A Bootstrap Method to Investigate the Variability of Overheating Risk Against the Future Climate Uncertainty in Dwellings

Cheng Cui, Rokia Raslan, Ivan Korolija
Institute for Environmental Design and Engineering,
University College London, London, UK

Abstract

Future overheating risk in dwellings can be potentially mitigated by minimising the variability of overheating hours against uncertainties in future climate via robust optimisation. However, the estimation of this variability value through the utilisation of percentile-based probabilistic weather data has yet to be sufficiently investigated. In this simulation-based study, the bootstrap method is used to quantify the accuracy of the variability estimation via percentile-based weather data. The results indicate significant overheating risk in regulation-compliant houses. An increased degree of difficulty is also suggested in obtaining accurate estimations when considering time periods further in the future and when assuming higher carbon emissions. In addition, the skew normal distribution can be used for a simpler and faster estimation, but the underlying uncertainties must be strengthened throughout its implementation.

Key Innovations

- Quantification of the estimation accuracy of the overheating variability against climate uncertainties under different weather scenarios;
- Proposed use of the skew normal distribution as an alternative for an adequately accurate estimation of overheating variability in the case of limited computational resources.

Practical Implications

The full range of future weather probabilities, rather than the medium one, should be considered in evaluating domestic overheating risk.

Introduction

Summertime overheating in dwellings has been a pressing issue for decades in heating-dominant regions such as the UK, where domestic mechanical cooling is not widely used. This issue has recently further intensified, as a result of both global warming and the unintended consequences of decarbonising the building stock. Links have been forged between overheating and its significant impact on occupant health and well-being (van Hooff et al., 2014; Sheridan and Allen, 2018). As the lifespan of the housing stock is normally considerable, accounting this issue is necessary at the design stage for both new-build and retrofit, in order to maintain expected performance in the long term. To achieve this, building performance simulation (BPS) alongside optimisation can be used to evaluate and mitigate overheating risk under future climate conditions.

Typical BPS processes consist of model parameters and the simulation engine, where weather data is one of the mandatory inputs. Even though BPS has gone through rapid development and increasing implementation over the past few decades, the handling of input uncertainties is still one of the major challenges facing the accurate prediction of building performance (Wang and Zhai, 2016; Hong et al., 2018). There are two common aspects in BPS that lead to this challenge. The first is the lack of knowledge in the quantification of these uncertainties, where many relevant studies have assumed arbitrary probability distributions (e.g. normal, uniform, etc.) for their evaluated inputs. The second concerns the integration of these quantified uncertainties into modelling, as existing modelling tools mainly require deterministic values as inputs. Consequently, these issues are inherited by simulation-based optimisation (Nguyen et al., 2014; Tian et al., 2018).

Issues of weather uncertainties

In assessing future overheating risk, future weather data and its use in BPS-based optimisation play a critical role. Climate uncertainties caused by natural and anthropogenic factors have seen considerable progress in their quantification (IPCC, 2014). The latest advance resulted in BPS-compatible probabilistic weather data within individual climate scenarios (year × emissions), such as the PROMETHEUS dataset (Eames et al., 2011), which is presented in different percentiles (i.e. 10^{th}, 33^{rd}, 50^{th}, 66^{th}, 90^{th}). On the other hand, challenges still remain in applying these probabilistic data in building design optimisation. One viable approach that has seen some initial investigations is robust optimisation. The philosophy of this approach is to reduce the sensitivity of the evaluated system performance against the uncertain boundary condition of interest, when it is infeasible to design a system that caters for all possible conditions (Gorissen et al., 2015). The capability of being insensitive against the uncertainty in boundary conditions (e.g. weather) for the target performance (e.g. thermal comfort) is termed as **robustness**, and this capability can be approached by seeking a certain **measure of robustness** implemented as the optimisation objective (Wright et al., 2016).

In existing efforts to investigate robust optimisation for building design under future climate, two types of measure of robustness can be found. One type involves the worst-case scenario approach (Kotireddy et al., 2017; Moazami et al., 2019), where a set of plausible future weather scenarios were evaluated to identify and mitigate the worst performing amongst them. However, this approach tends to yield what are referred to as 'fat solutions' that can excessively compromise design solution optimality due to high risk aversion (Kall and Wallace, 1994). The other is based on conventional statistical methods such as the standard deviation (Hoes et al., 2011). This approach is promising in avoiding 'fat solutions', but the number of weather scenarios is usually fairly small, and statistical inference on small-size sample is basically unreliable (Button et al., 2013; Wisz et al., 2008).

