What are the Implications of Building Simulation Algorithm Choice on Indoor Overheating Risk Assessment?

Giorgos Petrou[1], Anna Mavrogianni[2], Phil Symonds[2], Ivan Korolija[2],
Anastasia Mylona[3], Rokia Raslan[2], Dane Virk[4], Mike Davies[2]
[1]UCL Energy Institute, London, United Kingdom
[2]UCL Institute of Environmental Design and Engineering, London, United Kingdom
[3]Chartered Institution of Building Services Engineers, London, United Kingdom
[4]Atkins, London, United Kingdom

Abstract

Building performance simulation tools commonly offer several algorithm options for most heat transfer processes being modelled. The impact of this choice on indoor overheating risk, assessed using the criteria described in the CIBSE Technical Memorandum 59, was quantified for a naturally ventilated dwelling archetype in two popular tools. By selecting non-default algorithm options, the predicted overheating risk changed from high to low for 33% of the cases in tool A and doubled the predicted overheating hours in certain rooms in tool B. Given these findings, modellers should carefully decide on the algorithms being used and publish them for increased transparency.

Introduction

Indoor overheating is defined as the state at which the occupants of a property feel uncomfortably warm due to the indoor environment (CIBSE, 2013) and is a growing concern amongst the building research and industrial communities (Lomas and Porritt, 2017). With the projected increase in the frequency and magnitude of extreme heat episodes (Mora et al., 2017), the ability of dwellings to maintain a thermally comfortable summer environment is becoming a major concern (Mavrogianni et al., 2016). Modellers may commonly use pre-defined criteria, often in the form of thresholds, to decide whether their current design needs to be improved in terms of summer indoor thermal comfort. Technical Memorandum 59 (TM59), recently released by the Chartered Institution of Building Services Engineers (CIBSE), can aid modellers in predicting domestic indoor overheating risk and assessing mitigation options (CIBSE, 2017). This method is based on the use of Building Performance Simulation (BPS) tools.

BPS tools have found widespread applications within the construction industry (Guarino et al., 2016). As designers strive to optimise new and refurbished constructions across a number of objectives (including energy performance and thermal comfort) and achieve compliance with building regulations or planning requirements, BPS tools have become essential components in the industry's building analysis toolbox. Similarly, such tools are now commonly used within academia to assess the importance of various aspects of the indoor environment and identify possible solutions to poor building performance (Symonds et al., 2016; Mavrogianni et al., 2012).

Commonly, BPS tools are based on a series of heat balance equations, where the individual heat and mass transfer processes are approximations of physical laws (Crawley et al., 2008). This is due to the non-linear nature of certain equations, their complex interaction or the difficulty in analytically solving them. Taking the example of the simplest approximation for heat convection on a horizontal surface, the convection coefficient is a non-linear function of the position along the surface (Hens, 2012). When considering different air velocities and orientations, numerical approximations are hence required (Emmel et al., 2007). This results in a single heat process being replaced by a number of empirically based models, generated under specific scenarios, with their applicability depending on the modelled building (see Mirsadeghi et al. (2013) for a comprehensive review of exterior convection coefficient models). In other cases, the modelling options for a heat transfer process could represent different levels of detail which could also impact the model's outputs (Mantesi et al., 2018). Therefore, a modeller should intelligently choose the algorithm combination most appropriate for the building being modelled. The significance of this lies in the possible impact that the choice of algorithms has on the predicted indoor environment. This work aims to quantify this effect by predicting the indoor environment and associated overheating risk in two BPS tools whose algorithmic options will be varied while the building's physical description remains constant. The overheating risk assessment will follow the TM59 methodology (CIBSE, 2017).

A previous study that assessed the overheating risk prediction of nine model variations across two BPS tools established that the choice of software can influence the assessment result (Petrou et al., 2018). A possible consequence could be the intentional use of the tool that overall predicted a lower overheating risk by the modelling community. The importance of algorithm choice has been recognised in the previous work (Mantesi et al., 2018; Prada et al., 2014; Mirsadeghi et al., 2013). To the authors' knowledge, the possible impact of such choice has not been examined in relation to threshold criteria-based assessments, such as those described in TM59.

This paper aims to quantify the effect of BPS algorithm choice on indoor overheating risk prediction for two BPS tools using the TM 59 methodology (CIBSE, 2017). More specifically, the objectives are:

1. to quantify the discrepancies resulting from algorithm choice in the modelled indoor temperatures,
2. to translate these into the TM59 overheating risk metrics, and
3. to determine which modelling input factors are the most influential on the indoor temperature.

Building performance simulation uncertainty

Tian et al. (2018) classified BPS uncertainties into two broad categories: (a) *Model form* and (b) *parameter* uncertainties. Empirical validation work has demonstrated how the combination of both types could lead to significant discrepancies in the model and actual indoor environment (Strachan et al., 2016; Mateus et al., 2014). This work will focus on the magnitude of possible parameter uncertainties associated with the modellers' algorithm choice.

