Energy & Buildings 295 (2023) 113265

Contents lists available at ScienceDirect

Energy & Buildings

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

Characterisation and analysis of uncertainties in building heat transfer estimates from co-heating tests

Virginia Gori [a], David Johnston [b], Rémi Bouchié [c], Samuel Stamp [d,*]

[a] UCL Energy Institute, University College London, 14 Upper Woburn Place, London, WC1H 0NN, United Kingdom
[b] Leeds Sustainability Institute, Leeds Metropolitan University, Northern Terrace, 209, City Campus, Leeds, LS2 8AG, United Kingdom
[c] Centre Scientifique et Technique du Bâtiment (CSTB), Building Envelop Department, Expertise and Research Unit, 84 Avenue Jean Jaurès, 77420, Champs-sur-Marne, France
[d] UCL Institute for Environmental Design and Engineering, University College London, 14 Upper Woburn Place, London, WC1H 0NN, United Kingdom

A R T I C L E I N F O

MSC:
0000
1111

Keywords:
Co-heating test
Uncertainty characterisation
Heat Transfer Coefficient (HTC)
Field measurements
Thermal performance
Performance gap
Prediction gap

A B S T R A C T

In recent years, measurement protocols for the estimation of the total aggregate building heat transfer coefficient (HTC) have provided sufficient empirical evidence to indicate that buildings often do not perform as intended. However, little research has been carried out into the associated uncertainties. Within this context, this paper reviews sources of uncertainty associated with co-heating tests; characterises these uncertainties and their impact on HTC estimates; and devises a method for the calculation of HTC uncertainty. The method proposed was applied to 14 co-heating tests, showing estimated total uncertainty ranging between 2.2-21.1 W/K (or 4.6-26.7% of the measured value) with a mean of 10.1 W/K (or 8.7%). The natural variation of HTC and often-observed inaccuracy of design calculations (the 'prediction gap') suggest that more accurate measurements may be of little benefit. Additionally, results suggest that weather conditions, challenging building design and poor experimental technique can all significantly contribute to HTC uncertainty. However, when suitable buildings are tested by experienced technicians and under suitable weather conditions, HTC estimates from the co-heating protocol are likely to provide a useful tool to assess and understand real-world building fabric performance.

1. Introduction

In recent years, a series of field measurements have added to a growing body of empirical evidence that indicated that in-situ performance generally does not correspond to predicted performance and may vary significantly from predictions [1–4]. These measurements have also highlighted a lack of knowledge concerning the actual thermal performance of buildings and the processes, systems and materials that can act to undermine (or sometimes improve) it.

Measurement protocols to estimate total aggregate building heat loss, or a building's heat transfer coefficient (HTC), can provide a useful insight into fabric performance. Unlike discrete or disaggregate measurement methods (e.g., in-situ U-value measurements, pressurisation testing), aggregate methods are capable of capturing all of the complex inter-related heat transfer that occur across the entire building fabric of a building, accounting for thermal bridges, junctions, defects, convective bypasses, bulk air movement and other forms of non-uniform or more complex heat loss pathways. Although a number of aggregate measurement methods exist, such as ISABELE [5,6], the PSTAR method [7–9] and the QUB method [10], the only method that has seen significant application in the field to date is the co-heating test method [11,12]. It was first proposed in North America in the late 1970's by Socolow [13], with the first documented description of the test method being published by Sonderegger et al. in 1979 [14]. The earliest documented use of the test method in the UK was in the mid 1980's [15,16]. Since then the co-heating test method has been used frequently in building performance evaluations over the last two decades [e.g., 17–19], as well as for testing novel constructions [e.g., 20,21] and retrofit measures [e.g., 22]. Results have shown wide ranging performance, with estimates from previous tests undertaken in newly built dwellings indicating that fabric heat loss under unoccupied test conditions is an average of 1.6 times higher than predicted [4].