Furthermore, although it has been over a decade since the publication of UKCP09, very limited research to date has utilised probabilistic weather data to facilitate robust building design against climate uncertainties. Therefore, it still remains unclear how information on robustness can be derived from simulation outputs by applying the percentile-based weather data. Two key obstacles are (a) sample sizes being too small to support reliable conventional statistical inference, and (b) BPS models being too non-linear to enable an appropriate distribution assumption to facilitate parametric statistics.

Bootstrap and its Bayesian variant

The bootstrap, initiated by Efron (1979), is a non-parametric statistical approach to the accuracy quantification for the estimator of interest (e.g. standard deviation), when the population characteristics of the assessed sample data are unknown. It is a useful tool to consult with, on occasions that the size of sample data is relatively small, and its normality is questionable (Dekking et al., 2005). The philosophy of the bootstrap is to perform random sampling with replacement on sample data, to mimic sampling directly from its original population. Let $\mathbf{x} = (x_1, x_2, ..., x_n)$ denote the observed univariate sample, $\hat{\theta}(\mathbf{x})$ denote the estimator of interest, $\mathbf{x}^* = (x_1^*, x_2^*, ..., x_n^*)$ denote a bootstrap sample, the bootstrap procedure can be described as follows:

1. Generate a bootstrap sample $\mathbf{x}^*$ by sampling $\mathbf{x}$ with replacement;
2. Calculate the value of $\hat{\theta}(\mathbf{x}^*)$ to represent $\hat{\theta}(\mathbf{x})$;
3. Repeat the above two steps for B times.

Sampling with replacement means randomly selecting n original data from the observed sample via the Monte Carlo method, for instance, a bootstrap sample $\mathbf{x}^* = (x_1^*, x_2^*, x_3^*)$ from $\mathbf{x} = (x_1, x_2, x_3)$ can be $\mathbf{x}^* = (x_3, x_1, x_1)$. The bootstrap distribution of $\hat{\theta}$ can thus be obtained to investigate the accuracy of the estimator $\hat{\theta}$. It is suggested by Chernick (2008) that B should be at least 1,000 to construct accurate confidence intervals (CIs) of $\hat{\theta}$.

Integrating Bayesian inference, the Bayesian bootstrap is a variant of the aforementioned standard approach, introduced by Rubin (1981). One benefit of implementing the Bayesian bootstrap is its higher tolerance to the sample size (Gu et al., 2008), with an additional prior distribution. Its procedure mostly resembles that of the standard bootstrap, differing in the resampling step. Sampling with replacement only allows the original data to be selected, whilst a probability vector $\mathbf{g} = (g_1, g_2, ..., g_n)$ is assigned as weights to these data instead in the Bayesian bootstrap, where $\sum_{i=1}^{n} g_i = 1$. The flat Dirichlet distribution is usually used as the prior distribution, which can generate a given number (n) of uniformly random numbers between zero and one that sum up to one as $\mathbf{g}$. It should be noted that although an additional assumption is made, the flat Dirichlet distribution is a non-informative prior that is fairly weak, and the Bayesian bootstrap is still deemed as a non-parametric statistical approach. Despite of this, an important disadvantage for both the standard and the Bayesian bootstrap is their high computational cost, as a result of the large number of resampling simulation.

In recognition of the knowledge gap in deriving robustness information via the percentile-based probabilistic weather data to facilitate robust building design optimisation, this paper aims to apply the Bayesian bootstrap to investigate the variability of domestic overheating risk against the future climate uncertainty. Specific objectives are:

- to assess overheating risk in regulation-compliant dwellings under different future weather conditions via the PROMETHEUS dataset;
- to investigate the estimation accuracy of the variability of future overheating risk against the weather uncertainty via the Bayesian bootstrap;
- to compare results between bootstrap and conventional estimation methods in exploring a simplified approach with acceptable accuracy and less computational cost.