Imam et al. (2017) questioned the modelling literacy of design teams and its potential influence on the performance gap. 108 modellers were asked to rank the importance of 21 BPS input variables relating to the building's physical characteristics, with a quarter of them performing worse than a person choosing at random. Although the approach chosen by Imam et al. (2017) can be disputed, where an incorrect ranking of a list of variables that interact in a complex manner is not evidence of poor modelling capabilities, the question whether poor modelling judgement is widespread within the industry can be raised. The ability of modellers to differentiate and select algorithms could also contribute to the performance gap but was not investigated by Imam et al. (2017).

Mirsadeghi et al. (2013) presented a detailed comparison of the external convection coefficients used within a number of BPS tools and determined that their applicability depends on the modelled building's individual characteristics. For a simple cubic office modelled, deviations of up to 30% in the predicted yearly cooling energy demand compared to the average were observed due to the choice of convection coefficients alone (Mirsadeghi et al., 2013). Prada et al. (2014) compared the heat loss for two methods of modelling conduction. For a gaussian distribution of inputs (thermal conductivity, specific mass, specific heat and wall layer thickness) the maximum discrepancy in the expectation value of heat loss was 5%. However, the differences in the variance of the predicted outputs reached up to 60%. Mantesi et al. (2018) looked at the level of disagreement between two tools when the default and the most similar simulation options were used for each. This was performed for three types of thermal mass. The greatest levels of disagreement for the default options decreased from a normalised root mean square error of 26% for annual heating to less than 3%. It was further discovered that the type of thermal mass significantly influenced the discrepancy between tools. An interesting point raised by Mantesi et al. (2018) is that modellers will tend to rely on the default options of each tool. Based on the discrepancies observed, Mantesi et al. (2018) has called for modellers to make informed decisions on during the model design phase.

Indoor overheating assessment

Buildings can be considered to have a modifying effect on the external environment and may potentially lead to indoor thermal discomfort and poor air quality if not designed appropriately (Taylor et al., 2016). With TM59, CIBSE provides guidance on the assessment of overheating risk of new dwellings during the design phase (CIBSE, 2017). TM59 includes detailed recommendations on modelling input internal loads, occupancy, window and door operation to encourage consistency in industry and research practice of building overheating risk assessment. A high level of overheating risk is predicted if there is failure to meet any of the two following criteria (CIBSE, 2017).

1. Between May to September, the percentage of occupied hours during which $\Delta T = T_{op} - T_{max}$ is greater or equal to 1 °C should be less than 3%.
2. Annually, the threshold of 26 °C should not be exceeded by the bedroom operative temperature between 22:00-07:00 for more than 32 hours.

The operative temperature (T_{op}) is the weighted average of the room's radiant and air temperature, while the T_{max} is the maximum temperature predicted by the model of adaptive thermal comfort (CIBSE, 2013). CIBSE acknowledges that thermal discomfort is subjective. However, given that evidence suggests an increase in extreme heat episodes (Mora et al., 2017), a structured method of comparing overheating interventions could be deemed necessary.

As TM59 does not specify the BPS tool or algorithms to be used, it allows for a useful assessment of the importance of modelling choice. Given the widespread reach of CIBSE, TM59 may be widely adopted within the modelling community and any significant findings could apply to numerous modellers. Furthermore, the clear specification of building input parameters, occupancy-related schedules and a widely available weather file allows for the reproducibility of this work.

Simulation

To determine the importance of algorithm choice on predicted indoor temperatures and the potential exceedance of TM59 thresholds, a free-running and nat-

Table 1: Description of the model's physical properties.

Property	Description
Floor Level	Top floor at a height of 11.2 m.
Orientation	South-facing
Aspect	Single-Aspect
Construction	Lightweight: Timber frame, external brick layer and internal plasterboard.
U-values (W/m^2K)	Wall: 0.17, window: 1.28, floor: 0.18, roof: 0.13, doors: 3.00.
Solar Heat Gain Rate	0.5
Glazing Fraction	0.3
Infiltration	Constant air permeability of $5.0\,m^3/(h\,m^2)$ due to the building envelope, with additional $46.8\,m^3\,h^{-1}$ for the kitchen and $28.8\,m^3\,h^{-1}$ for the bathroom (exhaust fans).

Table 2: Summary of the number of simulation algorithm options assessed for each software. F.D refers to the finite differences method.

