Building and Environment 226 (2022) 109665

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

On the robustness of thermal comfort against uncertain future climate: A Bayesian bootstrap method

Cheng Cui *, Rokia Raslan, Ivan Korolija, Zaid Chalabi

UCL Institute for Environmental Design and Engineering, Central House, 14 Upper Woburn Place, London, WC1H 0NN, United Kingdom

ARTICLE INFO

Keywords:
Robustness
Bootstrap
Climate
Overheating
Uncertainty
Simulation

ABSTRACT

Climate change mitigation and adaptation warrants their synergetic consideration in the building design process, yet past decades have witnessed an unbalanced focus on the mitigation of energy and carbon. In redressing the imbalance, the major challenge lies in the accurate prediction of future building performance via building energy modelling, which is considerably hindered by uncertainties in future climate data. Robustness analysis is a promising technique to inform uncertainty-based decision-making, but its application to future thermal comfort has yet to be sufficiently explored in the built environment. From the perspective of domestic overheating, this paper represents an initial investigation into the implementation of the Bayesian bootstrap method, to quantitatively evaluate the robustness of thermal comfort against uncertain future climate. This is demonstrated using a case study of two typical post-retrofit dwellings in England, where the Bayesian bootstrap also enables the statistical comparison of their expected future overheating risk with climate uncertainty considered. The main findings reveal the magnitude of both overheating risk and its variability experienced during nocturnal occupancy in regulation-compliant dwellings, respectively comprising nearly 15 and 12 times greater than during daytime in extreme cases. Results also imply that adaptive ventilation is potentially the key measure to enhance the robustness of thermal comfort against climate uncertainty. Overall, the Bayesian bootstrap is shown to provide a systematically consistent approach to the robustness assessment of future thermal comfort, which can facilitate the comparability of design alternatives that is vital to the building design decision-making process integrating both mitigation and adaptation strategies.

1. Background

1.1. Climate change adaptation

The latest Intergovernmental Panel on Climate Change (IPCC) Working Group I report has once again highlighted the significant role that human activities have played in warming up the climate system [1]. Progress on emissions mitigation has been driven by national policies and international collaborations [2,3], however, the IPCC has emphasised that more substantial reduction in greenhouse gases is needed to avoid global temperature exceeding the 1.5 °C warming limit.

In addition to targeting ambitious climate change mitigation goals in upcoming years, adaptation is equally important, given the fact that the climate has changed, and its impact already been experienced. Heat exposure in the built environment, especially in homes, has been recognised as one of the highest priorities of climate change adaptation in the UK [4]. Summertime overheating has emerged as a persistent issue in dwellings here, as the region is historically heating-dominant, and its uptake of domestic air conditioning fairly rare. Due to the great amount of time people tend to spend in dwellings, which has been even further increased since the coronavirus disease 2019 (COVID-19) pandemic, this can strongly impact occupant health, wellbeing and productivity [5–7].

Whilst the warming climate has increased overheating risk in terms of both frequency and severity, indoor overheating has also been highlighted as one of the most significant unintended consequences of mitigation actions in the built environment [7,8]. For instance, high levels of insulation and airtightness can substantially decrease heating demand in winters but may exacerbate overheating risk in summers. This can be linked to the limited consideration of the prevention of excess heat in the current building regulations, whose main focus is on energy and carbon reduction (i.e. climate change mitigation) [2,9]. It is also recognised that these regulations are mostly developed based on historical data, whilst specific provisions for thermal comfort needs inspection of future building performance [10]. Given the present need for more rapid and ambitious decarbonisation, it is crucial that actions aiming to simultaneously conserve energy and prevent overheating are implemented, guided by a future-proof upgrade for

* Corresponding author.
E-mail addresses: cheng.cui@ucl.ac.uk (C. Cui), r.raslan@ucl.ac.uk (R. Raslan), i.korolija@ucl.ac.uk (I. Korolija), z.chalabi@ucl.ac.uk (Z. Chalabi).

https://doi.org/10.1016/j.buildenv.2022.109665
Received 9 August 2022; Received in revised form 23 September 2022; Accepted 29 September 2022
Available online 6 October 2022

Nomenclature

Latin Symbols

D	number of previous days
F	population
H_0	null hypothesis
H_1	alternative hypothesis
K	order of a distribution
R	bootstrap replication number
T	temperature
$\hat{a}$	acceleration factor
$\hat{b}$	bias-correction factor
d	time index
$\mathbf{g}$, g	probability vector and its element
$\hat{\mathbf{g}}$, $\hat{g}$	empirical probability vector and its element
$\mathbf{g}^*$, g^*	resampling vector and its element
n	sample size
$\mathbf{p}$, p	support of a distribution and its element
s_x^2	sample variance
t	test statistic
$\mathbf{x}$, x	observed sample and its element
$\mathbf{x}^\dagger$	jackknife sample
$\tilde{\mathbf{x}}$	translated sample
$\bar{x}$	sample mean

Greek Symbols

Δ	paired difference
α	type I error rate
$\boldsymbol{\gamma}$, γ	concentration parameter vector and its element
θ	statistic of interest
$\hat{\theta}$	statistic estimate
$\hat{\theta}^*$	bootstrap replication of the statistic estimate
$\hat{\theta}^{*(\alpha)}$	the 100αth percentile of $\hat{\theta}^*$
$\hat{\theta}^\dagger$	jackknife replication of the statistic estimate
λ	decay factor

Numeric Symbols

$\mathbf{1}$	vector of ones

Functions and Operations

$\cdot$	dot product of two vectors
$^\top$	transpose of a vector
$\Pr(\cdot)$	probability function
$S(\cdot)$	numeric evaluation procedure for a statistic
$T[\cdot]$	discrete temperature time series
$f(\cdot)$	probability density function
$\mathrm{B}(\cdot)$	beta function
$\Phi(\cdot)$	standard normal cumulative distribution function
$\Phi^{-1}(\cdot)$	inverse function of $\Phi(\cdot)$

Subscripts

i	index for the sample size
j	index for the bootstrap replication number
k	index for the order of a distribution
l	index for the number of previous days
ed	external daily mean
lo	lower confidence limit
max	maximum
op	operative
rm	external running mean
up	upper confidence limit

Abbreviations and Acronyms

BC_a	bias-corrected and accelerated
BEM	building energy modelling
CDF	cumulative distribution function
CI	credible/confidence interval
COVID-19	coronavirus disease 2019
DSY	Design Summer Year
IPCC	Intergovernmental Panel on Climate Change
SE	standard error
TM	Technical Memorandum
UHI	urban heat island
UKCP	UK Climate Projections
i.i.d.	independent and identically distributed

building regulations [4,7,11]. The UK government has set out a road map to implement a Future Buildings Standard for new buildings from 2025, and a new regulation guidance on domestic overheating has been introduced in the interim, but many have argued that it should be extended to existing buildings [12,13].

In conducting risk assessment of climate change for buildings to advise adaptation strategies, knowledge of future climate plays a critical role. This is especially true in addressing future overheating risk in both new-build and retrofit dwellings at the design stage, in order to avoid the 'lock-in'[1] risk due to the long lifespan of the housing stock [4,14]. This boundary condition is a major driver of building performance and a mandatory input when applying building energy modelling (BEM) for performance evaluation. Efforts have been made to quantify future climate uncertainties [15,16], and the resultant probabilistic climate data has since been adapted and adopted to investigate future building performance in numerous studies [9,14,17,18]. However, one common limitation amongst these studies is the use of probabilistic weather data in a deterministic fashion, which can be attributed to the difficulty of incorporating input uncertainty within the current BEM framework. Although this deterministic application can provide a glimpse of how buildings may perform under a given future climate scenario, it does not facilitate a robust risk assessment that can adequately inform the formulation of sound adaptation strategies. Therefore, risk assessment that recognises future climate uncertainty at the building design stage is a demanding yet challenging topic that warrants further investigation.

1.2. Robustness analysis

Robustness analysis is a risk assessment technique that is widely recognised in the risk-based decision-making field [19]. In this context, robustness is defined as a property of a given system that describes its capacity for resisting uncertainty-induced disturbance whilst maintaining desired functionality [20]. In the building domain, it can be stated as a process of analysing the sensitivity of some building performance metric against one or more uncertain boundary conditions, in an attempt to reduce such sensitivity so that the operational variation of the targeted metric is manageable [21]. For instance, in mitigating domestic overheating risk under uncertain future climate, robustness

[1] Lock-in refers to practically irreversible damages caused by decisions without considering long-term risks.

analysis can facilitate building design decision-making in such a way that summertime thermal comfort becomes less sensitive to climate uncertainty. Despite this advantage, the application of robust design to a system is usually accompanied by a trade-off between performance optimisation and uncertainty reduction, which is closely related to the selected measure of robustness [22].

Investigation into existing applications of robustness analysis in the building design field has identified two types of robustness measures:

The worst-case scenario: This is the most commonly used measure that discretises uncertainty into multiple scenarios and identifies the one with the worst performance [22]. In studies that implemented this measure, a set of plausible future climate scenarios were evaluated, amongst which the worst performing design candidate was determined [23,24]. This highly risk-averse strategy tends to yield 'fat solutions' that can excessively compromise on the optimality of other decision criteria. Perspective of such over-conservatism is reliant on decision makers' values and judgement [22]. In fields such as structural or aerospace engineering, the pursuit of extreme reduction of uncertainty is justified by the severe consequences of system failures; whilst in regard to energy and thermal comfort in buildings, it can be argued that a certain extent of 'uninsured' performance robustness should be allowed in exchange for better performance optimality.

The statistical dispersion: This measure caters for considering the aforementioned 'uninsured' performance robustness in building design decision-making, which is achieved by gauging the scatter of building performance across the uncertainty range using conventional statistics, such as the variance and the standard deviation [25]. Rather than guaranteeing the performance under the worst-case scenario, this statistical approach aims to lower the performance variability around an 'average' value (e.g. the mean or the median). Whilst this is a promising strategy that circumvents 'fat solutions', proper statistical inference can be obstructed by several factors in real-world problems. Specifically in the case of future climate, the availability of alternative scenarios is usually limited, and small sample inference using conventional statistics can be highly unreliable [26,27].

One method to overcome the small sample issue is to use a parametric model, which entails an assumption of the underlying population distribution for the observed data. In applying this, many researchers have arbitrarily adopted common distributions, such as the normal and the uniform, but very few have validated their choices. It has been recognised that an inappropriate selection of probability distributions can lead to extensive gaps in uncertainty analysis of building performance [28]. A non-parametric statistical approach is thus considered preferable to a parametric one in the absence of strong evidence for confident distribution assumptions.

1.3. Research aim

In fulfilling the need for climate change adaptation in the built environment, potential challenges lie in the robustness assessment of future thermal comfort in dwellings with a small sample of probabilistic future weather data. As such, this study aims to investigate the application of a non-parametric statistical approach, the Bayesian bootstrap, to robustness quantification, using a case study of two typical post-retrofit dwellings in England. Accompanied by analysing the performance robustness of the current building regulations to potentially inform its future-proof revision, the detailed objectives of the present study are:

- to investigate the robustness of future thermal comfort against climate uncertainty using the Bayesian bootstrap;
- to examine the performance of the Bayesian bootstrap by analysing statistical metrics regarding its implementation and application;
- to inspect the implications for the current building regulations based on the findings of the robustness analysis.