Methods

A UK dwelling archetype was selected for the case study of this research, using EnergyPlus as the BPS engine. Statistical inference on simulation results was

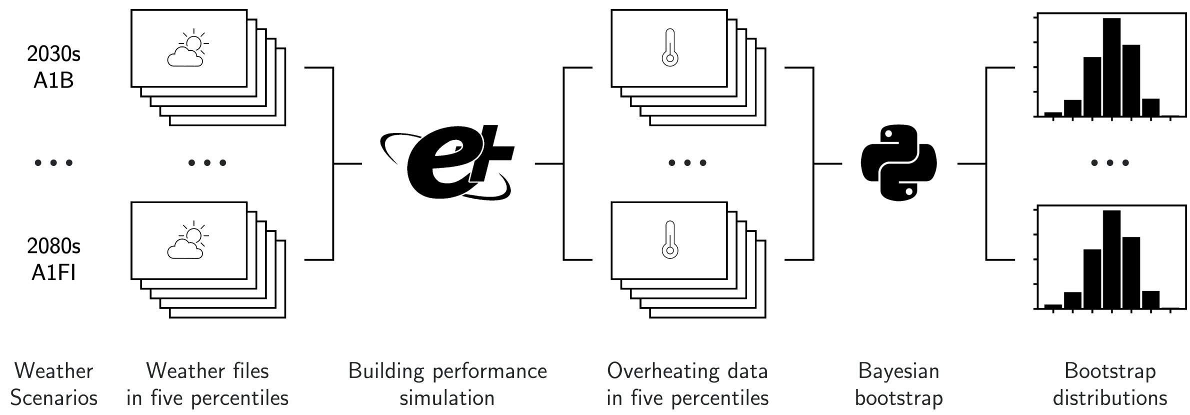

Figure 1: Workflow of research methods.

conducted using Python. For clarity, **variability** is only referred to as the variability of overheating risk against uncertainties in future climate (measure of robustness) throughout the whole paper, the accuracy of variability estimates is described in other terms like dispersion. Figure 1 illustrates the workflow of simulation and data processing, whose details are described as follows.

Models and overheating assessment

A two-storey mid-terraced dwelling archetype, initially developed by Oikonomou et al. (2012), was used for simulation due to its high incidence (14.5%) in the English housing stock (Figure 2). The dwelling layout includes a living room and a kitchen on the ground floor, and three bedrooms on the first.

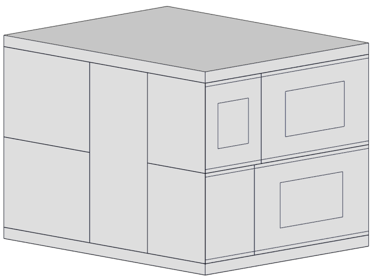

Figure 2: Case study mid-terraced house.

The thermal properties of the building envelope, summarised in Table 1, reflect the performance of a typical retrofitted dwelling of the selected archetype. They were developed in accordance with Approved Document L1B (HM Government, 2018), the regulation for dwelling retrofit in England & Wales. The model was set as free-running during summertime, as the majority of the UK houses currently have no mechanical cooling system.

TM 59 (CIBSE, 2017) was used for the overheating assessment, which provides a set of modelling parameters, internal gains profiles (Table 2) and overheating criteria. The reference occupant heat gain was assumed as 75 W sensible and 55 W latent per person. Windows and doors were modelled as open when internal dry bulb temperature exceeded 22 °C and occupants were present and awake. In this study, the suggested weather files in TM 59 were not followed, and the PROMETHEUS future Design Summer Year (DSY) weather dataset for London (Islington) was instead used. This is available in three time periods (2030s, 2050s, 2080s), two emissions scenarios (A1B, A1FI) and five percentiles (10$^{\text{th}}$, 33$^{\text{rd}}$, 50$^{\text{th}}$, 66$^{\text{th}}$, 90$^{\text{th}}$) (Eames et al., 2011). It should be noted that the UKCP09-based PROMETHEUS dataset was used as DSYs based on the latest UKCP18 were not yet available (at the time of writing). A full-factorial parametric analysis was executed, with weather being the only variable.

Table 1: Envelope U-values.

Envelope	Value ($\mathrm{W\,m^{-2}\,K^{-1}}$)
External wall	0.24
Roof	0.16
Ground floor	0.25
Window	1.58

Bayesian bootstrap and comparison

The implementation of the Bayesian bootstrap followed the aforementioned process, conducted individually under each weather scenario (year × emissions). In this study, the original sample size n is five, representing overheating assessment results varied against weather data in five percentiles $\mathbf{x} = \left(x_{10^{\text{th}}}, x_{33^{\text{rd}}}, x_{50^{\text{th}}}, x_{66^{\text{th}}}, x_{90^{\text{th}}}\right)$ within each weather scenario. The estimator $\hat{\theta}\left(\mathbf{x}\right)$ for the overheating variability was the standard deviation with no correction $\hat{\sigma}_{bootstrap}\left(\mathbf{x}\right)$ defined as follows:

$$\hat{\sigma}_{bootstrap}\left(\mathbf{x}\right) = \sqrt{\sum_{i=1}^{n} g_i (x_i - \hat{\mu}_{bootstrap})^2} \quad (1)$$

Where:

$$\hat{\mu}_{bootstrap}\left(\mathbf{x}\right) = \sum_{i=1}^{n} g_i x_i \quad (2)$$

The repeat number B was set to 1,000,000, considerably greater than 1,000 suggested by Chernick (2008),

Table 2: Internal gains profiles.