	No. of options	
Process	**Tool A**	**Tool B**
Conduction	2	1
F.D. Discretisation	2	1
Exterior Convection	5	2
Interior Convection	4	4
Ext. Longwave Radiation	1	2
Int. Air Emissivity	1	2
Solar Rad. Distribution	5	2
Air Heat Balance	3	1
Total	1200	64

urally ventilated dwelling model was simulated in two BPS tools. The model is based on a top-floor purpose-built flat archetype, representative of a typical London flat in the 1960-1979 age band, designed by Oikonomou et al. (2012). The model's layout is visualised in figure 1. The building's fabric and window thermal properties are summarised in table 1. Room doors were fully open between 08:00-23:00. Windows were fully open between 00:00-24:00 for the bedroom and 08:00-23:00 for the other rooms if the internal air temperature exceeded 22 °C and was lower than the external dry-bulb temperature. The Design Summer Year 1 (2020) weather file, with London Weather Centre location, was used (CIBSE, 2014). The internal gains recommended for a two-person, double bedroom flat by TM59 were used (CIBSE, 2017).

The algorithm options selected are summarised in table 2. A full-factorial analysis was performed where every possible algorithm combination was simulated while the building's physical characteristics remained unchanged. Such method covers the entire input space for the options selected, resulting in a detailed study of interaction effects between different simulation options. Since the options for each algorithm are not ordinal, with non-linear relationships being possible, any form of sampling could have hidden possible interactions. The options were chosen based on the algorithm options identified by the literature or sus-

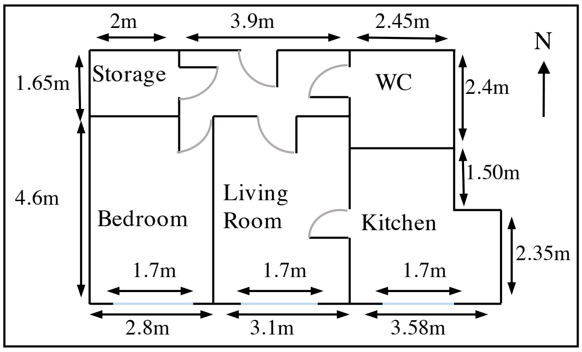

Figure 1: Layout of the model. The model's height is 2.8 m *and the window's height is* 1.6 m.

pected by the authors to be influential on the indoor environment but may be overlooked by modellers.

For this experiment, the simulation options were treated as categorical nominal variables which may be responsible (explanatory) for any observed changes to the response variables. The response variables included the hourly indoor operative temperature, the average daily maximum temperature ($T^{max}_{d,room}$[°C]) across the summer period and the number of predicted overheating hours. $T^{max}_{d,room}$ was estimated for every day (d) over the length of the summer period (D) using the hourly (h) temperature predictions ($T_{i,h}$ [°C]) of each day ($24h$) as:

$$T^{max}_{d,room} = \frac{\sum_{d=i}^{D} max_{h \in 24h}(T_{i,h})}{D} \quad (1)$$

A similar metric has been used previously to indicate a level of overheating risk (Symonds et al., 2016).

Multi-factor Analysis of Variance (ANOVA) was used to statistically determine whether the choice between the different algorithm options (levels) of each heat transfer process (explanatory variable or factor) can impact the indoor environment. The metric of choice was the $T^{max}_{d,room}$ for the bedroom. Although, this is not a metric included within TM59, it was deemed more appropriate to be used for multi-factor ANOVA compared to the number of recorded overheating hours because it is not threshold-dependent.

The null hypothesis explored was that for any heat transfer process, *the mean* $T^{max}_{d,room}$ *across the mod-*