Such findings have far-reaching consequences and would threaten to undermine efforts to cut energy demand in the domestic sector. However, interpreting measurements of a building's performance without any accompanying estimates of uncertainty risks either overstating or

* Corresponding author.
E-mail address: samuel.stamp@ucl.ac.uk (S. Stamp).

https://doi.org/10.1016/j.enbuild.2023.113265
Received 9 February 2023; Received in revised form 9 May 2023; Accepted 12 June 2023
Available online 16 June 2023

Nomenclature

ΔT	Temperature difference between the indoor and external environments .... [K]	n_s	Number of sensors measuring a given variable
ΔT_j	Temperature difference across party wall/floor j ..... [K]	P_h	Power supplied to heat up the indoor space .... [W]
$\sigma(T_i)$	Standard deviation between temperatures recorded throughout the test building	q_{sw}	Global solar irradiance .... [W/m^2]
ε	Additive error term	q_j	Estimated heat flow through party wall/floor j .. [W/m^2]
A_{sw}	Equivalent solar aperture .... [m^2]	t	Index of the time sample ($1 \le t \le n$)
A_j	Area party wall/floor j .... [m^2]	T_e	External temperature .... [K]
H	Heat transfer coefficient (HTC) of the building ... [W/K]	T_i	Indoor temperature .... [K]
n	Number of time samples (e.g., 24-hour periods) used for regression	$U(\cdot)$	Expanded uncertainty of the enclosed quantity (with a 95 % bilateral confidence interval)
		$u(\cdot)$	Standard uncertainty of the enclosed quantity
		u_{sensor}	Specified uncertainty of the respective sensor

understating the size of any potential performance gap. As stated by the Joint Committee for Guides in Metrology “a statement of measurement uncertainty is indispensable in judging the fitness for purpose of a measured quantity value” [23]. In the absence of reasonable uncertainty estimates, many of the existing measurements of building heat transfer cannot be placed in context. This may lead to mis-characterisation of the building fabric performance gap; unfair comparisons between buildings, materials and construction methods; or mis-interpretations of repeated measurements upon the same building under different stages of retrofit.

For example, in order to understand the scale of the problem, in 16 HTC estimates reported in two major building performance evaluation programmes, just 3 had any accompanying uncertainty estimates [18,19]. Even in these cases, uncertainty estimates were based purely on statistical uncertainty estimates, ignoring both simple measurement uncertainties (e.g., sensor accuracy) and more complex uncertainties and systematic bias that may exist. Similar tendency to neglect estimates of uncertainties associated with in-situ evaluation of building performance has also been observed more widely, including in in-situ U-value measurements [e.g., 18,24,25], air permeability [e.g., 19,25] and ventilation measurements [e.g., 26,27,25]. Methods attempting to begin to account for systematic uncertainties in HTC [6,28] and U-value [29] estimates have been developed recently, although they have not been widely adopted in industry.

The absence of uncertainty estimates may result from both a lack in established methods or guidance for estimating uncertainty, and more broadly a lack of understanding of the sources of uncertainty that may be present within test measurements [30,31]. This has resulted in a lack of confidence in the results obtained from co-heating measurements and has limited their application within industry [18,31], although other studies suggest there is scope for more widespread application [32,33]. These issues are not solely restricted to aggregate whole-house heat transfer measurements. Lack of clarity in overall methods for uncertainty calculations, characterisation, and comprehensive and suitable uncertainty budgets have also been observed for example in U-value estimation from in-situ measurements. While the ISO 9869-1 [34] standard lists the main sources of uncertainties affecting the measurements and quantifies their proportional effect on the U-value, it does not detail how these percentages (generally stated as a fixed value) were evaluated or describe how more accurate values could be quantified in specific circumstances [29]. Further examples of uncertainty analysis using infrared methods can be seen within the appendix of ISO 9869-2 [35].

Within this context, the research presented in this paper aims to address the issues identified by:

- Reviewing sources of uncertainty that may be associated with co-heating measurements used to estimate a building’s HTC;
- Characterising these uncertainties and their impact upon the interpretation of HTC estimates;
- Detailing a method for calculating uncertainties associated with co-heating HTC estimates;
- Estimating uncertainties for existing co-heating tests to determine typical ranges and influencing factors.