2. The Bayesian bootstrap

2.1. Principle

The Bayesian bootstrap [29], as a variant of the frequentist bootstrap[2] initiated by Efron [31], is a non-parametric approach to quantifying uncertainty of statistical estimates, using some measure of accuracy such as the standard error (SE) or the credible/confidence interval (CI). It alleviates the requirement of strong parametric assumptions, when the underlying population distribution of the assessed sample data is unknown and hard to be evidently determined. The philosophy of the bootstrap is to perform resampling on the empirical distribution constructed from the sample, in mimicking sampling directly from its population.

The resampling process for the Bayesian bootstrap can be described using the notion of the resampling vector. If $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is an independent and identically distributed (i.i.d.) sample[3] observed from its population F, $\mathbf{g}^* = (g_1^*, g_2^*, \dots, g_n^*)$ is the resampling vector where $\sum_{i=1}^{n} g_i^* = 1$ and $g_i^* \geq 0$ for $i \in \{1, 2, \dots, n\}$, θ is the statistic of interest, and $\hat{\theta}$ is its estimate given $\mathbf{x}$, the bootstrap replication of $\hat{\theta}$, denoted by $\hat{\theta}^*$, can be obtained by applying some numeric evaluation procedure $S(\cdot)$ that takes $\mathbf{g}^*$ as a variable and $\mathbf{x}$ as a constant,

$$\hat{\theta}^* = S(\mathbf{g}^*). \tag{1}$$

For instance, if $\hat{\theta}$ is the sample mean $\bar{x}$, (1) becomes the weighted mean of $\mathbf{x}$,

$$\hat{\theta}^* = \mathbf{g}^* \cdot \mathbf{x} = \mathbf{g}^* \mathbf{x}^\top = \sum_{i=1}^{n} g_i^* x_i, \tag{2}$$

where $\cdot$ is the dot product of two vectors, the superscript '$\top$' denotes the transpose of a vector, and $\mathbf{g}^*$ can be interpreted as a vector of weights attached to each element in the vector $\mathbf{x}$. Using the replication number R, the bootstrap algorithm can be described as follows.

1. Generate R random resampling vectors $\mathbf{g}_1^*, \mathbf{g}_2^*, \dots, \mathbf{g}_R^*$.
2. Evaluate the corresponding bootstrap replications $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_R^*$ to the resampling vectors.
3. Estimate the selected measure of accuracy for $\hat{\theta}$ based on the bootstrap replications.

The generation of $\mathbf{g}^*$ underpins the implementation of the Bayesian bootstrap. Assuming each observation in $\mathbf{x}$ is distinct for simplicity, $\mathbf{g}^*$ follows a uniform Dirichlet distribution,

$$\mathbf{g}^* \sim \boldsymbol{Dirichlet}(\mathbf{1}), \tag{3}$$

where $\mathbf{1}$ denotes an n-dimensional vector with all elements equal to one. Its derivation using the Bayesian inference, along with the definition of the Dirichlet distribution, is detailed in Appendix A.

An appropriate selection of R also plays an important role. The maximum of R is restricted by the available computational resources, as the ideal bootstrap estimate[4] can be exactly calculated when R approaches infinity. Its minimum, on the other hand, is related to the selection of the accuracy measure. As a general suggestion, an R of 50 is often enough for a good estimate of the SE, whilst at least 1000 bootstrap replications should be obtained to construct accurate CIs [32,33].

[2] Efron introduced this technique using the name 'the bootstrap', which is referred to as 'the frequentist bootstrap' in this study for differentiation, the latter was used in relevant literature, such as Lo [30], for the same purpose.

[3] Samples are all i.i.d. hereafter in this paper, unless otherwise specified.

[4] The ideal bootstrap estimate is a theoretical quantity, whose analytical computation is infeasible for virtually any statistic and measure of accuracy of interest [32].

Table 1
Bootstrap applications in the built environment.

Authors	Variant	Problem	Sample		Replication number	Statistic	Accuracy measure
			Data	Size			
Rastogi and Andersen [34]	Frquentist bootstrap	Weather file synthesis	Detrended climate time series	N/A	10	Spearman's ρ NMSE Degree days	N/A
Tian et al. [35]	Frquentist bootstrap	Sensitivity analysis Data-driven modelling	Simulated input–output pairs	110	500	SRC RMSE r^2	Percentile
Li et al. [36]	Frquentist bootstrap	Building stock modelling	Monitored energy consumption	1144	N/A	NMBE CV(RMSE) Percentage	SE
Rastogi and Andersen [37]	Simple block frquentist bootstrap	Weather file synthesis	Decomposed climate time series	3-day block	100	Mean Median Extremum Pearson's r Spearman's ρ	SE Percentile
Chen et al. [38]	Frquentist bootstrap	Sensitivity analysis	Simulated input–output pairs	100	1000	SRRC	CI
Geraldi et al. [39]	Frquentist bootstrap	Data-driven modelling	Monitored input–output pairs	N/A	N/A	NRMSE	N/A
Chaturvedi and Rajasekar [40]	Frquentist bootstrap	Sensitivity analysis	Simulated input–output pairs	15 000	1000	K–S statistic	CI

CV: coefficient of variance; K–S statistic: Kolmogorov–Smirnov statistic; NMBE: net mean bias error; NMSE: normalised mean square error; NRMSE: normalised root mean square error; Pearson's r: Pearson correlation coefficient; RMSE: root mean squared error; Spearman's ρ: Spearman correlation coefficient; SRC: standardised regression coefficient; SRRC: standardised rank regression coefficient; r^2: coefficient of determination.

2.2. Application in the built environment

Including its Bayesian analogue, the bootstrap has witnessed active development and wide application in multiple domains since its debut, because non-parametric statistics is more robust in solving real-world problems with fewer assumptions required. However, its importance has not been adequately acknowledged in the building sector, whilst parametric models, such as the normal, the log-normal, the uniform and other analytical distributions, were arbitrarily assumed in the majority of uncertainty-related studies [28]. Consequently, a very limited number of applications of such non-parametric tools as the bootstrap can be seen in existing literature, as summarised in Table 1. Some studies were excluded due to the low interpretability of their bootstrap implementation.

Despite varied building-related problems they intended to solve, all the studies applied the frequentist bootstrap, mostly via the standard implementation paradigm as described in Section 2.1 or equivalent. This was either explicitly stated in or deduced from these studies. For instance, Li et al. [36] applied the bootstrap on monitored energy consumption data, to validate their building stock models. Likewise, Tian et al. [35] investigated the accuracy of sensitivity and regression analysis based on a set of BEM input–output pairs. Similar methods have also been adopted in several later studies on sensitivity analysis or data-driven modelling [38–40], amongst which Geraldi et al. [39] used an empirical dataset instead. In these studies, the bootstrap enabled the uncertainty quantification of their sensitivity or error indicators that were conventionally presented in deterministic terms. Its application in building data-driven models also allowed a reduced number of the time-consuming physics-based simulations, which were substituted by resampling.

In addition to the standard paradigm of the bootstrap, several alternatives have been developed over the past decades to accommodate more complicated data structures. One such example is the block bootstrap that deals with time series, where observations cannot be resampled individually due to their dependence (i.e. not i.i.d.) [41]. In creating weather files, Rastogi and Andersen [37] used the simple block bootstrap to resample the noise component of the decomposed weather time series. The stochasticity of the weather conditions was thus able to be captured with limited data availability, and the uncertainty of building performance evaluated accordingly. A follow-up study of theirs adapted the same method to future weather synthesis [42].

One issue arisen from the literature is that the bootstrap may be confused with the Monte Carlo based forward uncertainty propagation concerning multiple variables, as both involve the Monte Carlo method for sampling. Four such cases were found in the built environment literature [43–46]. When multiple variables are involved, each observation in the original sample can be regarded as a vector of values, representing a realised combination of the assessed variables. A key principle in applying the bootstrap to multivariate problems is that such observation should be resampled as a whole, and no new observations are generated during the resampling process. A typical example is the regression analysis by Tian et al. [35]. In contrast, the Monte Carlo based forward uncertainty propagation draws a single realisation from each variable individually based on some theoretical or empirical distribution, followed by combining them as a new observation, and then repeats this process to generate a required number of new observations for further evaluation. Nevertheless, it is worth noting that the inappropriate use of terminology does not necessarily invalidate the methods and the findings in these studies. Excluding them, however, the existing applications of the non-parametric statistics for uncertainty-related analysis are even fewer in the building domain.

3. Methods

Two representative English dwelling archetypes, a mid-terraced house and a top-floor high-rise flat, were selected for the case study in this paper. Fig. 1 illustrates the data flow from the probabilistic weather conditions through the robustness of thermal comfort via BEM and the Bayesian bootstrap. EnergyPlus [47], an open-source building simulation engine, was coupled with Python [48], an open-source programming language, to facilitate parametric modelling and statistical inference. After the development of dwelling models, an exhaustive parametric study was carried out solely on the variations of weather conditions. Within each climate scenario, building simulations were executed, and the bootstrap inference was drawn consecutively. By this means, the mean overheating rate was estimated as the expected overheating risk, and its estimation accuracy evaluated to indicate the robustness of thermal comfort.

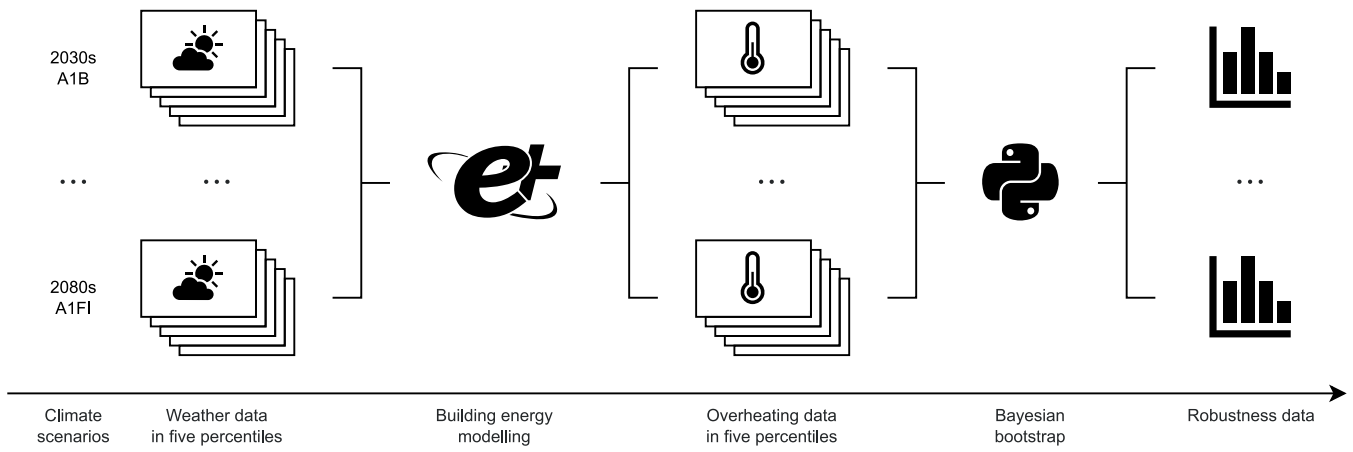

Fig. 1. Data flow diagram of building simulation and statistical inference. Each climate scenario contains weather data in five percentiles, which are individually translated into coincident overheating data via EnergyPlus, followed by robustness evaluation using the Bayesian bootstrap via Python.