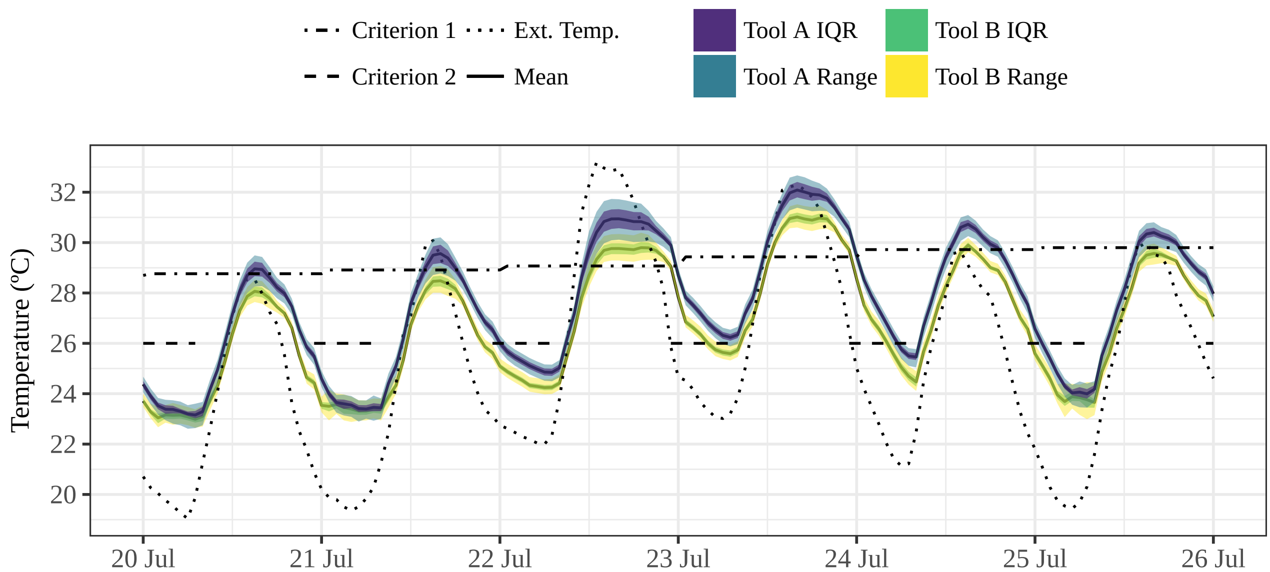

Figure 2: Line chart of the indoor bedroom temperatures predicted by both tools during the hottest period. The shaded areas indicate the range and inter-quartile range (IQR) of the distribution of predicted hourly indoor operative temperatures due to the choice of simulation options.

els with the same option would be equal to the mean $T^{max}_{d,room}$ across all other options of that process at a significance level of 5%. For instance, it is expected that the mean $T^{max}_{d,room}$ predicted by the simulations performed using Tool A and conduction process (a) will be the same as those predicted by the group which used conduction process (b), with a significance level of 5%. Since the choice of algorithms within tools may influence the compliance assessment, the tools and their algorithms will be kept anonymous to prevent the misuse of the results.

Results analysis

The time-series plot in figure 2 displays the mean and spread (due to algorithm differences) of hourly predicted indoor operative temperatures over the hottest six-day period included within the weather file. The maximum hourly spread fluctuates over this period between 0.5–1.7 °C for tool A and 0.4–1.5 °C for tool B. The spread is greatest when the operative temperature plateaus while it diminishes when the rate of temperature change is high. The greatest hourly temperature range predicted by tool A over the summer period was 2.5 °C while for tool B it was 2.1 °C. Similar patterns of indoor temperature were observed for the other rooms as well. The possible implications of such levels of spread on the predicted overheating risk, can be observed within figure 2 where in at least three out of the six days, the Criterion 1 threshold lies within the spread of predicted indoor temperatures for both tools. In such cases it may be possible that the choice of algorithms could influence whether an overheating hour is recorded.

The full impact of this effect on this model over the entire summer period has been quantified and visualised in figures 3 and 4. The distribution of predicted overheating hours are comprised of the individual output of each combination of algorithms. The spread in overheating hours for tool A is visibly greater, especially for the living room with a range of 2.6% compared to 0.7% for the living room in tool B. The red lines shown in figures 3 and 4 represent the level of overheating risk predicted by the default options. For tool A, these predictions are near or above the upper quartile of the simulation results for this tool. Comparing the generated results to the criterion thresholds and by taking the default combination as the reference point, 65% of the simulation combinations would change the end result from a high overheating risk to a low for the living room and 33% for the kitchen. For Criterion 2, as seen in fig. 4, the default options result in a predicted risk near the upper quartile and approximately 10% of the predictions are below the threshold line. For tool B, the entire distribution of overheating risk lies below the threshold. Notably, the default options result in the lowest overheating risk predictions, with all other simulation options predicting an equal or higher overheating risk.

To determine which of the algorithms assessed are influencing the predicted indoor temperatures and overheating risk, an analysis of variance for the main effects was performed with results summarised in table 3. For tool A, the choice of conduction or discretisation appear to have no influence on the average maximum daily temperature. Every other parameter of Tool A has a p-value < 0.001 suggesting a statistically significant result. Therefore, there is enough evidence to suggest that mean average daily

