limited to the ones listed in Table 1. These features include inputs of the common PMV and Adaptive comfort models as well as additional parameters, such as the building type, its cooling strategy, the geographical location, and the prospective mean age of occupants. These features were selected based on previous analysis of the ACD data. In this analysis it was observed that whereas the building type is somehow considered when choosing between PMV and Adaptive model, those conventional models do not consider the building use. Analysis of the available data clearly indicated that occupants perceive their thermal environment differently, depending on the use of the building, e.g., office, educational or residential. The influence of the building location on thermal sensation could also be identified, for example occupants in warmer climates tend to prefer higher indoor temperature ranges. To compensate for differences in perceived thermal comfort between different age groups, the mean occupant age was also considered as an additional parameter. Furthermore, to allow the use of any customised dataset derived from the ACD, the whole ML process was summarised into Python modules for pre-processing, model training and target prediction.

2.2.1. Pre-processing. Using the building-specific information from the simulation, the dataset is refined to samples that match the main characteristics and geographical factors of the targeted building, data collected in e.g., divergent regions, climates or building types is not considered. Next, a simplified pre-processing routine is done, that includes the removal of samples with missing data and outliers outside the triple interquartile range. For testing purposes, common steps like resampling and advanced feature engineering were mostly omitted in this study, so that the raw and original data is passed on to the model training process. The only exceptions are normalisation and one-hot-encoding of feature values that are numerical or categorical, respectively. The normalisation of numerical features to a range between 0 and 1 and the process of one-hot-encoding categorical values allows the use of a variety of algorithms, and in many cases is indispensable for appropriate generalisation of the model.

2.2.2. Model training & validation. The refined and pre-processed the data is used to train several ML algorithms like Random Forest (RF) or Support Vector Machines and test and validate their respective performance using k-fold cross-validation. At this stage, the basic ML algorithms without further hyperparameter tuning were used to establish a baseline scenario for a later comparison to conventional comfort models.

2.3. Comfort Prediction

In the last stage of the workflow, the trained ML comfort model is used to make a comfort prediction using the previously simulated building indoor climate data and returns predicted TSVs to LT to visualise the results in Grasshopper. The algorithm selection is made based on a comparison of common metrics for regression or classification algorithms. To simplify the process and dynamically link the thermal simulation in LT/Grasshopper with the ML model development in scikit-learn/Python, a custom, Python-based Grasshopper component was created. This component obtains all necessary parameters from the thermal simulation and user input and automatically performs the data refinement

and ML development process to make a TSV prediction, which again can be used, for example, to visualise results in Grasshopper (Figure 2).

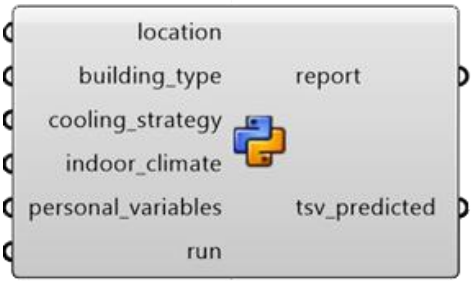


Figure 2. Custom Python component in Grasshopper.

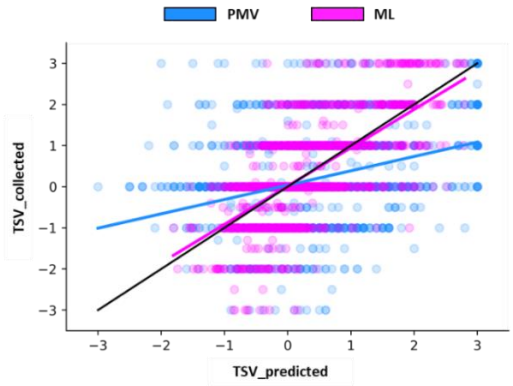
Table 1. ML comfort model input features.

Category	Features
(Indoor-) climate	Air temperature, relative humidity, air velocity, PMV, outdoor temperature (running mean)
Geographical	City, country
Building	Cooling strategy, building type
Personal	Age, met, clo

To fully test the proposed workflow, a simplified case study was conducted. Indoor climate data for one floor of an exemplary air-conditioned office building in Sydney was generated by thermal simulation. Next, instead of following the common approach of using the generalised PMV model for comfort assessment, the proposed workflow featuring a customised ML comfort model was applied. The main objective of the study was to compare the comfort predictions of the different models for a full year and varying indoor conditions, represented mainly by a changing indoor air temperature range. The temperature ranges were defined by a constant heating setpoint of 21˚C and cooling setpoints in range between 22˚C and 25˚C, respectively. This allowed to assess the comfort predictions for varying indoor climates established by the changing setpoints. Additional thermal simulation settings like internal loads or occupancy schedules essentially corresponded to typical settings for office buildings. Sydney was chosen as the location due to its extensive and dense representation in the original dataset.