Room	Occupancy	Equipment
Kitchen	3 people, 25% gains, 9 am - 10 pm.	300 W, 6 pm - 8 pm; 50 W, 9 am - 6 pm & 8 pm - 9 am.
Living room	3 people, 75% gains, 9 am - 10 pm.	150 W, 6 pm - 10 pm; 60 W, 9 am - 6 pm & 10 pm - 12 am; 35 W, 12 am - 9 am.
Bedroom	2 people, 70% gains, 11 pm - 8 am; 2 people, full gains, 8 am - 9 am & 10 pm - 11 pm; 1 person, full gains, 9 am - 10 pm.	80 W, 8 am - 11 pm; 10 W, 11 pm - 8 am.

so as to guarantee the convergence of bootstrap iterations. The interquartile range (IQR) was used as CIs to denote the acceptable dispersion of the overheating variability estimates.

Comparisons of the bootstrap results were undertaken against estimates using two distribution assumptions and using the direct calculation of the standard deviation with correction. The distributions assessed were skew normal and normal , both applying the maximum likelihood estimation (MLE) method to fit the five data points. The estimator by direct calculation $\hat{\sigma}_{direct}\left(\mathbf{x}\right)$ was defined as follows:

$$\hat{\sigma}_{direct}\left(\mathbf{x}\right) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_{direct})^2} \quad (3)$$

Where:

$$\hat{\mu}_{direct}\left(\mathbf{x}\right) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (4)$$

Results

Overheating risk of regulation-based dwellings

The overheating assessment results for the case study dwelling for two typical weather scenarios are illustrated in Figure 3. In each individual graph, it is evident that an overall upward trend in the percentage of overheating hours can be seen as the percentile increases, in spite of some slight anomalies that are mainly between the 50$^{\text{th}}$ and 66$^{\text{th}}$ ones within the 2030s A1B scenario. Bedrooms and non-bedrooms also show marked differences in the magnitude of overheating risk, which is deemed reasonable as bedrooms are regulated by more strict criteria. The fact that they are on different floors may also contribute to this dissimilarity.

The inter-comparison between the two cases indicates that more distant (future) years and higher emissions can significantly worsen the indoor thermal environment in summer, which is in line with common expectations under climate change. Even more important is the fact that with lower emissions in 2030s, overheating still has a significant opportunity to occur in this regulation-compliant house, suggesting the potential need to revise building regulations in order to better consider thermal comfort.

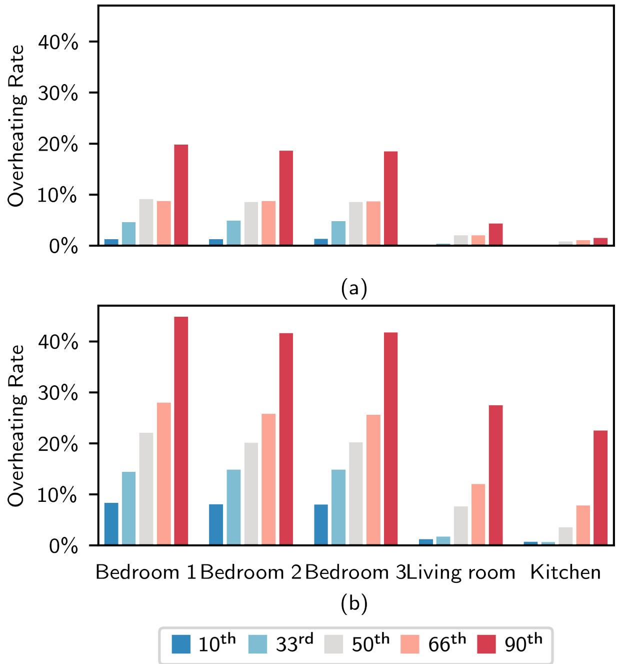

Figure 3: Overheating rate of assessed rooms in (a) 2030s A1B and (b) 2080s A1FI weather scenarios.