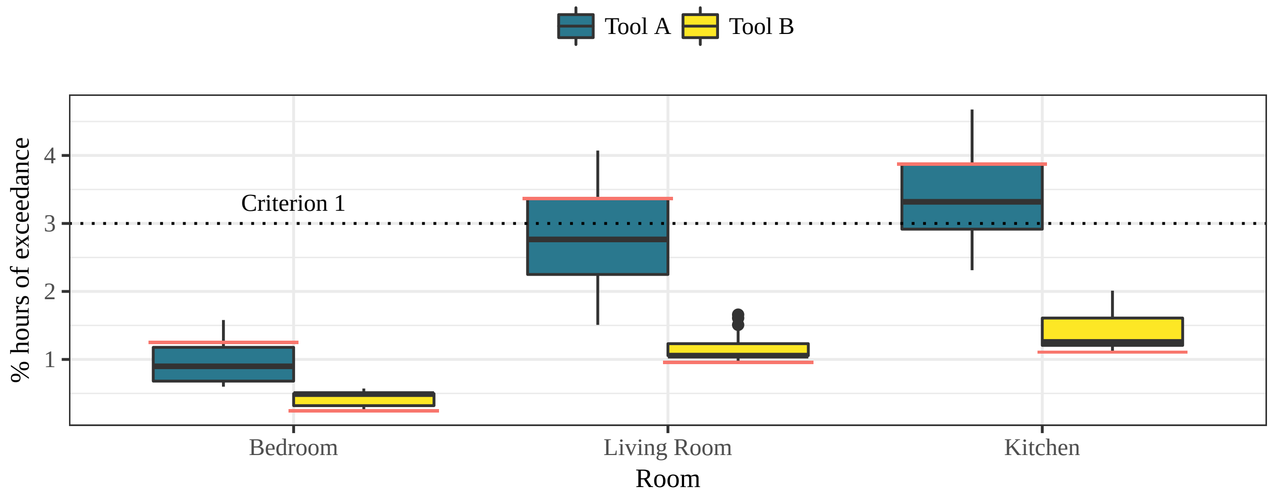

Figure 3: Distributions of the percentage of occupied hours that the CIBSE TM59 Criterion 1 was exceeded. Each box plot is comprised of the individual predictions resulting from every combination of simulation options of each tool (1200 for Tool A, 64 for Tool B). The red lines indicate the prediction of the default options.

maximum temperature of each group differs at the 5% significance level. For tool B, there is not enough evidence to reject the null hypothesis for the exterior convection coefficient or the solar radiation distribution. The contrary is true for the interior convection, exterior longwave radiation and interior air emissivity models, with p-values < 0.001 providing enough support for the rejection of the null. The above statistical analysis has not accounted for any interaction effects. These are visualised for a small set of statistically significant options in figure 5. The interaction between the algorithm options specified by each plot's x-axis and heading results in distributions of predicted $T^{max}_{d,room}$ with differing means. The spread is due to the algorithm options that have not been specified and its non-constant value indicates that heat transfer processes may interact in a complex manner.

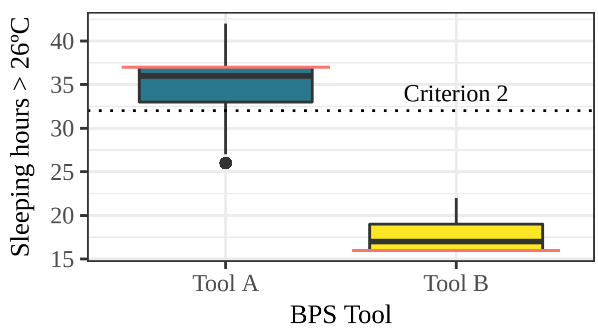

Figure 4: Distributions of the overheating hours predicted by the algorithm combinations of each tool, according to Criterion 2 of CIBSE TM59. The red lines indicate the prediction of the default options.

Discussion

This work investigated whether the choice of algorithms within BPS tools can influence the predicted indoor temperatures and the associated overheating risk, using the threshold method described by CIBSE TM59. Subsequently, the work looked at which of the heat transfer process examined could have a statistically significant impact on the predicted bedroom's mean average daily maximum temperature.

It was determined that the choice of algorithms within BPS tools could appreciably impact the predicted indoor temperatures, as could be hypothesised from previous work (Mirsadeghi et al., 2013). The algorithm combinations resulted in a distribution of predicted hourly indoor temperatures for both tools, whose spread fluctuated and possibly related to the rate of temperature change. Therefore, a change in algorithms is not likely to cause a linear shift in the predicted indoor temperatures but may instead have a dynamic effect. Note that for any single hour, the distribution of predicted indoor temperatures within each tool is not representative of the distribution of indoor temperatures that would be observed if a group of modellers tried to model the pre-defined dwelling using the TM59 criteria[1]. Instead, the distribution is most probably being influenced by (a) the default options (as suggested by Mantesi et al. (2018)), (b) the most accurate choice of options or (c) the options that ensure compliance based on assessment criteria. Here lies the importance of the use of range as a measure of spread. It quantifies the maximum influence on the hourly bedroom temperature that a modeller may have on this specific modelling exercise if they were to consciously or unconsciously alter the choice of algorithms.

The predicted temperature spreads for this model were sufficient to influence the overheating risk as-

[1]This would only be true, if a great number of modellers were to randomly allocate the algorithms within a BPS tool.