3. Results

After refining the dataset based on the characteristics of the case study model, for this specific case it was found that the RF regression algorithm showed the best overall performance on the test dataset, which is why the RF regressor was used for the rest of the study. To visualise and compare its performance, the TSV predictions of both the PMV and the ML(RF)-based comfort model on the test dataset are plotted against the actual thermal sensation votes given by occupants in the ACD field studies in Figure 3.



	PMV	ML (RF)
MAE	0.90	**0.54**
MSE	1.34	**0.57**
R^2	-0.21	**0.42**

Figure 3. Scatter plot comparing PMV and ML model predictions with actual TSV collected for field studies in ACD; black diagonal line represents an ideal prediction behaviour or a perfect model.

The ML model regression line fits the ideal prediction scenario better than PMV. Even though the wide scatter of points for both models clearly indicates significant potential for further improvement, quantitative analysis of the data shows much lower mean absolute error (MAE) and mean squared error (MSE) for the ML model. The relatively low R^2 score for the latter can be considered acceptable when analysing field study data involving a human component. However, the negative R^2 score for the PMV indicates a prediction performance lower than taking the mean value.

The ML model was also tested when implemented into the full workflow to compare the comfort predictions with the results obtained by the conventional PMV model approach. Using the simulated indoor climate data, both models were provided with the exact same input data. A summary of the main results is presented below (Figure 4 and 5). To streamline the analysis and compare the predictions in different indoor climate conditions, only the "extreme" scenarios of a low as well as a relatively high cooling set point are shown in the 3D plots of Figure 4. By comparing the plots of PMV and ML predictions for the two temperature ranges displayed, it can be seen that these predictions are in a similar range. However, they differ substantially for certain parts of the year, especially for the winter months (June to August). In addition, the histogram in Figure 5 emphasises these observations by showing clear differences in the distribution of TSV predictions. In the summer (January) the ML model tends to predict a lower TSV than the PMV model, whereas in the winter (July), despite operating in the same temperature range, the opposite is the case.

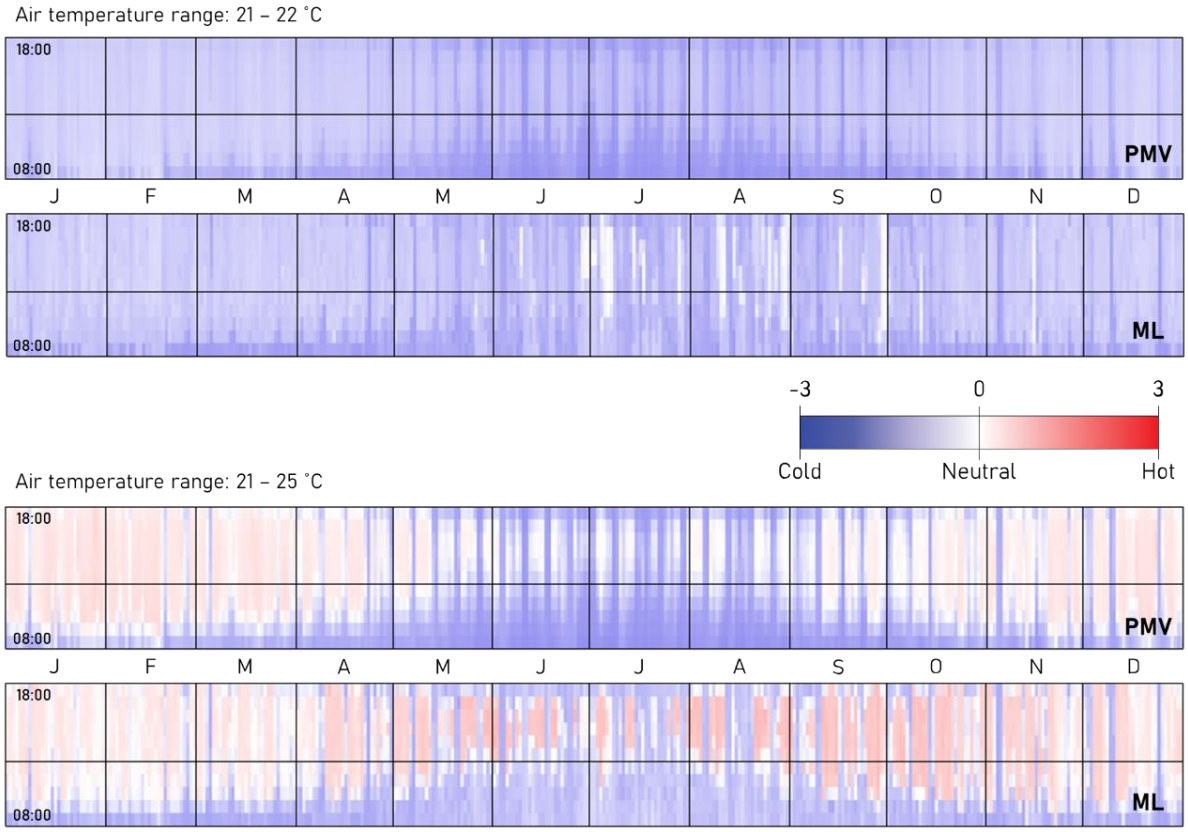


Figure 4. 3D plots for two temperature ranges. Plots visualise the predicted TSV values for the simulated office environment between 08:00h and 18:00h, where every coloured pixel indicates the TSV prediction on the 7-point ASHRAE comfort scale for one specific hour of the year.