2. The co-heating test and heat transfer coefficient evaluation

As previously stated, one of the most common experimental techniques to evaluate the whole-building aggregate heat transfer (both fabric and background ventilation [36]) is the co-heating test [12,37]. During the test, which is normally performed in winter, thermostatically controlled (using proportional, integral and derivative (PID) control) electric resistance heaters are deployed throughout the dwelling to maintain a constant indoor temperature (typically in the range of 20-25 °C) and achieve an average temperature difference of at least 10 K between the internal and external environment. Electric air circulation fans are simultaneously laid out to ensure a good air mix and minimise temperature stratification and dead zones. The electrical energy required during the test is monitored, as well as indoor temperatures and a range of external weather parameters, such as: air temperature; wind speed and direction; and incident solar radiation, ideally measured on the vertical plane of the building facade expected to receive the highest proportion of solar gains. Additional measurements (e.g., heat flux density through building elements, relative humidity, local wind conditions, etc.) may also be collected to gain further information on the thermophysical behaviour of the building fabric.

Owing to the quasi-stationary nature of the test, linear regression models are generally adopted to analyse the data collected and evaluate the aggregate heat transfer coefficient of the building. These models are based on a steady-state energy balance of the building:

$$P_h = H\,\Delta T - A_{sw}\,q_{sw} \tag{1}$$

where P_h is the power supplied to heat up the indoor space [W], H is the heat transfer coefficient of the building [W/K] (including both fabric and infiltration heat losses - typically, ventilation openings are sealed during testing), ΔT is the temperature difference between the indoor (T_i) and external (T_e) environments [K], A_{sw} is the equivalent solar aperture [m^2], q_{sw} is the global solar irradiance [W/m^2].

Among the most common data analysis methods for the estimation of the whole-building aggregate heat transfer coefficient are the Siviour and multiple linear regression analysis methods [12]. The Siviour analysis adopts a bi-axial regression approach. At each daily sample t, the daily average global solar irradiance (independent variable) is plotted against the daily average electrical heating power (dependent variable), with both terms divided by the daily average temperature difference between the inside and outside [15]:

$$\frac{P_{h,t}}{\Delta T_t} = H - \frac{A_{sw}\,q_{sw,t}}{\Delta T_t} + \varepsilon_t. \tag{2}$$

The y-axis intercept of the regression line represents the heat transfer coefficient of the building, while the gradient is the 'solar aperture' – i.e. a quantity equivalent to a totally transparent area letting in the same solar energy as the whole building [38].

In the multiple linear regression (MLR) analysis method, the dependent variable is represented by the daily averaged electrical power, while the independent variables are the daily average global solar irradiance and the internal-to-external daily average temperature difference [12,16]:

$$P_{h,t} = H\,\Delta T_t - A_{sw}\,q_{sw,t} + \varepsilon_t. \tag{3}$$

In this method, the heat transfer coefficient is estimated by plotting the daily average temperature differences (independent variable) against the total daily average heating power (dependent variable), which includes both the electrical power and the solar heat input (independent variable) and calculated as the solar aperture multiplied by the mean global irradiance for each day. The linear regression line is forced through the origin, while the correlation coefficient between the global solar irradiance and the electrical power provides an estimate of the solar aperture.

Variations and extensions to this method exist. In some cases, MLR analysis includes daily averaged global solar irradiance and both internal and external daily average temperatures as independent variables. The HTC value is then calculated by weighting linear regression coefficients identified for both internal and external temperatures [39]. In most cases, the linear regression coefficients are determined generally assuming an unbiased estimate, i.e. assuming that the heating power is nought if there are no temperature difference between internal and external conditions nor solar radiation, although systematically testing the significance of introducing a bias has been proposed [6].

Researchers have also developed dynamic test methods and forms of analysis, potentially offering some improvements on the steady-state method - particularly in terms of test length [6,9,40–42]. In addition, with the advent of smart-meters, there has also been interest in determining the in-use HTC – i.e. determined from normally occupied dwellings [43–47]. While this paper addresses primarily steady-state methods that are deployed in unoccupied dwellings, there is significant cross-over between these different approaches to estimating a building's HTC. As a result, outcomes of this work will be relevant to both dynamic and in-use test methods.