Estimation accuracy of overheating variability

Figure 4 is a compilation of results related to the bootstrap estimation of the overheating variability under varied weather scenarios across the five rooms. Both blue and grey histograms represent the same frequency distributions, but depicted in different scales. The blue ones are all plotted in the same scale on their shared bottom x axis, to facilitate their intercomparisons; the grey ones are stretched to varied scales on their individual top x axis, so that their statistical characteristics like IQRs (dark grey) can be more clearly shown. Besides, the three cross marks indicate the variability estimates (top x axis) using the parametric statistics assuming the skew normal (green) and the normal (orange) distribution, and using the direct calculation (red).

Regarding the bootstrap distribution of the variability estimates, dissimilar patterns for different weather-room cases can be clearly seen in blue histograms across the graph matrix. Vertically in weather scenarios, the locations of all bootstrap distributions gradually shift towards right from near to

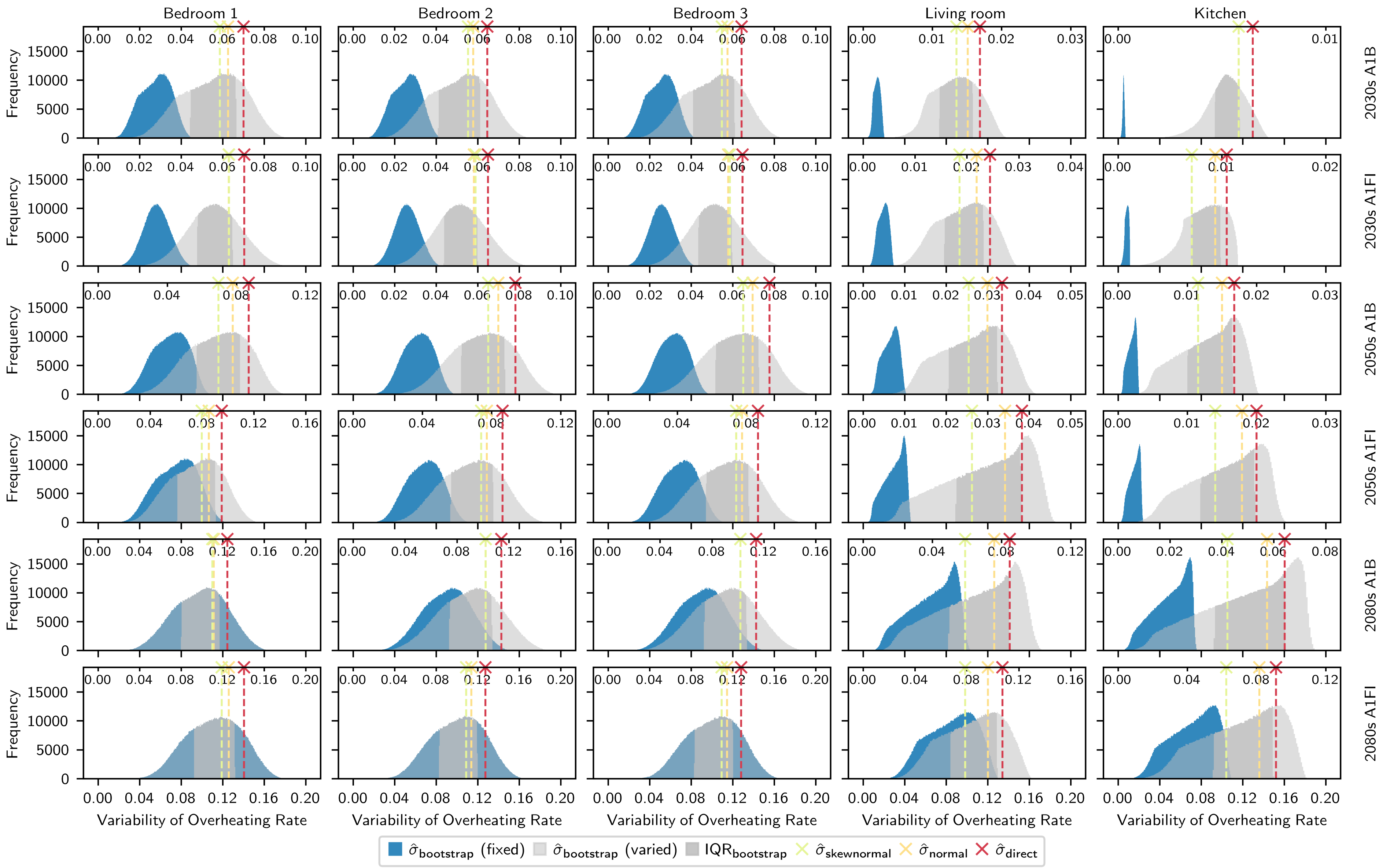

Figure 4: Bootstrap distributions and their CIs with point-wise estimates via three conventional approaches in assessed weather scenarios and rooms (for full explanation see Estimation accuracy of overheating variability).