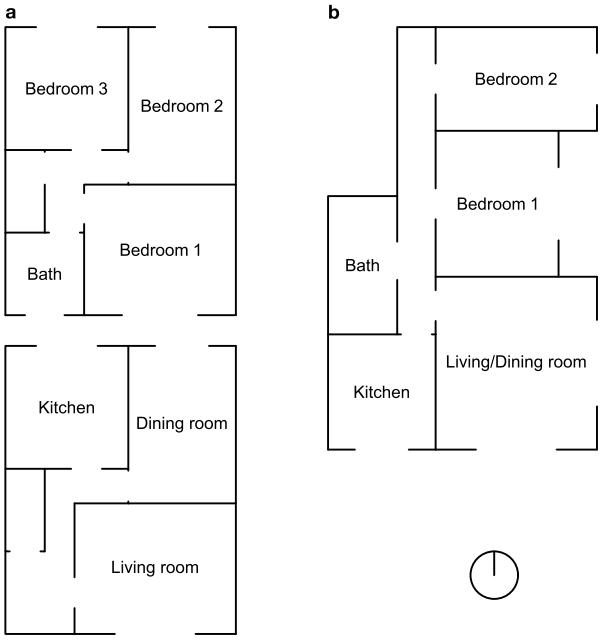

Fig. 2. Floor plans of case study dwellings. Layouts of rooms, their windows and doors in the mid-terraced house (**a**) and the top-floor high-rise flat (**b**).

3.1. Building characteristics and overheating assessment

Initially developed by Oikonomou et al. [49], the two dwelling archetypes are illustrated in Fig. 2 using their floor plan. They were selected considering that the two-storey mid-terraced house has the highest incidence in the English housing stock, and the top-floor high-rise flat is widely recognised with the highest overheating risk. A few modelling assumptions in the original archetypes were adjusted as per the purpose of this study.

To examine the current building regulations, thermal properties of the building envelope were configured accordingly. Specifically, Approved Document L Volume 1 [50], the regulation guidance for dwellings in England, was applied to curate a typical regulation-compliant post-retrofit state for both case study dwellings. Typical constructions were derived from Guide A [51] to meet the regulation requirements, as summarised in Table 2 along with their U-value.

Table 2
Envelope thermal properties.

Envelope	U-value W m^{-2} K^{-1}	Materials
External wall	0.28	Render, 19 mm Concrete block, 200 mm Polyurethane, 80 mm Plasterboard, 12.5 mm
Roof	0.15	Waterproof covering, 2 mm Polyurethane, 150 mm Screed, 75 mm Cast concrete, 150 mm Plasterboard, 12.5 mm
Ground floor	0.25	Cast concrete, 150 mm Extruded polystyrene, 50 mm Screed, 75 mm Vinyl covering, 2 mm
Window	1.32	Low-e glass, 6 mm Argon, 12 mm Clear glass, 6 mm

Since the summertime overheating assessment is the focus of this study, only the summer period that is defined in Technical Memorandum (TM) 59 [52] (i.e. May to September inclusive) was simulated. Both models were set as free-running, as a reflection of limited presence of residential air conditioners in the UK.

TM59 is a widely-used standardised approach to overheating assessment, most of its methodology was followed in this study, with a certain adaptation to better accommodate the research aim. In TM59, overheating is measured during the summer period by a quantitative metric — hours of exceedance, based on the operative temperature, denoted by T_{op}, in each room. T_{op} can be obtained directly from EnergyPlus simulation, which is computed as the average of the air and the radiant temperature. The overheating rate can thus be calculated as

$$\text{Overheating rate} = \frac{\text{Hours of exceedance}}{\text{Hours of occupancy}}, \tag{4}$$

where Hours of occupancy is a fixed and strictly positive integer for a given type of room as elaborated later.

Two criteria of the temperature exceedance and their respective threshold are defined in TM59 for naturally ventilated dwellings, one focuses on occupied hours in general, the other specifically on sleeping hours. As the basis of the first criterion is occupants' 'adaptivity', it seems not reasonable to be used in assessing nocturnal bedrooms, albeit required by TM59; it has also been argued by Mourkos et al. [53] that

Table 3
Internal gain profiles.

Room	Occupancy	Equipment
Kitchen	3 people, 25 %, 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 %, 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 %, 11 pm – 8 am 2 people, 100 %, 8 am – 9 am & 10 pm – 11 pm 1 person, 100 %, 9 am – 10 pm	80 W, 8 am – 11 pm 10 W, 11 pm – 8 am

duplicate evaluation of nocturnal bedrooms is unnecessary. Regarding the second criterion, TM59 intends to apply it in the annual context, but preliminary analysis (not presented in this paper) showed that simulated overheating hours rarely occur outside the summer months; on the other hand, the definition of the summer period may need revision otherwise. As such, both criteria were applied only to the summertime, contingent on the type of the occupied period:

Diurnal occupied period: This criterion applies to all main rooms during non-sleeping hours. An adaptive thermal comfort model, recommended in TM52 [54], is employed to define the limiting maximum acceptable temperature

$$T_{\max} = 0.33T_{\mathrm{rm}} + 21.8, \tag{5}$$

where T_{rm} is the external running mean temperature, whose calculation is specified in Appendix B. It is required that T_{op} shall not exceed $T_{\max}$ by 1 °C for over 3 % of diurnal occupied hours.

Nocturnal occupied period: This criterion applies only to bedrooms during sleeping hours. An absolute threshold advised by Guide A [51] is employed, due to limited opportunities for occupants to 'adapt' at nights. It is required that T_{op} shall not exceed 26 °C for more than 1 % of nocturnal occupied hours.

Apart from the overheating criteria, detailed guidance on modelling assumptions for thermal and internal gain profiles provided by TM59 was also consulted. In the presence of occupants in non-sleeping hours, windows were set operable in case the internal temperature exceeded 22 °C. Interior doors were kept open during the daytime, and otherwise when occupants were sleeping. Natural ventilation was enabled by window operation and infiltration, no mechanical system was included, neither were any dedicated shading devices. Since no residential heating system was modelled in this study, heat loss from pipework was not considered, and overheating in communal corridors was not assessed. As to internal gain, tabulated in Table 3, a metabolic rate consisting of 75 W sensible and 55 W latent per person was used as the baseline, which was reduced by 30 % during sleeping. TM59 assumes constant daytime occupancy in all main rooms to formulate a worse-case scenario, this facilitates the comparability of overheating hours and rates between different rooms and dwellings under the diurnal criterion. Lighting was assumed as 2 W m^{-2} for the period between 6 pm and 11 pm, whilst good daylighting was deemed reasonable during the assessment period.

London was selected as the case study location, where a shift from an oceanic climate (warm summer) at present to a humid subtropical climate (hot summer) later in this century has been projected under the Köppen–Geiger classification [55]. It should be noted that the TM59 guidance on weather files was not followed, substituted by the PROMETHEUS future Design Summer Year (DSY) weather dataset [56] due to more data points available under individual climate scenarios. Specifically, the London (Islington) weather files were used for the case study. This dataset was generated based on the UK Climate Projections (UKCP) 2009, provided in combinations of three time periods (2030s, 2050s, 2080s), two emissions scenarios (A1B, A1FI) and five percentiles (10th, 33rd, 50th, 66th, 90th), where the percentiles were adopted to describe the climate uncertainty under a given climate scenario (year × emissions). UKCP09 has been updated into UKCP18, but DSYs based on the latest climate projections have not yet been available at the time of writing.

3.2. Bootstrap specifications and robustness evaluation

The implementation of the Bayesian bootstrap mainly followed the procedure introduced in Section 2.1, executed individually under each climate scenario.

Specifically in this study, each sample comprised five simulated overheating rates, $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$, coincident with weather data in the five percentiles. The statistic of interest was the sample mean $\bar{x}$, namely the mean overheating rate, whose bootstrap replication $\hat{\theta}^*$ was calculated as per (2), with the resampling vector $\mathbf{g}^*$ generated following (3), and the replication number R set as 1×10^6. The robustness of the thermal comfort was indicated using the $1-2\alpha$ CI of $\hat{\theta}^*$ with $\alpha = 0.05$ (i.e. 90 % credible level). To compute the CI, the bias-corrected and accelerated (BC_a) method was used, as recommended by Efron and Tibshirani [32] for its capability of alleviating impacts from bias and skewness. The BC_a interval can be denoted as

$$[\hat{\theta}^{*(\alpha_{\mathrm{lo}})}, \hat{\theta}^{*(\alpha_{\mathrm{up}})}], \tag{6}$$

where the subscripts 'lo' and 'up' respectively represent the lower and the upper limit of the CI, and $\hat{\theta}^{*(\alpha)}$ denotes the 100αth percentile value of the ordered bootstrap replications. The detailed computation of α_{lo} and α_{up} is given in Appendix C, and an example implementation of the Bayesian bootstrap is provided in Appendix E.

As discussed in Section 2.1, a reasonable R is needed to obtain a good approximation of the ideal bootstrap estimate. Using a considerably larger R of 1×10^6, in comparison with the suggestions from literature, is to guarantee convergence in the bootstrap exercise. This is practical for this study, as such inference is only conducted on two case study dwellings with fixed building characteristics under six climate scenarios. However, it will be significantly more time consuming when applied at scale, such as its potential integration into design optimisation. To both confirm the Monte Carlo convergence with the selected R in this study, and inform future work where a multitude of building design candidates are evaluated, a convergence analysis was also performed. This involved conducting the same statistical procedure using a range of values of R from 50 to 2×10^6, with an increment step of 10 before 1×10^3, 50 before 1×10^4, 100 before 1×10^5, 500 before 1×10^6 and 1000 afterwards, and the evaluation of each was repeated 100 times.

The bootstrap method also enables the execution of a non-parametric two-sample hypothesis test [32,57], facilitated by which a comparison of overheating risk between the two dwelling archetypes was drawn under each climate scenario. As overheating risk was evaluated under coincident weather conditions for both dwellings, a paired two-sample t-test was applied. The null hypothesis H_0 was that the mean overheating rate in the mid-terraced house is equal to or higher than that in the top-floor high-rise flat, whereas the alternative hypothesis H_1 was that the former is lower than the latter. The test statistic is

$$t = \frac{\overline{\Delta x}}{\sqrt{s^2_{\Delta x}/n}}, \tag{7}$$

where $\overline{\Delta x}$ and $s^2_{\Delta x}$ are respectively the sample mean and the sample variance of $\Delta\mathbf{x}$, which is the observed difference between the overheating risk of a given room in the mid-terraced house and that in the top-floor high-rise flat. The calculation of the p-value using the Bayesian bootstrap is presented in Appendix D.