3. Review of uncertainties

In this section, evidence of the various sources of uncertainties in HTC estimates are systematically reviewed. These are classified into either measurement (section 3.2) or model (section 3.3) uncertainties. Before this, previous studies directly investigating uncertainty and self-consistency are reviewed (section 3.1).

3.1. Studies into uncertainty and self-consistency

Few studies have directly aimed to investigate uncertainty in the co-heating method itself, focusing instead on the results of measurements. In the earliest known work on self-consistency of co-heating measurements, Everett [16] reported a range of 21% in HTC estimates across 9 consecutive tests in the same dwelling, largely thought to be the result of unsuitable testing conditions. A series of tests on the same test dwelling under the UK National House Building Council (NHBC) field trial, reported the results obtained for 6 tests that were within 15% of the mean [28,30]. However, these estimates were not fully blind, obtained from different test organisations using variations of the same test method and again across varying seasons and test conditions. Alzetto et al. [48] conducted a series of co-heating tests on a test house inside a controlled environmental chamber, concluding that retrofit measures could only be clearly observed when they represented at least a 10% change to the HTC. In another study [49], also conducted within a controlled environmental chamber, no statistically significant difference in the HTC was observed over three co-heating test phases with different internal/external temperature differences. Finally, across single night measurements, Lloyd et al. [50] reported uncertainties ranging between 4-22% on tests at different stages of retrofit, again resulting in smaller improvements being not statistically significant.

Due to their shorter nature, the self-consistency of short term, dynamic, test methods have been evaluated more frequently. Repeated measurements using the PSTAR method on test cells, both outdoors and within an indoor controlled environment, reported a standard deviation within 5% of the mean [51,52]. However, these measurements were often on simplified, lightweight constructions, limiting the influence of some types of uncertainties. More recently, the QUB method has been found to show a maximum deviation of 4% in static laboratory conditions [40], 11% in a numerical study, and 11% from co-heating based methods when carried out in real buildings and under full outdoor environmental conditions [53]. A further study assessing repeated QUB measurements of the same dwelling's HTC, showed a deviation of 21% between 6 valid measurements [54]. Whilst these studies provide an indication of the reproducibility and self-consistency of measurements, investigations into the underlying causes of uncertainty are limited. Developing an understanding of these issues therefore remains key for both limiting and defining suitable estimates of uncertainty.

To gauge the performance of real dwellings, measurements must be made in an uncontrolled external environment. This inevitably introduces a number of uncertainties which are discussed on an individual basis in sections 3.2 and 3.3.

3.2. Experimental uncertainties in input variables

Initially, uncertainties associated with the measurement and determination of input parameters (i.e. P_h, q_{sw}, T_i, T_e) for equation (1) are reviewed. In all cases, improved experimental techniques can reduce these measurement uncertainties, although not eliminate them entirely.

3.2.1. Measurement of input parameters

Most simply, there are uncertainties associated with each set of sensors used to provide the inputs required in equation (1). For temperature measurements (T_i, T_e), measurement uncertainties may relate to the accuracy of instruments, sensor positioning, influence of radiation and the accuracy of data logging systems [12]. Similarly, the accuracy of metering equipment (P_h) will influence the accuracy of HTC estimates, although as long as suitable equipment is used (including the resolution of meters), the greater risk of uncertainty is likely to be associated with experimental mistakes that can lead to un-metered equipment, metering of equipment outside the building fabric or meter failures. Uncertainties within solar radiation (q_{sw}) measurements include calibration uncertainties, angular uncertainties, overshading and dirt on the sensor. However, as q_{sw} is included as an independent regression variable in the analysis and co-heating test measurements are undertaken in the winter months when q_{sw} is at its lowest, any systematic uncertainties are likely to have a negligible impact upon HTC estimates [55].