far future and from low to high emissions. This indicates that the increment of the year and emissions can not only increase overheating risk per se, but also the degree of difficulty to accurately predict it, because of its raised variability within each weather scenario. In addition, the spread of the bootstrap distribution also tends to increase vertically on the graphs (mainly in year variations), further suggesting considerable uncertainties in long-term climate projections and their percentile representatives.

As to horizontal variations between different rooms in Figure 4, remarkable contrasts between bedrooms and non-bedrooms exist in their bootstrap estimates, in line with Figure 3. The bootstrap distribution locations are closer to zero for both non-bedrooms than those for bedrooms coincidently, indicating lower overheating risk. The spreads of their bootstrap distributions are also much less significant, however, this disparity diminishes considerably in later years, probably due to increased uncertainties in long-term climate projections.

In further interpreting these results quantitatively, the definition of statistical significance in this context should be considered, as relevant results are all fairly small in their absolute values. The TM 59 overheating indicator is expressed as percentages, which is a real number between zero and one. Further, the TM 59 threshold of overheating is 1% and 3% for nocturnal and diurnal occupancy respectively. As such, variations larger than 0.005 and 0.015 in variability values of overheating risk can be considered statistically significant, because the standard deviation shares the same unit of measurement as the data whereby it is calculated.

The key benefit of implementing the bootstrap method is its capability of quantifying the accuracy of estimated statistics via CIs. The dark grey areas depict IQRs for corresponding bootstrap distributions, used as CIs in this study. As analysed above, a general trend can be found that the spread of bootstrap distributions rises from upper right cases to lower left ones in Figure 4. The minimum IQR is 0.001 (kitchen, 2030s A1B), while the maximum one is 0.039 (bedroom 1, 2080s A1FI). The dispersion ranges for non-bedrooms are all less than 0.016 in 2030s and 2050s, suggesting that the variability of overheating risk in these cases can be accurately estimated with fairly small errors. In contrast, all bedrooms, whose IQRs are greater than 0.005 in all the weather scenarios, and non-bedrooms in 2080s can hardly be **point-wise** estimated via simulation informed by only the percentile-based weather data.

Comparisons with conventional statistics

Despite of difficulties in their point-wise estimation, these estimates for the overheating variability can still be trusted with CIs, which can also be used to evaluate the validity of other conventional estimation methods. Inferred from the positions of the three vertical lines, the direct calculation method yields estimates that are all outside the bootstrap CIs, which clearly strengthen its inappropriate usage with only five data points. Comparatively, both distribution assumptions can lead to variability results all within the coincident CIs.

To further evaluate their discrepancies from the bootstrap method, the mean of the bootstrap distribution is used as a benchmark, illustrated in Figure 5. The top graph 5-(a) using the raw error values reveals an outstanding positive skewness in assuming a normal distribution or using the direct calculation, whilst a fairly limited one in assuming skew normal distributions. The bottom graph 5-(b) using their moduli indicates that the skew normal distribution also leads to a smaller mean absolute error, with a similar dispersion when compared to the normal one.

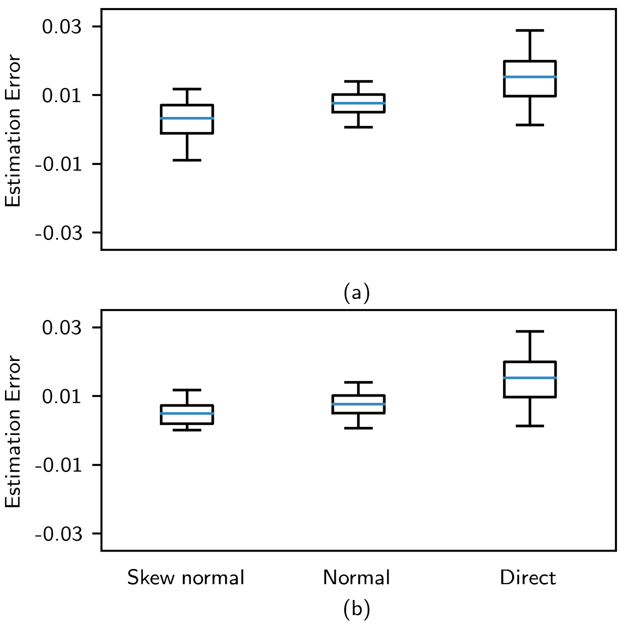

Figure 5: Errors of the variability estimates of overheating risk via other methods against the bootstrap method in (a) raw and (b) absolute values.