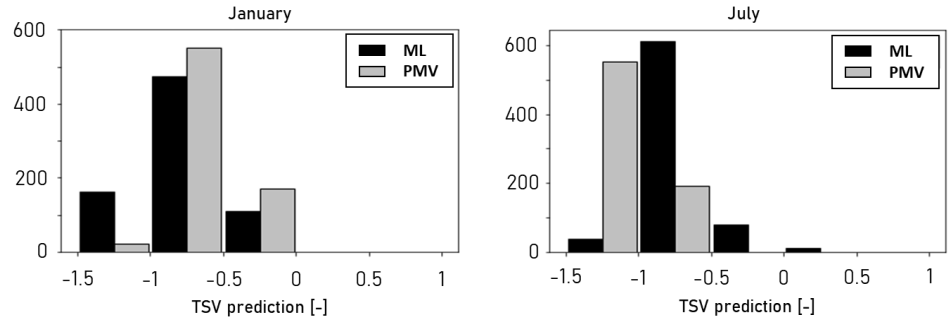


Figure 5. Distribution of TSV predictions of PMV and ML model for the months of January and July and an air temperature range of 21-22°C (08:00h-18:00h).

4. Conclusion

In this paper, computational simulation software was coupled with ML algorithms to generate a customised, building-specific thermal comfort model that is based on real-world data. The presented work shows the potential of using emerging technologies like ML for more advanced comfort predictions that make use of comfort data already available. Dynamically linking thermal simulations with Data Science & AI allows for customised comfort predictions with the capability to improve energy-efficiency as well as thermal comfort of building occupants by enhancing the accuracy of comfort models.

The workflow was tested only for some locations which are quantitatively well represented in the ACD. The study focused on air-conditioned buildings only. Differences between the ML and the Adaptive Model will be assessed in future studies.

The presented workflow is part of a research project investigating the influence of climate change on indoor thermal comfort. Further work will include adding inputs for high-quality future weather data with an increased temporal and spatial resolution and a parameter for long-term human thermal adaptation. To do so, it is intended to collect thermal comfort data in the field.

References

[1] P. O. Fanger, *Thermal comfort: analysis and applications in environmental engineering*. New York, NY, USA: McGraw-Hill, 1972.

[2] T. Parkinson, R. de Dear, and G. Brager, "Nudging the adaptive thermal comfort model," *Energy and Buildings,* vol. 206, p. 109559, 2020/01/01/ 2020.

[3] G. Baird and C. Field, "Thermal comfort conditions in sustainable buildings - Results of a worldwide survey of users' perceptions," *Renewable Energy*, vol. 49, pp. 44-47, 2013.

[4] T. Cheung, S. Schiavon, T. Parkinson, P. Li, and G. S. Brager, "Analysis of the accuracy on PMV – PPD model using the ASHRAE Global Thermal Comfort Database II," *Building and Environment,* vol. 153, pp. 205-217, 2019/04/15/ 2019.

[5] V. Földváry Ličina, T. Cheung, H. Zhang, R. J. de Dear, T. Parkinson, E. Arens, *et al.*, "Development of the ASHRAE Global Thermal Comfort Database II," *Building and Environment,* vol. 142, pp. 502-512, 2018.

[6] Z. Wang, H. Zhang, Y. He, M. Luo, Z. Li, T. Hong, *et al.*, "Revisiting individual and group differences in thermal comfort based on ASHRAE database," *Energy and Buildings,* vol. 219, p. 110017, 2020/07/15/ 2020.

[7] F. Zhang and R. de Dear, "Impacts of demographic, contextual and interaction effects on thermal sensation—Evidence from a global database," *Building and Environment,* vol. 162, p. 106286, 2019/09/01/ 2019.

[8] Y. Li, Y. Rezgui, A. Guerriero, X. Zhang, M. Han, S. Kubicki, *et al.*, "Development of an adaptation table to enhance the accuracy of the predicted mean vote model," *Building and Environment,* vol. 168, p. 106504, 2020/01/15/ 2020.

[9] R. J. de Dear, T. Akimoto, E. Arens, G. S. Brager, C. Candido, K. W. D. Cheong, *et al.*, "Progress in thermal comfort research over the last twenty years," *Indoor Air,* vol. 23, pp. 442-461, 2013.

[10] M. Luo, J. Xie, Y. Yan, Z. Ke, P. Yu, Z. Wang, *et al.*, "Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II," *Energy and Buildings,* vol. 210, 2020.

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, *et al.*, "Scikit-learn: Machine learning in Python," vol. 12, pp. 2825-2830, 2011.