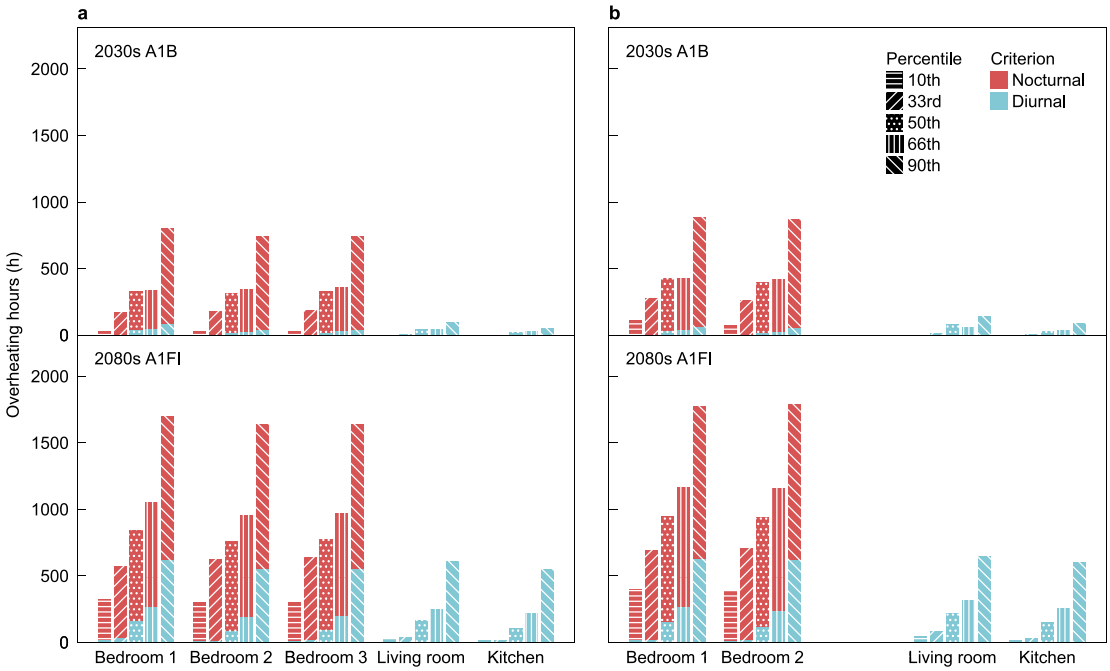

Fig. 3. Breakdown of overheating hours. Simulated overheating hours of the assessed rooms in the mid-terraced house (**a**) and the top-floor high-rise flat (**b**) during nocturnal and diurnal occupancy in the five percentiles under the 2030s A1B (top) and the 2080s A1FI (bottom) climate scenario.

4. Results

4.1. Overheating rate

The simulated overheating hours under two typical climate scenarios, representing a nearer future with lower emissions (2030s A1B) and a more distant future with higher emissions (2080s A1FI), are depicted in Fig. 3, categorised by their diurnal and nocturnal occurrences.

Each individual graph illustrates a clear trend that as the percentile increases, the aggregated overheating hours rise accordingly. A few anomalies exist in the transition between some percentiles, such as those between the 50th and the 66th under the 2030s A1B scenario, where a certain non-monotonicity is shown, plausibly as a collective result of the non-linearity in BEM and the ranking scheme used in weather file synthesis. The latter is solely temperature-based for the PROMETHEUS dataset [56], but domestic overheating is also driven by other climatic variables such as solar and wind. Its impact, along with alternative schemes, was discussed by Liu et al. [58]. In contrast to the similarity in changes over percentiles, a distinct difference regarding the magnitude of the aggregated overheating hours in bedrooms and non-bedrooms (i.e. the living room and the kitchen) can be identified. Judging from the diurnal and nocturnal breakdown, it is evident that this difference mostly stems from the sleeping period, during which overheating can exceed 500 hours even in the mildest situation; yet overheating severity during daytime is quite comparable across rooms. Notably, the bedroom 2 and the bedroom 3 in the mid-terraced house share fairly similar overheating rates under all climate scenarios, this is likely attributed to their same orientation (i.e. north) and room function, which respectively leads to comparable thermal and occupancy profiles. As such, the bedroom 3 is excluded in the following analysis.

Across different climate scenarios for the same dwelling, the comparison indicates a marked increase in summertime overheating hours as the time period progresses from the 2030s to the 2080s and the emissions scenario worsens from A1B to A1FI, which is a common expectation as a result of climate change. However, it is noticeable that even in the 2030s with a relatively optimistic emissions scenario, regulation-compliant dwellings such as both case study homes still have a non-negligible chance of overheating occurring in bedrooms. Non-bedrooms that are moderately thermally comfortable in the near future will also gradually reach an overheating rate that exceeds the 3% threshold towards the 2080s. As previously analysed, the quantity of diurnal overheating hours is nearly identical across different rooms, whilst it is during the nighttime that bedrooms experience outstandingly excessive heat. As indicated in Fig. 3, the nocturnal dominance in overheating declines from the 2030s to the 2080s, and Fig. 4 provides a more holistic view of this decline, using the bedroom 1 as an example. Here, the overheating dominance is expressed in terms of the percentage that overheating during a given occupied period comprises. The heat maps show that the overheating dominance during nights peaks at roughly the 10th percentile of the 2030s in both dwellings, and decreases over years and percentiles, though there seems to be no distinct divergence between different emissions assumptions within the same time period. It is also observable that nocturnal overheating is slightly more dominant in the top-floor flat overall.

To better illustrate the difference in overheating severity between two dwellings, Fig. 5 depicts their coincident overheating rates under the same combinations of climate scenarios, percentiles and rooms. Visually, despite being fairly over the identity line[5], data points for daytime occupancy in all rooms are scattered across both sides, which suggests that their overheating level in both dwellings shares a certain similarity between one another. In contrast, nocturnal overheating is significantly severer in the top-floor flat, with none of the sixty data points below the identity line. This will be further analysed later in a more quantitative manner using the bootstrap hypothesis test.

4.2. Overheating variability

4.2.1. Thermal comfort robustness

Overlaid with the bootstrap distributions, Fig. 6 illustrates the estimated mean values of the overheating rate (i.e. the expected

[5] The identity line in this context is a reference to an identical overheating rate in both dwellings.

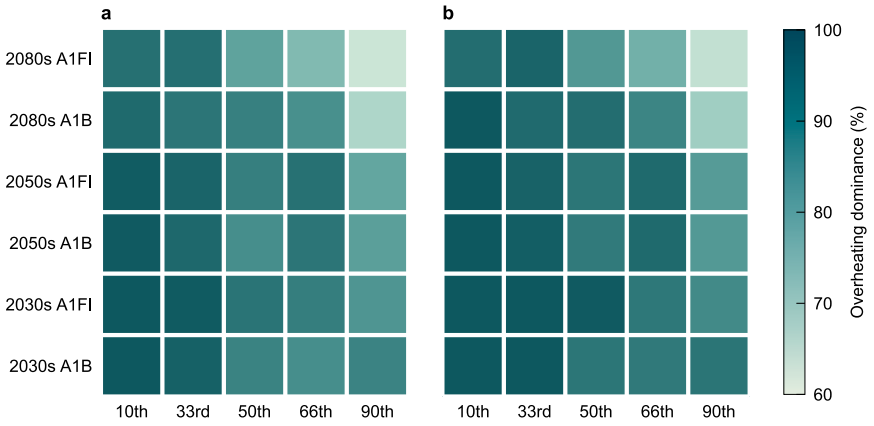

Fig. 4. Dominance of nocturnal overheating. Trends of changes in the percentage that overheating during nocturnal occupancy comprises of the bedroom 1 in the mid-terraced house (**a**) and the top-floor high-rise flat (**b**) over climate scenarios and percentiles.

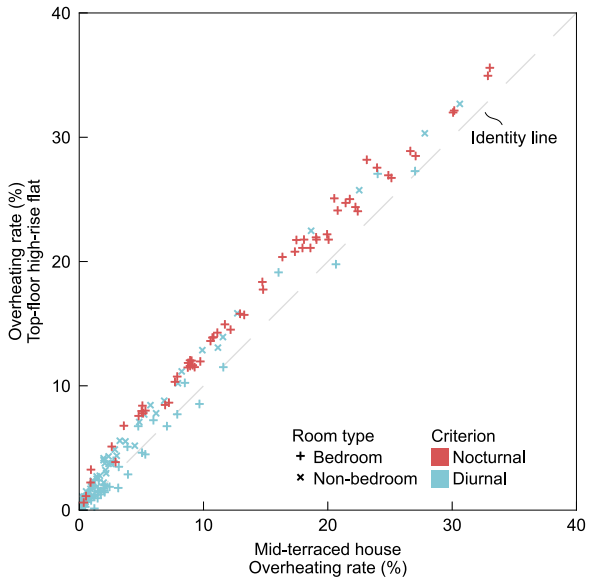

Fig. 5. Comparison of overheating rates. Coincident overheating rates of the same rooms in the mid-terraced house (x axis) against the top-floor high-rise flat (y axis) under the same combinations of climate scenarios and percentiles.

overheating risk) and their 90 % BC_a intervals (i.e. the thermal comfort robustness) for each room and climate scenario in the two case study dwellings. As aforementioned, bedroom evaluation is separated into sleeping and non-sleeping periods, along with the exclusion of the bedroom 3 in the mid-terraced house.

Similar patterns to previous analysis can be observed when inspecting the estimated mean. In contrast to fairly low or even no expected overheating risk during daytime before mid-century, occupants may have already been enduring intense overheating (around 10 %) in recent years when sleeping in regulation-compliant homes. Although the mean overheating rates demonstrate a general upward trend over the years in all rooms and both occupied periods, they appear to be quite stable within the same time period in the 2030s and the 2050s, with an average increase by slightly over 0.3 %. However, the mean estimates are expected to have at least 2.3 % rise from A1B to A1FI in the 2080s, and this figure can reach above 4.5 % in some diurnal occupancy cases. Such difference between the 2080s and the nearer time periods might be related to climate system inertia, where current emissions will not fully reveal its impact until the second half of the century [15]. Comparing the two dwellings, there is also a clear difference in the nocturnal overheating rates, which are 1.9 – 3.5 % higher in the top-floor flat across climate scenarios; but relatively more comparable during daytime.

With regard to the evaluated robustness, a significant distinction is shown between diurnal and nocturnal occupancy. When occupants are awake, relatively narrow BC_a intervals that begin at roughly 0.8 % can be seen in coming decades, but gradual increment over the climate scenarios can lead to a variability of over 16.3 %, which is extremely high compared to the magnitude of the diurnal threshold. This suggests that the thermal comfort during daytime tends to stay fairly robust against climate uncertainty in the short term, whereas such robustness is expected to be significantly weakened over the years. Conversely, the variability of overheating risk in bedrooms during nighttime has been considerably high since the very beginning, but it is notably less prone to be time varying, fluctuating by merely 2.1 %. It is also observable that the CIs at daytime are more extended under higher emissions than at nights, which may imply that the emissions assumption is more impactful on the overheating variability during non-sleeping hours. Besides, overheating variability in the two dwellings is fairly analogous in magnitude during sleeping hours; but it seems comparatively more significant in the top-floor flat during daytime, especially in later years, whose difference can become greater than 2.0 %.

4.2.2. Monte Carlo convergence

To examine the convergence of the bootstrap results in Fig. 6, two examples of the bootstrap cases in this study are selected, namely the mid-terraced living room under the 2030s A1B and the nocturnal high-rise bedroom 2 under the 2080s A1FI. Fig. 7 illustrates the variation of their resultant mean overheating rates and both limits of the 90 % BC_a intervals using different R in multiple runs.

It can be firstly seen that both statistics reach a plateau before 1×10^6, with very limited variation between different runs, advising that it should be reasonable to deem the Monte Carlo errors eliminated in the Bayesian bootstrap exercise of this study. From the other side, the varied fluctuation patterns indicate that the mean estimate is less prone to the stochasticity in the generation of $\mathbf{g}^*$ and converges considerably faster than the robustness indicator; the latter appears not to stabilise until 2×10^4 bootstrap repetitions, whereafter the maximum difference between varied runs remains well below 0.4 %. This is in agreement with the suggestion by Efron and Tibshirani [32] that an R of 1000 is only a start point to construct good CIs. Thus, 2×10^4 may be considered as a balanced R in future work.