Discussion

Applying the variability of overheating risk as the measure of robustness in robust optimisation provides a beneficial means to avoid 'fat solutions', where performance gaps between the average-case scenario and others are minimised, and the worst-case scenarios (which are very unlikely to occur) may be allowed to not strictly meet the overheating criteria. This study overall suggests that the underlying uncertainties in the long-term climate projections significantly influence not only the variability of overheating risk under individual weather scenarios, but also the ability to accurately estimate it via percentile-based weather data. The former supports the importance of adopting techniques, such as robust optimisation, to mitigate future overheating risk due to warming climate, ironically, the latter implies great challenges to achieve that.

Such dilemma can be considered to be a result of the high non-linearity in building energy models. This leads to two important consequences in the context of this study, which are also implied in the results of Kershaw et al. (2011), that (1) percentiles of climate inputs do not necessarily correspond to those of outputs, and (2) a certain skewness may arise during this non-linear physical process. Therefore, it is not reasonable to rely on percentile-based weather data to point-wise infer the variability of the targeted building performance against climate uncertainties, only their value range (max/min) should be trusted. In other words, the performance variables simulated with weather data in five percentiles should only be regarded as five data points with no percentiles assigned (not even strictly monotonic, see Figure 3), which brings difficulties to study their probability distributions and characteristics as demonstrated in detail above. It should also be noted that the bootstrap is no magic to the scarcity of sample data, yet it provides a useful means to quantify the accuracy of the statistical inference on them.

In comparison of overheating risk and its variability between bedrooms and non-bedrooms, their disparity is remarkable despite its reduction over the time, where TM 59 criteria play a crucial role as speculated above. In detail, as aforementioned, non-bedrooms are only required to be assessed for their diurnal use with a 3% threshold, whilst bedrooms are assessed for 24 hours along with a stricter 1% threshold during sleeping. Furthermore, unlike the diurnal assessment with an adaptive thermal comfort model, the nocturnal one for bedrooms applies a fixed temperature limit of 26 °C. Such deterministic threshold is apparently easier to exceed than a fluctuated one. It may be worth separating the overheating assessment for bedrooms by day and night for further investigation of this subject in the future research, whose findings may be able to inform a more robust alternative procedure for overheating assessment.

The implications of the comparison between the bootstrap method and other conventional ones also merit further investigation. Firstly, it can be firmly concluded that the standard deviation with correction is an invalid estimator with only five data points, because its results are outside the CIs in all cases. This corresponds to the findings from several literature that statistics point-wise derived from small-size samples are intrinsically unreliable, due to their high sensitivity to random errors (Button et al., 2013; Wisz et al., 2008). In regard to parametric statistics, the skew normal distribution has a slightly better estimation than the normal one in terms of the error mean and dispersion. However, the skew normal estimator shows a considerably better performance in the error skewness in the case study. This can be plausibly reasoned by the discussion above, that a certain skewness may be introduced in assessing some building performance variables due to the model non-linearity. On occasions where a relatively accurate deterministic estimation for the overheating variability is desired but computational resources are limited, the skew normal distribution can be used as a simple approach to this value. However, the uncertainties of this simple approach should be highlighted throughout its implementation.

Conclusion

This paper initiates the investigation into the estimation of the variability of overheating risk within individual weather scenarios, which is intended to inform better implementation of robust optimisation regarding future heat resilience against climate uncertainties in dwellings. Three levels of values for future domestic thermal comfort were evaluated via the Bayesian bootstrap, namely overheating risk, the variability of overheating risk, the estimation accuracy of the variability. Results suggest that all three values are expected to increase, following the increase of the year and emissions. Considerable dissimilarities exist in different rooms in the case study dwelling, where bedrooms tend to have greater overheating risk in general. Suggestion has also been made that the skew normal distribution can be used to estimate the overheating variability in a simpler and faster fashion, but uncertainties should be acknowledged throughout its implementation.