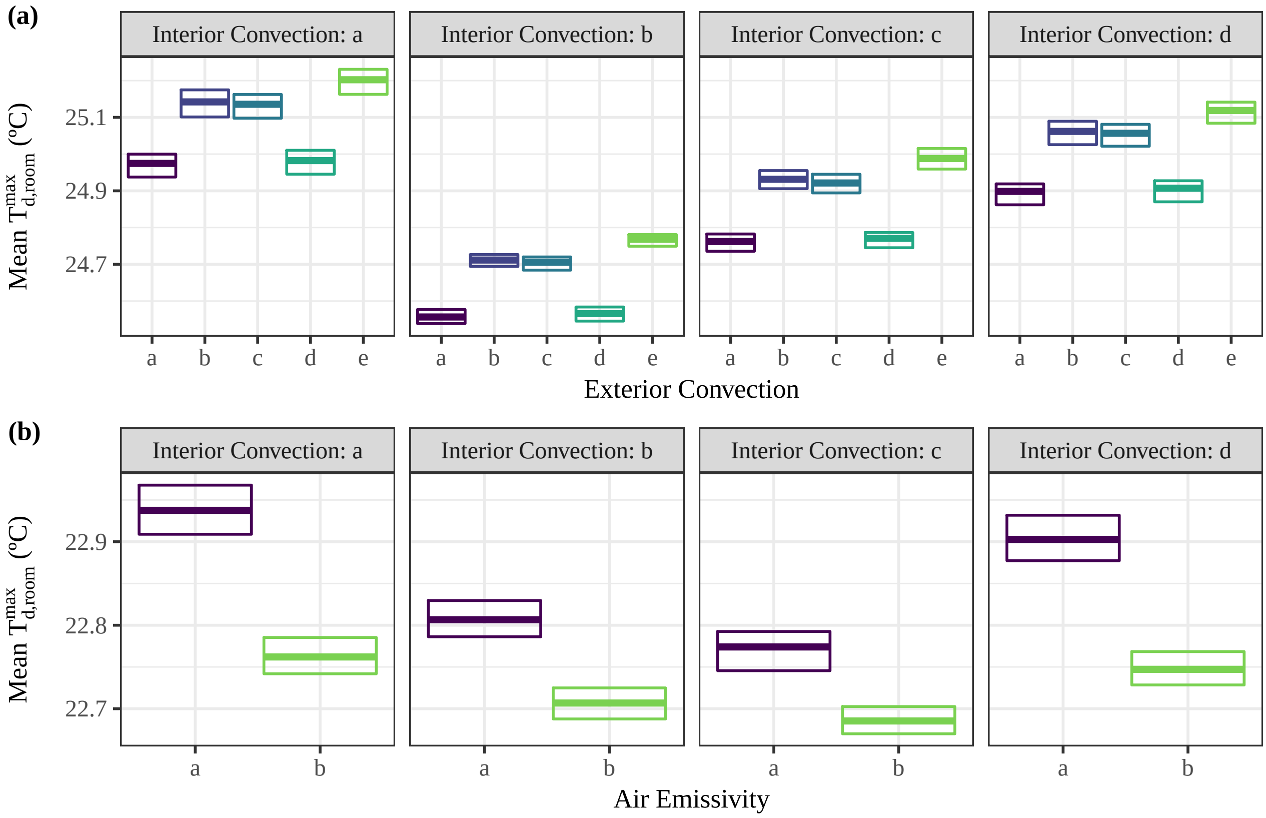

Figure 5: Crossbar plots displaying the interaction effects of the exterior with the interior convection for tool A (part a) and the interior convection with the air emissivity for tool B (part b). Each distribution results from the variation in all the algorithm options besides the two specified by the plot's x-axis value and heading.

sessment in both tools, leading to a distribution of exceedance hours. For tool A, it was possible to change the compliance result without altering the building's physical design. Although this was not the case for tool B, this was an outcome of the overall lower indoor temperatures. Nevertheless, the choice of algorithms led to doubling of the predicted overheating risk for Criterion 1 in tool B. Strikingly, the predictions of default options led to a greater discrepancy between the two tools than most other algorithm combinations, reinforcing the doubts surrounding the frequent use of default options (Mantesi et al., 2018).

Table 3: Statistical results (p-values) of the Analysis of Variance, testing the hypothesis of similar mean $T^{max}_{d,room}$ for each option of a heat transfer process.

	P-values	
Process	**Tool A**	**Tool B**
Conduction	1	-
Discretisation	1	-
Exterior Convection	<0.001	0.104
Interior Convection	<0.001	<0.001
Ext. Longwave Radiation	-	<0.001
Int. Air Emissivity	-	<0.001
Solar Rad. Distribution	<0.001	0.724
Air Heat Balance	<0.001	-

Finally, the statistically significant options were identified and some of their interaction effects were visualised. Given the deterministic nature of BPS tools, the relations discovered can be used to pursue a specific outcome. As a demonstration, choosing option *b* for interior convection in tool A would yield lower temperature than any other option for this model. If this choice is supplemented with option *a* for exterior convection, the predicted temperatures are the lowest that can be achieved due to only adjusting these two algorithms. To instead obtain maximum average daily maximum indoor temperatures over the summer period using tool B, the air emissivity option should be set to *a* whilst using algorithm *a* for interior convection. These relations hold true for the building fabric used for this investigation, however, other interactions may appear for different building designs. (Mantesi et al., 2018; Prada et al., 2014).