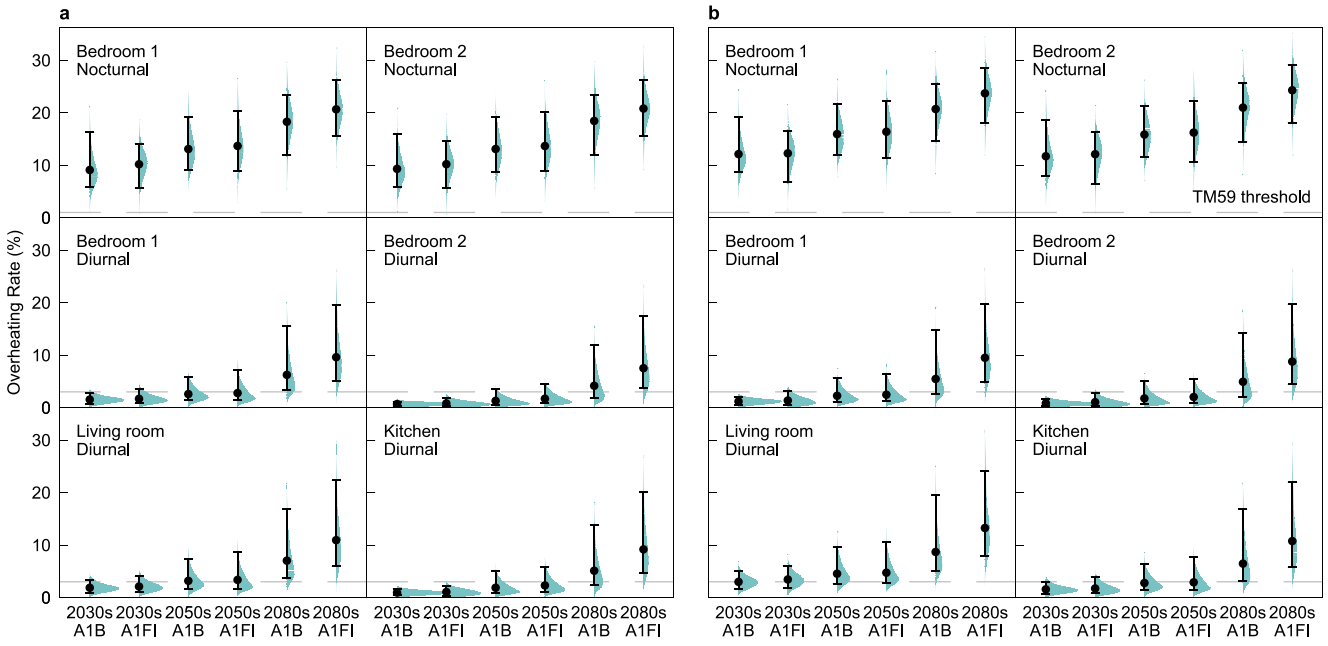

Fig. 6. Variability of expected overheating risk. The bootstrap estimates, the 90 % BC_a intervals and the bootstrap distributions of mean overheating rates of the nocturnal bedroom 1 (top left), the nocturnal bedroom 2 (top right), the diurnal bedroom 1 (centre left), the diurnal bedroom 2 (centre right), the living room (bottom left) and the kitchen (bottom right) in the mid-terraced house (**a**) and the top-floor high-rise flat (**b**) under each climate scenario.

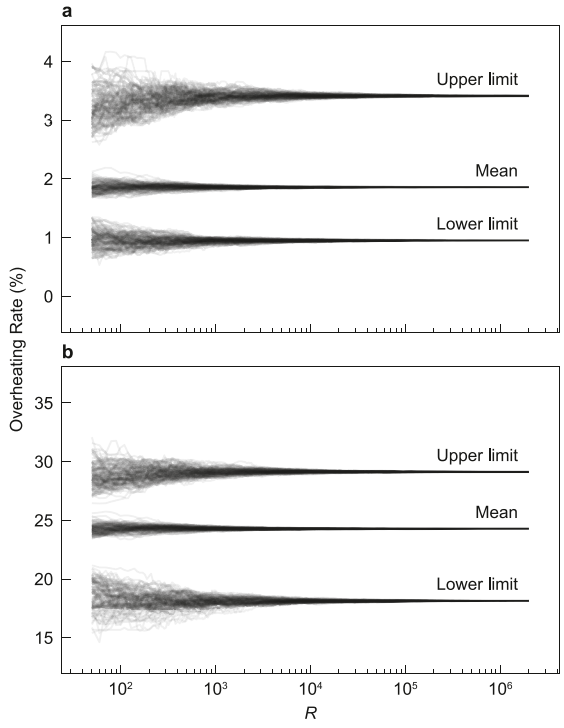

Fig. 7. Convergence of robustness estimation. The bootstrap estimates and the 90 % BC_a intervals of mean overheating rates of the living room in the mid-terraced house under the 2030s A1B climate scenario (**a**) and of the nocturnal bedroom 2 in the high-rise top-floor flat under the 2080s A1FI climate scenario (**b**) with a variety of bootstrap replication numbers of multiple runs.

Table 4
Resultant *p*-values in comparing overheating rates in the two dwellings.

Climate scenario	Nocturnal		Diurnal			
	Bedroom 1	Bedroom 2	Bedroom 1	Bedroom 2	Living room	Kitchen
2030s A1B	< 0.001	< 0.001	0.943	0.260	0.018	0.077
2030s A1FI	< 0.001	0.001	0.914	0.238	0.007	0.100
2050s A1B	< 0.001	< 0.001	0.988	0.169	0.016	0.007
2050s A1FI	< 0.001	< 0.001	0.882	0.147	< 0.001	0.129
2080s A1B	< 0.001	< 0.001	0.999	0.156	0.004	0.051
2080s A1FI	< 0.001	< 0.001	0.800	0.020	< 0.001	0.003

4.2.3. *Dwelling archetype comparison*

In comparing the expected overheating risk between the two dwellings using the bootstrap hypothesis test, the resultant *p*-values are tabulated in Table 4 for different rooms under each climate scenario, categorised by diurnal and nocturnal occupancy. In a hypothesis test, the smaller the *p*-value, the stronger the evidence against H_0. Conventionally, a critical value of 0.1 is considered as borderline evidence, and that of 0.05, reasonably strong evidence [32].

It can be clearly seen that *p*-values for the sleeping period are all less than or equal to 0.001, strongly suggesting that the mid-terraced bedrooms are expected to experience dominantly lower overheating risk than the top-floor ones at future nights. This is in great alignment with the depiction in Fig. 5, where the nocturnal data points are all above the identity line. On the contrary, the bedroom 1 during daytime has a *p*-value fairly close to one under all climate scenarios, which may imply higher diurnal overheating risk in the mid-terraced house, attributed to its south-oriented window as compared to the east-oriented one in the top-floor bedroom 1. It should be emphasised that this implication does not necessarily hold true given the current results, but can be strictly tested (to be either proven or disproven) by modifying H_0 and H_1 accordingly. With less such distinction in thermal profiles, *p*-values for the bedroom 2, despite a trend of decrease over climate scenarios, still provide fairly weak evidence against H_0.

Whilst in the living room and the kitchen, two thirds of the resultant *p*-values demonstrate strong evidence, to varied degrees, for greater overheating rates in the high-rise dwelling, with only one above the borderline critical value. This seems to differ from the findings for diurnal bedrooms, given that they are both occupied during non-sleeping

hours; but they plausibly share the same driver in relation to thermal profiles. Unlike the mid-terraced ones, these non-bedrooms in the top-floor flat have roofs absorbing solar gain during the daytime, which leads to significantly more heat gain indoors.

5. Discussion

5.1. Bootstrap application and probabilistic weather data

Applying the variability of overheating risk as the measure of robustness in decision-making can potentially avoid excessive compromise of building performance optimality for extreme events that are unlikely to occur. Yet its estimation is challenging, which may fundamentally be ascribed to the inherent non-linearity in the modelling of the physical processes in buildings. Percentiles are quite informative as a descriptive statistic, and they are often used as a robust measure of statistical dispersion; as a matter of fact, the BC_a interval applied as the robustness indicator in this study is percentile-based. However, the input uncertainty described by a limited number of percentiles, as is the case of the future weather data, cannot facilitate proper forward uncertainty propagation in non-linear models. That is, the percentiles of climate inputs, even being the sole variation, do not necessarily correspond to those of outputs. This can also be interpreted from an uncertainty analysis study by Kershaw et al. [59] on several other building performance metrics, using the 3000 weather years (inaccessible due to force majeure) from which the percentile weather data was derived. In other words, corresponding simulated performance metrics to weather data in the five percentiles may only be analysed as five data points with no assigned percentiles. Furthermore, a sample of five observations is less than enough for the law of large numbers to take effect using conventional statistical inference. Therefore, the robustness quantification method aided by the non-parametric Bayesian bootstrap is proposed, to provide a systematically consistent means of future thermal comfort evaluation under climate uncertainty.

Aside from this, it is important to acknowledge that the bootstrap still has its limitation as to small samples. The measure of robustness for future thermal comfort in this study is the variability of the estimated mean overheating rate, which is denoted by the width of the BC_a interval. Three key factors that determine such variability are a) the population variability, b) the sample size (n) and c) the bootstrap replication number (R) [32]. The population variability is the ideal value of the thermal comfort robustness, which can be well approximated[6] by simulation results using all 3000 weather years; this quantity is instead indicated by means of the Bayesian bootstrap, embedded with uncertainty stemmed from sampling errors in relation to both n and R. As shown in the convergence analysis, the Monte Carlo error can be deemed eliminated using a sufficiently large R, but n is a constant (i.e. five) in this context, which is subject to the delivery of the probabilistic future weather data and leads to a unneglectable part of errors. Their collective uncertainty can be further measured by techniques such as the double bootstrap and the jackknife-after-bootstrap [60].

Nonetheless, the conjecture can be made that the bootstrap BC_a interval, with its adjustment for bias and skewness, is monotonically correlated to the population variability. As such, albeit the existence of embedded uncertainty in absolute values of the robustness indicators, their comparability between design candidates is established, which is key to robustness-informed decision-making. Besides, another benefit of adopting the Bayesian bootstrap can be revealed by the comparison between the two case study homes. Existing literature, such as Gupta and Gregg [17], applied the median weather data to comparing future overheating risk between different buildings. However, amongst 24 data points of the 50th percentile during diurnal occupancy in Fig. 5, a third present higher overheating rate in the mid-terraced house, so it would seem quite arbitrary to draw the comparison using such deterministic approach. Overall, the Bayesian bootstrap demonstrates a good capability of handling uncertainty in future overheating assessment.

Recognising that the percentile-based delivery of probabilistic weather data is the root cause of the uncertainty in robustness estimation, it is strongly recommended that more details should be included rather than merely five percentiles in the future delivery with updated climate projections, so that more sound quantitative analysis can be performed with less difficulty. However, even with the original dataset (i.e. the 3000 weather years) available, the Bayesian bootstrap should still be an essential tool for robustness evaluation of building performance, as it is not practical to run building simulation under all 3000 weather years, especially when a large number of building design candidates are involved in the decision-making process. Rather, a full dataset would be more useful in providing an opportunity for further validation of the bootstrap results.