The present study provides some initial findings as part of a larger doctoral research project, and is intended to better inform future research involving the use of robust building design optimisation in mitigating overheating risk via probabilistic weather data. Ideally, future work should undertake a systematic investigation of the original full weather dataset, whereby the percentile-based one was derived, to further evaluate the implications of this study. The UKCP18-based weather files will be used in future work when available, but it should be highlighted that the use of UKCP09 does not compromise the main findings of this paper, since the method introduced here can be easily implemented with 'newer' climate projections. The validation of simulation results using these future weather files will be investigated in the future. It is also attempting to investigate other dwelling archetypes to evaluate the generalisation of the findings in this paper. Besides, of the two types of measure of robustness in existing literature, this research only discussed the conventional statistics approach, as the worst-case scenario approach cannot be directly compared. Future work is expected to draw this comparison on the level of optimisation results.

References

Button, K. S., J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, and M. R. Munafo (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience 14*(5), 365–376.

Chernick, M. R. (2008). *Bootstrap Methods: A Guide for Practitioners and Researchers* (Second ed.). Hoboken, New Jersey: John Wiley & Sons.

CIBSE (2017). *TM59: Design Methodology for the Assessment of Overheating Risk in Homes.* London, England: The Chartered Institution of Building Services Engineers.

Dekking, F. M., C. Kraaikamp, H. P. Lopuhaa, and L. E. Meester (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How.* London, England: Springer-Verlag London.

Eames, M., T. Kershaw, and D. Coley (2011). On the creation of future probabilistic design weather years from UKCP09. *Building Services Engineering Research and Technology 32*(2), 127–142.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics 7*(1), 1–26.

Gorissen, B. L., I. Yanikoglu, and D. den Hertog (2015). A practical guide to robust optimization. *Omega 53*, 124–137.

Gu, J., S. Ghosal, and A. Roy (2008). Bayesian bootstrap estimation of ROC curve. *Statistics in Medicine 27*(26), 5407–5420.

HM Government (2018). Approved document L1B: Conservation of fuel and power in existing dwellings.

Hoes, P., M. Trcka, J. L. M. Hensen, and B. H. Bonnema (2011). Optimizing building designs using a robustness indicator with respect to user behavior. In *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association*, Sydney, Australia, pp. 1710–1717. IBPSA.

Hong, T., J. Langevin, and K. Sun (2018). Building simulation: Ten challenges. *Building Simulation 11*(5), 871–898.

Intergovernmental Panel on Climate Change (2014). *Climate Change 2014: Synthesis Report*.

Kall, P. and S. W. Wallace (1994). *Stochastic Programming.* Chichester, England: John Wiley & Sons.

Kershaw, T., M. Eames, and D. Coley (2011). Assessing the risk of climate change for buildings: A comparison between multi-year and probabilistic reference year simulations. *Building and Environment 46*(6), 1303–1308.

Kotireddy, R., P.-J. Hoes, and J. L. M. Hensen (2017). Simulation-based comparison of robustness assessment methods to identify robust low-energy building designs. In *Proceedings of the 15th IBPSA Conference*, San Francisco, CA, USA, pp. 892–901. IBPSA.

Moazami, A., S. Carlucci, V. M. Nik, and S. Geving (2019). Towards climate robust buildings: An innovative method for designing buildings with robust energy performance under climate change. *Energy and Buildings 202*, 109378.

Nguyen, A.-T., S. Reiter, and P. Rigo (2014). A review on simulation-based optimization methods applied to building performance analysis. *Applied Energy 113*, 1043–1058.

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.

Rubin, D. B. (1981). The Bayesian bootstrap. *Annals of Statistics 9*(1), 130–134.

Sheridan, S. C. and M. J. Allen (2018). Temporal trends in human vulnerability to excessive heat. *Environmental Research Letters 13*(4), 043001.

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

van Hooff, T., B. Blocken, J. L. M. Hensen, and H. J. P. Timmermans (2014). On the predicted effectiveness of climate adaptation measures for residential buildings. *Building and Environment 82*, 300–316.

Wang, H. and Z. J. Zhai (2016). Advances in building simulation and computational techniques: A review between 1987 and 2014. *Energy and Buildings 128*, 319–335.

Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, A. Guisan, and NCEAS Predicting Species Distributions Working Group (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions 14*(5), 763–773.

Wright, J., E. Nikolaidou, and C. J. Hopfe (2016). Exhaustive search: Does it have a role in explorative design? In *Proceedings of the 3rd IBPSA-England Conference BSO 2016*, Newcastle, UK, pp. 1074. IBPSA-England.