Implications

Within the industry, modellers should carefully select the algorithms most suitable for the building being modelled. Conscious or unconscious erroneous[2] choice of algorithms may contribute to the perfor-

[2] Such as the use of a convection algorithm developed in the case of high levels of forced air when this is not true for the building being modelled.

mance gap. Default options should not be considered the optimum choice of each tool and the community could possibly benefit by eliminating them and asking users to choose each option. The developers of such tools are responsible to provide detailed information on the algorithms used within their packages, something which is often not the case. To ensure an adequate level of knowledge within the industry, it may also be valuable to develop BPS training courses on the appropriate use of such tools. In the case that the "correct" choice is not clear, an argument for predicting the worst case scenario can be made in order to minimise the risk on the occupant's health and well-being. As discussed by Raslan and Davies (2010), although there are clear benefits in pass/fail assessments, the variability in outputs between and within tools hinders their usefulness. Therefore, it may be beneficial for assessments to move away from binary results unless the level of such uncertainties can be quantified. Importantly, for increased levels of transparency modellers should publish the algorithms and BPS tools used for any assessment.

Academia could contribute directly to the improvement of BPS tools. Research in empirical validation should clearly state the BPS tool and algorithms used. For any model, different algorithm combinations could be tested to determine how that influences their predictive ability. Academia should also try to refine the current set of algorithms within BPS tools and expand the available options where necessary.

To maximise the utility of BPS tools in informing policy, there is a need for increased accuracy and transparency. All of the aforementioned suggestions could improve the trust on such tools and their outcomes.

Limitations

The use of a single model with constant physical parameters limits the generalisability of the detected effects. As the literature has demonstrated, the choice of algorithms along with the variation in a model's physical design can have a synergistic effect (Mantesi et al., 2018; Prada et al., 2014). Thus, although this work allowed for the discovery of interaction effects, their magnitude in terms of changes to the indoor temperature could vary depending on the physical characteristics of the building.

Finally, an indoor temperature based metric which was not included in TM59 had to be used for the analysis of variance. This is due to the nature of the assessment. Significant changes in the indoor operative temperature (T_{op}) may exist due to the choice of algorithms, however, if they are not near the maximum temperature threshold (T_{max}), they will not be detected. Although ΔT is not threshold-related, the combination of positive and negative values prevented it from being a useful metric for statistical analysis.

Conclusion

This paper investigated the importance of algorithm options within building performance simulation (BPS) tools on predicted indoor temperatures. A purpose-built, naturally ventilated, top-floor flat was modelled in two BPS tools. Within each tool, the algorithm options were varied in a full-factorial analysis resulting in 1,200 simulations for tool A and 64 simulations for tool B. The different algorithm combinations led to a spread in the predicted hourly temperature for both tools, reaching a maximum of 2.5 °C for tool A and 2.1 °C for tool B. These discrepancies were translated into overheating risk prediction, following the guidance in the CIBSE Technical Memorandum 59. The choice of options was shown to impact the number of predicted overheating hours in either tool. The effects were most pronounced for tool A, where the predicted overheating risk changed from high (default options) to low for 33% of the algorithm combinations. It was also discovered that the default options resulted in greater discrepancies between the tools than most other combinations. The individual algorithms which had a statistically significant impact on the indoor temperature were identified and their potential to provide a specific bias on the predictions was demonstrated. In light of the results, the choice of algorithms could also be contributing to the performance gap, especially in the case of threshold-based assessments. Increased transparency through the publishing of tools and algorithms used could be the first action taken against this problem.

Future work

Future work will include a number of physical model variations, exploring the relationship of the observed differences and interaction effects to the model's physical design. In addition, the effect of algorithm selection on other metrics or thresholds will be examined.

Acknowledgement

This research was made possible by support from the EPSRC Centre for Doctoral Training in Energy Demand (LoLo), grant numbers EP/L01517X/1 and EP/H009612/1 and the Chartered Institution of Building Services Engineers. We would also like to thank Ms Eleni Oikonomou for providing the characteristics for a base-case model.

References

CIBSE (2013, July). The limits of thermal comfort: Avoiding overheating in European buildings,TM52: 2013. Technical report, Chartered Institution of Building Services Engineers, London.

CIBSE (2014). Design Summer Years for London - TM49:2015. Technical report, Chartered Institution of Building Services Engineers.

CIBSE (2017). Design methodology for the assessment of overheating risk in homes, TM59: 2017. Technical report, Chartered Institution of Building Services Engineers, London.

Crawley, D. B., J. W. Hand, M. Kummert, and B. T. Griffith (2008). Contrasting the capabilities of building energy performance simulation programs. *Building and Environment 43*(4), 661 – 673.