5.2. Building regulations and energy efficiency measures

In the course of exploring the application of the Bayesian bootstrap in robustness assessment, this study modelled two post-retrofit residential buildings, to examine the comfort-related impact of the current energy-oriented building regulations. A clear observation from the previous analysis is that occupants are expected to experience relatively mild overheating risk when they are awake, but quite severe one when sleeping even in the 2030s. This may further imply the high likelihood of occupants having been experiencing considerable nocturnal overheating in their homes during hot summers in recent years.

This phenomenon can be traced back to various assumptions that TM59 makes in terms of occupancy profiles and overheating thresholds. It may seem quite pessimistic to assume continuous occupancy in its scheduling guidance, as is also argued by Mourkos et al. [53]; but recent trends have shown that working from home becomes increasingly popular, especially since the COVID-19 pandemic [5,6]. On the other hand, it is fairly realistic to consider the difficulty of occupants being 'adaptive' to their thermal comfort during sleep, as windows should remain closed unless security measures are present. With windows closed at nights, excessive heat can hardly be released to the outside through highly insulated fabric in the presence of lower external temperature, contributing to severe overheating risk. In light of this, more attention should be paid to bedrooms in future-proof revision of building regulations because of their nocturnal susceptibility to overheating, so that a better balance can be struck between thermal comfort and energy efficiency.

As highlighted in the results, nocturnal overheating is more of a concern in both dwellings, regarding not only its magnitude but more importantly its variability. This can be mainly attributed to the lack of ventilation, which has been widely discussed by researchers in their work on indoor thermal comfort, such as Gupta et al. [61] who suggested that window opening can make an extensive difference in internal air temperature at nights. In contrast, diurnal overheating risk is both lower and less dispersed, albeit its worsening in later years. Collectively, these indicate that more efforts are required to expel heat outwards rather than to prevent from absorbing it inwards, as measures focussing on the latter, such as shading, mainly work during daytime. It may also suggest that such adaptive measures as window operation is a key factor towards high robustness of future thermal comfort. This can be seen as one of the underlying reasons of high-level insulation inducing unintended consequences of climate change mitigation actions in buildings, as it hampers heat transfer in both directions; consequently, the existence of internal gain leads indoor air temperature to accumulate in the absence of efficient heat release.

A further noteworthy observation relates to the different efforts needed for the two dwelling archetypes to achieve robust thermal

[6] This is only an approximation as further uncertainty sources associated with climate projections exist, but a comparatively better one due to the law of large numbers.

comfort. As discussed before, window opening at nights is subject to security consideration. Although windows in both dwellings are kept closed for comparability, top-floor bedrooms are less likely to be associated with security concern [62], which facilitates the utilisation of night ventilation. In contrast, to introduce nocturnal ventilation, dwellings such as mid-terraced houses are supposed to either invest in extra security measures or resort to mechanical ventilation, both of which would involve higher life cycle cost and energy consumption. However, this should not be interpreted as cooling high-rise homes being easier. It is well recognised that the high-rise is more susceptible to overheating under current climatic conditions [14,62–65], and this seems to remain the case in some rooms and occupied periods in the future according to the previous comparative analysis using the bootstrap hypothesis test. It has been shown that the extensive overheating in top-floor flats is closely linked to their distinctive thermal profiles, but that may not be the only driver. Predominantly located in urban areas, high-rise dwellings are also strongly impacted by the urban heat island (UHI) phenomenon, which was not accounted in this study. Oikonomou et al. [49] indicated that an urban location contributes more to overheating than a lack of shading devices. UHI is a critical factor due to its full-spectrum influence on both diurnal and nocturnal occupancy, as the increment of ambient temperature can not only raise the daytime heat intake but suppress nighttime heat release.

6. Conclusion

From the perspective of domestic overheating, this paper investigated the robustness evaluation of thermal comfort against uncertain future climate, using a non-parametric statistical tool termed the Bayesian bootstrap. The results revealed that the distinction in both expected overheating risk and its variability mainly lies between nocturnal and diurnal occupancy, rather than different rooms in each dwelling. The dissimilarity in thermal profiles was considered as a major driver of the difference in overheating risk between mid-terraced houses and high-rise top-floor flats. Whilst formulation of explicit recommendations for building regulations is beyond the scope of this study, a significant implication was highlighted that adaptive ventilation may be a key measure to enhance the robustness of future thermal comfort. Most importantly, the application of the Bayesian bootstrap demonstrated its good capability to facilitate systematically consistent robustness evaluation, which potentially can better inform risk-based decision-making in building design.

One limitation in this study is the lack of consideration of UHI, which can have considerable impact on the overheating assessment. Integration of UHI in BEM weather data is an emerging research topic, accounting which would be ideal, but the use of directly available weather files is deemed to suffice the aim of this study. In future work, the UKCP18 weather data will be adopted when available. The main findings drawn with UKCP09 in this study are expected to hold, but the more recent dataset would potentially facilitate further validation of the bootstrap results. Investigation is being undertaken into incorporating the bootstrap-aided robustness assessment into building design optimisation to inform future-proof housing stock.

CRediT authorship contribution statement

Cheng Cui: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Rokia Raslan:** Conceptualization, Methodology, Writing – review & editing. **Ivan Korolija:** Conceptualization, Methodology, Writing – review & editing. **Zaid Chalabi:** Methodology, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Formulation of the Bayesian bootstrap

An important concept in formulating the Bayesian bootstrap is the Dirichlet distribution, whose probability density function is defined as

$$f(\mathbf{p};\boldsymbol{\gamma}) = \frac{1}{\mathrm{B}(\boldsymbol{\gamma})}\prod_{k=1}^{K} p_k^{\gamma_k - 1}, \tag{A.1}$$

where K is the order of the distribution, $\mathbf{p} = (p_1, p_2, \dots, p_K)$ is its support with $\sum_{k=1}^{K} p_k = 1$ and $p_k \geq 0$ for $k \in \{1, 2, \dots, K\}$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_K)$ is the concentration parameter with $\gamma_k > 0$ for $k \in \{1, 2, \dots, K\}$, and $\mathrm{B}(\cdot)$ is the multivariate beta function. The formula of $\mathrm{B}(\cdot)$ is omitted herein, as $\mathrm{B}(\boldsymbol{\gamma})$ is a normalisation constant for a given $\boldsymbol{\gamma}$.

As per Rubin [29], mimicking sampling from the population F by resampling the observed sample $\mathbf{x}$ indicates a fundamental assumption underlying the bootstrap — all distinct values in F have been observed in $\mathbf{x}$. If each observation in $\mathbf{x}$ is also distinct, which is a special case of the theory by Rubin [29] and is the case for this study, the probability of their occurrence in F can be denoted by $\mathbf{g} = (g_1, g_2, \dots, g_n)$. The corresponding empirical probability is therefore $\hat{\mathbf{g}} = (\hat{g}_1, \hat{g}_2, \dots, \hat{g}_n)$, where $\hat{g}_i = 1/n$ for $i \in \{1, 2, \dots, n\}$, which can be alternatively denoted in terms of frequency as $n\hat{\mathbf{g}} = (n\hat{g}_1, n\hat{g}_2, \dots, n\hat{g}_n)$ with $n\hat{g}_i = 1$ for $i \in \{1, 2, \dots, n\}$, meaning that the number of times that each distinct value in $\mathbf{x}$ is observed is identical and equal to one.

As per (A.1), taking the Dirichlet distribution $\boldsymbol{Dirichlet}(\boldsymbol{\gamma})$ as the prior gives

$$\mathbf{g} \propto \prod_{i=1}^{n} g_i^{\gamma_i - 1}, \tag{A.2}$$

and the likelihood for the sample is given by

$$\hat{\mathbf{g}} \mid \mathbf{g} = \prod_{i=1}^{n} g_i^{n\hat{g}_i}. \tag{A.3}$$

Applying Bayes' theorem, the posterior is derived as proportional to the prior times the likelihood, leading to another Dirichlet distribution $\boldsymbol{Dirichlet}(\boldsymbol{\gamma} + n\hat{\mathbf{g}})$,

$$\mathbf{g} \mid \hat{\mathbf{g}} \propto \prod_{i=1}^{n} g_i^{\gamma_i + n\hat{g}_i - 1}. \tag{A.4}$$

The generation of $\mathbf{g}^*$ is essentially drawing random samples from such posterior, hence

$$\mathbf{g}^* \sim \boldsymbol{Dirichlet}(\boldsymbol{\gamma} + n\hat{\mathbf{g}}). \tag{A.5}$$

In the limit of $\gamma_i \to 0$ for $i \in \{1, 2, \dots, n\}$ gives the Haldane's prior, a commonly used non-informative prior [66]. Along with $n\hat{\mathbf{g}} = \mathbf{1}$, the posterior $\boldsymbol{Dirichlet}(\boldsymbol{\gamma} + n\hat{\mathbf{g}})$ becomes the uniform Dirichlet distribution $\boldsymbol{Dirichlet}(\mathbf{1})$, and (A.5) becomes (3) accordingly.

Appendix B. Calculation of the running mean temperature

As per TM52 [54], the exponentially weighted running mean temperature is defined as

$$T_{\mathrm{rm}}[d] = (1 - \lambda)\sum_{l=1}^{D} \lambda^{l-1} T_{\mathrm{ed}}[d - l], \tag{B.1}$$

where d is the time index corresponding to the current day, $d - 1$ the previous day, and so on; specially, $d = 0$ refers to the first day. As such, $T_{\mathrm{ed}}[d]$ and $T_{\mathrm{rm}}[d]$ respectively denote the daily mean and the running mean external temperature for the day of the time index d. $\lambda \in (0, 1)$ is a decay factor that describes the impact of the past, and D is the number of previous days to be accounted.

To calculate T_{rm} for a series of annual weather data more efficiently, (B.1) can be first applied to compute $T_{rm}[0]$ for initialisation, using T_{ed} for the last D days in the same year; then $T_{rm}[1], T_{rm}[2], \ldots, T_{rm}[364]$ (assuming a non-leap year) can be ordinally obtained by

$$T_{rm}[d] = (1-\lambda)T_{ed}[d-1] + \lambda T_{rm}[d-1], \tag{B.2}$$

which is equivalent to (B.1) given the initial condition $T_{rm}[0]$. Whilst D can be as large as possible, $D = 7$ with a λ of 0.8 was used in this study, as recommended in TM52 [54].

Appendix C. Computation of the BC$_a$ interval

As per Efron and Tibshirani [32], α_{lo} and α_{up} for the BC$_a$ interval are defined as

$$\begin{aligned}\alpha_{lo} &= \Phi\Big(\hat{b} + \frac{\hat{b}+\Phi^{-1}(\alpha)}{1-\hat{a}(\hat{b}+\Phi^{-1}(\alpha))}\Big),\\ \alpha_{up} &= \Phi\Big(\hat{b} + \frac{\hat{b}+\Phi^{-1}(1-\alpha)}{1-\hat{a}(\hat{b}+\Phi^{-1}(1-\alpha))}\Big),\end{aligned} \tag{C.1}$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution, $\Phi^{-1}(\cdot)$ is the inverse function of $\Phi(\cdot)$, $\hat{b}$ is the bias-correction factor computed by

$$\hat{b} = \Phi^{-1}(\Pr(\hat{\theta}^* < \hat{\theta})), \tag{C.2}$$

and $\hat{a}$ is the acceleration factor that can be approximated via

$$\hat{a} = \frac{\sum_{i=1}^{n}(\bar{\hat{\theta}}^{\dagger} - \hat{\theta}_i^{\dagger})^3}{6[\sum_{i=1}^{n}(\bar{\hat{\theta}}^{\dagger} - \hat{\theta}_i^{\dagger})^2]^{3/2}}, \tag{C.3}$$

where $\hat{\theta}_i^{\dagger}$ is obtained via the jackknife, an alternative resampling method predating and closely linked to the bootstrap.

A jackknife sample $\mathbf{x}_i^{\dagger} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ is constructed as a subset of the original sample $\mathbf{x}$ by removing the ith observation, $\hat{\theta}_i^{\dagger}$ is the jackknife replication of $\hat{\theta}$ corresponding to $\mathbf{x}_i^{\dagger}$. For instance, if the sample has five observations $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$ and $\hat{\theta}$ is the sample mean, the jackknife replication $\hat{\theta}_3^{\dagger}$ is calculated by $(x_1 + x_2 + x_4 + x_5)/4$.

Appendix D. Execution of the bootstrap hypothesis test

As per Efron and Tibshirani [32] and Noreen [57], the paired two-sample bootstrap t-test can be conducted as follows.

1. Calculate the paired difference between the two samples as $\Delta\mathbf{x}$.
2. Translate the paired difference to have a mean of zero by
$$\widetilde{\Delta\mathbf{x}} = \Delta\mathbf{x} - \overline{\Delta x}. \tag{D.1}$$
3. Generate R random resampling vectors $\mathbf{g}_1^*, \mathbf{g}_2^*, \ldots, \mathbf{g}_R^*$ that corresponds to $\widetilde{\Delta\mathbf{x}}$.
4. Evaluate the corresponding bootstrap replications of the test statistic $t_1^*, t_2^*, \ldots, t_R^*$ to the resampling vectors using (7)
$$t^* = \frac{\overline{\Delta x}^*}{\sqrt{s_{\Delta x}^{2\,*}/n}}. \tag{D.2}$$
5. Estimate the p-value applying the inverse of the BC$_a$ interval in Appendix C, with the observed value of the test statistic calculated by (7) on $\Delta\mathbf{x}$.

Appendix E. Implementation of the Bayesian bootstrap

An example implementation for the Bayesian bootstrap is demonstrated via pseudocode in Algorithm 1, using the sample mean as the statistic of interest. Functions not herein defined are accessible via either Python standard libraries or NumPy [67] and SciPy [68], two established Python packages for scientific computing. The mathematical symbols used are the same as in the main paper.

Algorithm 1 The Bayesian bootstrap of the sample mean.

```
function Main(x, R, α)
    n ← Length(x)
    γ ← Ones(n)
    for j ← 1, R do
        g_j ← RandomDirichlet(γ)
        θ̂*_j ← WeightedMean(x, g_j)
    end for
    return BCa(x, α, (θ̂*_1, θ̂*_2, ..., θ̂*_R))
end function
function BCa(x, α, (θ̂*_1, θ̂*_2, ..., θ̂*_R))
    b̂ ← BiasCorrection(x, (θ̂*_1, θ̂*_2, ..., θ̂*_R))
    â ← Acceleration(x)
    α_lo ← CumulativeProbability(α, b̂, â)
    α_up ← CumulativeProbability(1 − α, b̂, â)
    return Quantile((θ̂*_1, θ̂*_2, ..., θ̂*_R), (α_lo, α_up))
end function
function BiasCorrection(x, (θ̂*_1, θ̂*_2, ..., θ̂*_R))
    θ̂ ← Mean(x)
    count ← 0
    for j ← 1, R do
        if θ̂*_j < θ̂ then
            count ← count + 1
        end if
    end for
    return NormalInverseCDF(count / R)
end function
function Acceleration(x)
    (θ̂†_1, θ̂†_2, ..., θ̂†_n) ← Jackknife(x)
    θ̄̂† ← Mean((θ̂†_1, θ̂†_2, ..., θ̂†_n))
    numerator ← 0
    sum ← 0
    for i ← 1, n do
        difference ← θ̄̂† − θ̂†_i
        numerator ← numerator + Power(difference, 3)
        sum ← sum + Power(difference, 2)
    end for
    denominator ← Power(sum, 3/2) * 6
    return numerator / denominator
end function
function Jackknife(x)
    n ← Length(x)
    for i ← 1, n do
        x†_i ← Delete(x, i)
        θ̂†_i ← Mean(x†_i)
    end for
    return (θ̂†_1, θ̂†_2, ..., θ̂†_n)
end function
function CumulativeProbability(α, b̂, â)
    numerator ← b̂ + NormalInverseCDF(α)
    denominator ← 1 − â * numerator
    return NormalCDF(b̂ + numerator / denominator)
end function
```

References

[1] IPCC, Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom & New York, United States, 2021.

[2] CCC, Progress in Reducing Emissions: 2021 Report to Parliament, Climate Change Committee, London, United Kingdom, 2021.

[3] EEA, Annual European Union Greenhouse Gas Inventory 1990–2019 and Inventory Report 2021: Submission to the UNFCCC Secretariat, European Environment Agency, Copenhagen, Denmark, 2021.

[4] CCC, Independent Assessment of UK Climate Risk: Advice to Government for the UK's Third Climate Change Risk Assessment (CCRA3), Climate Change Committee, London, United Kingdom, 2021.

[5] A. Ambrose, W. Baker, G. Sherriff, J. Chambers, Cold comfort: Covid-19, lockdown and the coping strategies of fuel poor households, Energy Rep. 7 (2021) 5589–5596, http://dx.doi.org/10.1016/j.egyr.2021.08.175.

[6] M. Awada, B. Becerik-Gerber, S. Hoque, Z. O'Neill, G. Pedrielli, J. Wen, T. Wu, Ten questions concerning occupant health in buildings during normal operations and extreme events including the COVID-19 pandemic, Build. Environ. 188 (2021) 107480, http://dx.doi.org/10.1016/j.buildenv.2020.107480.

[7] S. Kovats, R. Brisley, Health, Communities and the Built Environment, Climate Change Committee, London, United Kingdom, 2021.

[8] M. Collins, S. Dempsey, Residential energy efficiency retrofits: potential unintended consequences, J. Environ. Plan. Manag. 62 (12) (2019) 2010–2025, http://dx.doi.org/10.1080/09640568.2018.1509788.

[9] H. Elsharkawy, S. Zahiri, The significance of occupancy profiles in determining post retrofit indoor thermal comfort, overheating risk and building energy performance, Build. Environ. 172 (2020) 106676, http://dx.doi.org/10.1016/j.buildenv.2020.106676.

[10] P. de Wilde, D. Coley, The implications of a changing climate for buildings, Build. Environ. 55 (2012) 1–7, http://dx.doi.org/10.1016/j.buildenv.2012.03.014.

[11] A. Pathan, A. Mavrogianni, A. Summerfield, T. Oreszczyn, M. Davies, Monitoring summer indoor overheating in the London housing stock, Energy Build. 141 (2017) 361–378, http://dx.doi.org/10.1016/j.enbuild.2017.02.049.

[12] DLUHC, The Future Buildings Standard: Summary of Responses, and Government Response, Department for Levelling Up, Housing & Communities, London, United Kingdom, 2021.

[13] HM Government, Approved Document O, Overheating, Department for Levelling Up, Housing & Communities, London, United Kingdom, 2021.

[14] A. Mavrogianni, P. Wilkinson, M. Davies, P. Biddulph, E. Oikonomou, Building characteristics as determinants of propensity to high indoor summer temperatures in London dwellings, Build. Environ. 55 (2012) 117–130, http://dx.doi.org/10.1016/j.buildenv.2011.12.003.

[15] J.M. Murphy, D.M.H. Sexton, G.J. Jenkins, P.M. Boorman, B.B.B. Booth, C.C. Brown, R.T. Clark, M. Collins, G.R. Harris, E.J. Kendon, R.A. Betts, S.J. Brown, T.P. Hinton, K.A. Howard, M.P. McDonald, R.E. McCarthy, A. Wood, C. Wallace, R. Warren, R. Wilby, R.A. Wood, UK Climate Projections Science Report: Climate Change Projections, Met Office Hadley Centre, Exeter, United Kingdom, 2009.

[16] J.M. Murphy, G.R. Harris, D.M.H. Sexton, E.J. Kendon, P.E. Bett, R.T. Clark, K.E. Eagle, G. Fosser, F. Fung, J.A. Lowe, R.E. McDonald, R.N. McInnes, C.F. McSweeney, J.F.B. Mitchell, J.W. Rostron, H.E. Thornton, S. Tucker, K. Yamazaki, UKCP18 Land Projections: Science Report, Met Office Hadley Centre, Exeter, United Kingdom, 2018.

[17] R. Gupta, M. Gregg, Assessing energy use and overheating risk in net zero energy dwellings in UK, Energy Build. 158 (2018) 897–905, http://dx.doi.org/10.1016/j.enbuild.2017.10.061.

[18] J. Taylor, P. Symonds, C. Heaviside, Z. Chalabi, M. Davies, P. Wilkinson, Projecting the impacts of housing on temperature-related mortality in London during typical future years, Energy Build. 249 (2021) 111233, http://dx.doi.org/10.1016/j.enbuild.2021.111233.

[19] T. Aven, Risk assessment and risk management: review of recent advances on their foundation, Eur. J. Oper. Res. 253 (1) (2016) 1–13, http://dx.doi.org/10.1016/j.ejor.2015.12.023.

[20] B. Roy, Robustness in operational research and decision aiding: a multi-faceted issue, Eur. J. Oper. Res. 200 (3) (2010) 629–638, http://dx.doi.org/10.1016/j.ejor.2008.12.036.

[21] J. Wright, E. Nikolaidou, C.J. Hopfe, Exhaustive search: does it have a role in explorative design? in: Proc. BSO Conf. 2016 Third Conf. IBPSA Engl., International Building Performance Simulation Association England, Newcastle, United Kingdom, 2016, pp. 49–56.

[22] V. Gabrel, C. Murat, A. Thiele, Recent advances in robust optimization: an overview, Eur. J. Oper. Res. 235 (3) (2014) 471–483, http://dx.doi.org/10.1016/j.ejor.2013.09.036.

[23] R. Kotireddy, P.-J. Hoes, J.L.M. Hensen, Simulation-based comparison of robustness assessment methods to identify robust low-energy building designs, in: Proc. Build. Simul. 2017 15th Conf. IBPSA, International Building Performance Simulation Association, San Francisco, United States, 2017, pp. 892–901, http://dx.doi.org/10.26868/25222708.2017.240.

[24] A. Moazami, S. Carlucci, V.M. Nik, S. Geving, Towards climate robust buildings: an innovative method for designing buildings with robust energy performance under climate change, Energy Build. 202 (2019) 109378, http://dx.doi.org/10.1016/j.enbuild.2019.109378.

[25] P.-J. Hoes, M. Trcka, J.L.M. Hensen, B.H. Bonnema, Optimizing building designs using a robustness indicator with respect to user behavior, in: Proc. Build. Simul. 2011 12th Conf. IBPSA, International Building Performance Simulation Association, Sydney, Australia, 2011, pp. 1710–1717, http://dx.doi.org/10.26868/25222708.2011.1560.