Emmel, M. G., M. O. Abadie, and N. Mendes (2007, March). New external convective heat transfer coefficient correlations for isolated low-rise buildings. *Energy and Buildings 39*(3), 335–342.

Guarino, F., M. Cellura, S. Longo, T. Gulotta, M. Mistretta, G. Tumminia, M. Ferraro, and V. Antonucci (2016, November). Integration of Building Simulation and Life Cycle Assessment: A TRNSYS Application. *Energy Procedia 101*, 360–367.

Hens, H. S. L. (2012). *Building Physics - Heat, Air and Moisture: Fundamentals and Engineering Methods with Examples and Exercises* (2nd ed.). Ernst & Sohn.

Imam, S., D. A. Coley, and I. Walker (2017, May). The building performance gap: Are modellers literate? *Building Services Engineering Research and Technology 38*(3), 351–375.

Lomas, K. J. and S. M. Porritt (2017, February). Overheating in buildings: Lessons from research. *Building Research & Information 45*(1-2), 1–18.

Mantesi, E., C. J. Hopfe, M. J. Cook, J. Glass, and P. Strachan (2018, March). The modelling gap: Quantifying the discrepancy in the representation of thermal mass in building simulation. *Building and Environment 131*, 74–98.

Mateus, N. M., A. Pinto, and G. C. da Graça (2014, June). Validation of EnergyPlus thermal simulation of a double skin naturally and mechanically ventilated test cell. *Energy and Buildings 75*, 511–522.

Mavrogianni, A., A. Pathan, E. Oikonomou, P. Biddulph, P. Symonds, and M. Davies (2016). Inhabitant actions and summer overheating risk in London dwellings. *Building Research & Information 45*(1-2), 119–142.

Mavrogianni, A., P. Wilkinson, M. Davies, P. Biddulph, and E. Oikonomou (2012, September). Building characteristics as determinants of propensity to high indoor summer temperatures in London dwellings. *Building and Environment 55*, 117–130.

Mirsadeghi, M., D. Cóstola, B. Blocken, and J. L. M. Hensen (2013, July). Review of external convective heat transfer coefficient models in building energy simulation programs: Implementation and uncertainty. *Applied Thermal Engineering 56*(1), 134–151.

Mora, C., B. Dousset, I. R. Caldwell, F. E. Powell, R. C. Geronimo, C. R. Bielecki, C. W. W. Counsell, B. S. Dietrich, E. T. Johnston, L. V. Louis, M. P. Lucas, M. M. McKenzie, A. G. Shea, H. Tseng, T. W. Giambelluca, L. R. Leon, E. Hawkins, and C. Trauernicht (2017, July). Global risk of deadly heat. *Nature Climate Change 7*(7), 501–506.

Oikonomou, E., M. Davies, A. Mavrogianni, P. Biddulph, P. Wilkinson, and M. Kolokotroni (2012). Modelling the relative importance of the urban heat island and the thermal quality of dwellings for overheating in London. *Building and Environment 57*, 223 – 238.

Petrou, G., A. Mavrogianni, P. Symonds, A. Mylona, D. Virk, R. Raslan, and M. Davies (2018). Can the choice of building performance simulation software significantly alter the level of predicted indoor overheating risk? Manuscript submitted for publication.

Prada, A., F. Cappelletti, P. Baggio, and A. Gasparella (2014, March). On the effect of material uncertainties in envelope heat transfer simulations. *Energy and Buildings 71*, 53–60.

Raslan, R. and M. Davies (2010, March). Results variability in accredited building energy performance compliance demonstration software in the UK: An inter-model comparative study. *Journal of Building Performance Simulation 3*(1), 63–85.

Strachan, P., K. Svehla, I. Heusler, and M. Kersken (2016, July). Whole model empirical validation on a full-scale building. *Journal of Building Performance Simulation 9*(4), 331–350.

Symonds, P., J. Taylor, Z. Chalabi, A. Mavrogianni, M. Davies, I. Hamilton, S. Vardoulakis, C. Heaviside, and H. Macintyre (2016). Development of an England-wide indoor overheating and air pollution model using artificial neural networks. *Journal of Building Performance Simulation 9*(6), 606–619.

Taylor, J., M. Davies, A. Mavrogianni, C. Shrubsole, I. Hamilton, P. Das, B. Jones, E. Oikonomou, and P. Biddulph (2016, April). Mapping indoor overheating and air pollution risk modification across Great Britain: A modelling study. *Building and Environment 99*, 1–12.

Tian, W., Y. Heo, P. de Wilde, Z. Li, D. Yan, C. S. Park, X. Feng, and G. Augenbroe (2018, October). A review of uncertainty analysis in building energy assessment. *Renewable and Sustainable Energy Reviews 93*, 285–301.