[26] M.S. Wisz, R.J. Hijmans, J. Li, A.T. Peterson, C.H. Graham, A. Guisan, NCEAS Predicting Species Distributions Working Group, Effects of sample size on the performance of species distribution models, Divers. Distrib. 14 (5) (2008) 763–773, http://dx.doi.org/10.1111/j.1472-4642.2008.00482.x.

[27] K.S. Button, J.P.A. Ioannidis, C. Mokrysz, B.A. Nosek, J. Flint, E.S.J. Robinson, M.R. Munafo, Power failure: why small sample size undermines the reliability of neuroscience, Nat. Rev. Neurosci. 14 (5) (2013) 365–376, http://dx.doi.org/10.1038/nrn3475.

[28] W. Tian, Y. Heo, P. de Wilde, Z. Li, D. Yan, C.S. Park, X. Feng, G. Augenbroe, A review of uncertainty analysis in building energy assessment, Renew. Sustain. Energy Rev. 93 (2018) 285–301, http://dx.doi.org/10.1016/j.rser.2018.05.029.

[29] D.B. Rubin, The Bayesian bootstrap, Ann. Stat. 9 (1) (1981) 130–134, http://dx.doi.org/10.1214/aos/1176345338.

[30] A.Y. Lo, A Bayesian bootstrap for a finite population, Ann. Stat. 16 (4) (1988) 1684–1695, http://dx.doi.org/10.1214/aos/1176351061.

[31] B. Efron, Bootstrap methods: another look at the jackknife, Ann. Stat. 7 (1) (1979) 1–26, http://dx.doi.org/10.1214/aos/1176344552.

[32] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, first ed., in: Monographs on Statistics and Applied Probability, vol. 57, Chapman & Hall, New York, United States, 1993, http://dx.doi.org/10.1201/9780429246593.

[33] M.R. Chernick, Bootstrap Methods: A Guide for Practitioners and Researchers, second ed., John Wiley & Sons, Hoboken, United States, 2008.

[34] P. Rastogi, M. Andersen, Generation of weather files using resampling techniques: an exploratory study, in: Proc. Build. Simul. 2013 13th Conf. IBPSA, International Building Performance Simulation Association, Chambéry, France, 2013, pp. 1422–1429, http://dx.doi.org/10.26868/25222708.2013.1081.

[35] W. Tian, J. Song, Z. Li, P. de Wilde, Bootstrap techniques for sensitivity analysis and model selection in building thermal performance analysis, Appl. Energy 135 (2014) 320–328, http://dx.doi.org/10.1016/j.apenergy.2014.08.110.

[36] Q. Li, S.J. Quan, G. Augenbroe, P.P.-J. Yang, J. Brown, Building energy modelling at urban scale: integration of reduced order energy model with geographical information, in: Proc. Build. Simul. 2015 14th Conf. IBPSA, International Building Performance Simulation Association, Hyderabad, India, 2015, pp. 190–199, http://dx.doi.org/10.26868/25222708.2015.2706.

[37] P. Rastogi, M. Andersen, Embedding stochasticity in building simulation through synthetic weather files, in: Proc. Build. Simul. 2015 14th Conf. IBPSA, International Building Performance Simulation Association, Hyderabad, India, 2015, pp. 963–970, http://dx.doi.org/10.26868/25222708.2015.2321.

[38] X. Chen, H. Yang, K. Sun, Developing a meta-model for sensitivity analyses and prediction of building performance for passively designed high-rise residential buildings, Appl. Energy 194 (2017) 422–439, http://dx.doi.org/10.1016/j.apenergy.2016.08.180.

[39] M.S. Geraldi, M.V. Bavaresco, E. Ghisi, Bayesian network for predicting energy consumption in schools in Florianópolis – Brazil, in: Proc. Build. Simul. 2019 16th Conf. IBPSA, International Building Performance Simulation Association, Rome, Italy, 2019, pp. 4188–4195, http://dx.doi.org/10.26868/25222708.2019.210484.

[40] S. Chaturvedi, E. Rajasekar, Application of a probabilistic LHS-PAWN approach to assess building cooling energy demand uncertainties, Build. Simul. 15 (3) (2022) 373–387, http://dx.doi.org/10.1007/s12273-021-0815-6.

[41] S.N. Lahiri, Resampling Methods for Dependent Data, first ed., in: Springer Series in Statistics, Springer-Verlag, New York, United States, 2003, http://dx.doi.org/10.1007/978-1-4757-3803-2.

[42] P. Rastogi, M. Andersen, Incorporating climate change predictions in the analysis of weather-based uncertainty, in: ASHRAE IBPSA-USA SimBuild 2016 Build. Perform. Model. Conf., International Building Performance Simulation Association USA, Salt Lake City, United States, 2016, pp. 181–188.

[43] Y.-S. Chiou, K.M. Carley, C.I. Davidson, M.P. Johnson, A high spatial resolution residential energy model based on American Time Use Survey data and the bootstrap sampling method, Energy Build. 43 (12) (2011) 3528–3538, http://dx.doi.org/10.1016/j.enbuild.2011.09.020.

[44] Q. Li, G. Augenbroe, R. Muehleisen, A framework for empirical validation of building performance simulation under uncertainty, in: Proc. Build. Simul. 2017 15th Conf. IBPSA, International Building Performance Simulation Association, San Francisco, United States, 2017, pp. 1770–1779, http://dx.doi.org/10.26868/25222708.2017.466.

[45] Y. Xie, V. Mendon, M. Halverson, R. Bartlett, J. Hathaway, Y. Chen, M. Rosenberg, T. Taylor, B. Liu, Assessing overall building energy performance of a large population of residential single-family homes using limited field data, J. Build. Perform. Simul. 12 (4) (2019) 480–493, http://dx.doi.org/10.1080/19401493.2018.1477833.

[46] T. Ostergard, R.L. Jensen, F.S. Mikkelsen, The best way to perform building simulations? One-at-a-time optimization vs. Monte Carlo sampling, Energy Build. 208 (2020) 109628, http://dx.doi.org/10.1016/j.enbuild.2019.109628.

[47] D.B. Crawley, L.K. Lawrie, F.C. Winkelmann, W. Buhl, Y. Huang, C.O. Pedersen, R.K. Strand, R.J. Liesen, D.E. Fisher, M.J. Witte, J. Glazer, EnergyPlus: creating a new-generation building energy simulation program, Energy Build. 33 (4) (2001) 319–331, http://dx.doi.org/10.1016/S0378-7788(00)00114-6.

[48] G. van Rossum, F.L. Drake, Python 3 Reference Manual, CreateSpace, Scotts Valley, United States, 2009.

[49] E. Oikonomou, M. Davies, A. Mavrogianni, P. Biddulph, P. Wilkinson, M. Kolokotroni, Modelling the relative importance of the urban heat island and the thermal quality of dwellings for overheating in London, Build. Environ. 57 (2012) 223–238, http://dx.doi.org/10.1016/j.buildenv.2012.04.002.

[50] HM Government, Approved Document L, Conservation of Fuel and Power, Volume 1: Dwellings, Department for Levelling Up, Housing & Communities, London, United Kingdom, 2021.

[51] CIBSE, Guide A: Environmental Design, Chartered Institution of Building Services Engineers, London, United Kingdom, 2015.

[52] CIBSE, TM59: Design Methodology for the Assessment of Overheating Risk in Homes, Chartered Institution of Building Services Engineers, London, United Kingdom, 2017.

[53] K. Mourkos, R.S. McLeod, C.J. Hopfe, C. Goodier, M. Swainson, Assessing the application and limitations of a standardised overheating risk-assessment methodology in a real-world context, Build. Environ. 181 (2020) 107070, http://dx.doi.org/10.1016/j.buildenv.2020.107070.

[54] CIBSE, TM52: The Limits of Thermal Comfort: Avoiding Overheating in European Buildings, Chartered Institution of Building Services Engineers, London, United Kingdom, 2013.

[55] H.E. Beck, N.E. Zimmermann, T.R. McVicar, N. Vergopolan, A. Berg, E.F. Wood, Present and future Köppen–Geiger climate classification maps at 1-km resolution, Sci. Data 5 (1) (2018) 180214, http://dx.doi.org/10.1038/sdata.2018.214.

[56] M. Eames, T. Kershaw, D. Coley, On the creation of future probabilistic design weather years from UKCP09, Build. Serv. Eng. Res. Technol. 32 (2) (2011) 127–142, http://dx.doi.org/10.1177/0143624410379934.

[57] E.W. Noreen, Computer-Intensive Methods for Testing Hypotheses: An Introduction, first ed., John Wiley & Sons, New York, United States, 1989.

[58] C. Liu, T. Kershaw, M.E. Eames, D.A. Coley, Future probabilistic hot summer years for overheating risk assessments, Build. Environ. 105 (2016) 56–68, http://dx.doi.org/10.1016/j.buildenv.2016.05.028.

[59] T. Kershaw, M. Eames, D. Coley, Assessing the risk of climate change for buildings: a comparison between multi-year and probabilistic reference year simulations, Build. Environ. 46 (6) (2011) 1303–1308, http://dx.doi.org/10.1016/j.buildenv.2010.12.018.

[60] B. Efron, Jackknife-after-bootstrap standard errors and influence functions, J. R. Stat. Soc. Ser. B Methodol. 54 (1) (1992) 83–111, http://dx.doi.org/10.1111/j.2517-6161.1992.tb01866.x.

[61] R. Gupta, A. Howard, M. Davies, A. Mavrogianni, I. Tsoulou, N. Jain, E. Oikonomou, P. Wilkinson, Monitoring and modelling the risk of summertime overheating and passive solutions to avoid active cooling in London care homes, Energy Build. 252 (2021) 111418, http://dx.doi.org/10.1016/j.enbuild.2021.111418.

[62] P. Symonds, J. Taylor, A. Mavrogianni, M. Davies, C. Shrubsole, I. Hamilton, Z. Chalabi, Overheating in English dwellings: comparing modelled and monitored large-scale datasets, Build. Res. Inf. 45 (1–2) (2017) 195–208, http://dx.doi.org/10.1080/09613218.2016.1224675.

[63] K.J. Lomas, S.M. Porritt, Overheating in buildings: lessons from research, Build. Res. Inf. 45 (1–2) (2017) 1–18, http://dx.doi.org/10.1080/09613218.2017.1256136.

[64] R. Mitchell, S. Natarajan, Overheating risk in Passivhaus dwellings, Build. Serv. Eng. Res. Technol. 40 (4) (2019) 446–469, http://dx.doi.org/10.1177/0143624419842006.

[65] H. Thomson, N. Simcock, S. Bouzarovski, S. Petrova, Energy poverty and indoor cooling: an overlooked issue in Europe, Energy Build. 196 (2019) 21–29, http://dx.doi.org/10.1016/j.enbuild.2019.05.014.

[66] D. Alvares, C. Armero, A. Forte, What does objective mean in a Dirichlet-multinomial process? Int. Stat. Rev. 86 (1) (2018) 106–118, http://dx.doi.org/10.1111/insr.12231.

[67] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T.E. Oliphant, Array programming with NumPy, Nature 585 (7825) (2020) 357–362, http://dx.doi.org/10.1038/s41586-020-2649-2.

[68] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods 17 (3) (2020) 261–272, http://dx.doi.org/10.1038/s41592-019-0686-2.