In most experimental setups, multiple sensors are typically used to measure P_h and T_i, such that many of these uncertainties will reduce with the number of sensors deployed. As a result, with typically fewer sensors deployed and potentially a higher risk of errors from radiation and micro-climates, in many cases it will be the measurement of T_e that has the most significant impact upon HTC estimates [55].

3.2.2. Internal temperature drifts and fluctuations

The steady-state model assumes constant internal temperatures, although experimentally this can only be approximately achieved. Difficulties may occur in achieving constant temperatures, particularly from solar gains rising temperatures above experimental set points [54]. This can be particularly problematic in low energy dwellings, such as Passivhaus [12,56].

Further, it is important that the analysis is only conducted after a test building has been sufficiently heat soaked and heated to quasi-steady state conditions – i.e. discarding the heating up period where the thermal mass has not fully charged. Identifying this point, however, may not always be obvious from heating loads and temperature traces alone [55], with heat flow into heavier elements potentially missed by air temperature measurements. The addition of heat flux density measurements may assist in identifying when a building has reached equilibrium [55,57].

3.2.3. Achieving uniform internal temperatures

The energy balance described in section 2 assumes uniform internal temperatures and a single zone model of heat transfer. In practice, temperatures are likely to vary between zones, depending upon experimental technique, external conditions and the characteristics of the test dwelling, with ± 1 K thought to be typical [58]. Stratification between floors [54] or poor mixing between restricted zones can act to increase this variation. This can introduce bias if the measured and averaged internal temperature does not match that experienced by all the different heat loss elements, particularly if sensor positions are not representative [6]. Test dwellings in which heat transfer is unevenly distributed across the building fabric may then act to highlight this non-uniformity [10]. Using different internal temperature weightings may vary HTC estimates by 2-8% when internal zonal temperatures vary [55]. Bauwens and Roels [37] state that tailored equipment is indispensable in order to avoid such issues.

3.2.4. Party wall heat transfer

The vast majority of dwellings in most regions have some form of attached neighbouring dwellings. Ideally, these are 'guarded' during the test by heating the neighbouring dwelling to the same internal set-point temperature as the test dwelling [12]. In practice, this can never be perfectly achieved and some heat transfer will inevitably still occur. This can become particularly significant in highly connected dwellings (e.g., apartments, terraced houses) under conditions of poor control (e.g., no access or significant solar gains in adjoining dwellings), or when complex heat pathways exist in the party elements (e.g., convective bypasses). Corrections may be applied to the test data based upon temperature traces or measured heat flux density [12,57]. However, even with these corrections and guarded heat transfer, the magnitude of this uncertainty within apartments has suggested they cannot be tested at scale [18,55].

3.3. Model uncertainties

Models are always abstractions of the natural system, with some less important variables and interactions left out whilst other relationships are given in simplified forms. Within the physical model adopted to represent co-heating tests (section 2), steady-state aggregations and simplified heat flows are both adopted. Uncertainties resulting from such simplifications and approximations are termed 'model uncertainties' and are reviewed below.

3.3.1. Measured solar radiation

Solar radiation measurements will provide an imperfect representation of the solar radiation received across a test building. This model uncertainty may lead to systematic errors in estimations of the HTC. Most significantly, the use of global horizontal radiation can fail to suitably distinguish between overcast and sunny days, causing a significant underestimate in the estimated HTC, as much as 20 W/K [38]. Any systematic error should be reduced by positioning solar radiation sensors in a vertical plane aligned with the facade exposed to the highest expected solar gains [38]. However, this will always be an imperfect representation of the solar radiation across multiple building surfaces with different properties. Whilst vertical measurements are suggested in the protocol set out by Johnston et al. [12], evidence suggests this is not consistently followed [55].

3.3.2. Stored heat

Linear regression models assume the independence of aggregated data points (typically downsampled to 24-hour intervals, to minimise dynamic effects). Diurnal and day-to-day variations in internal temperatures are known to impact short-term tests and overnight methods [59]. More significantly, in longer-term tests heat inputs from solar gains can result in heat being stored from one day to subsequent days [60]. This can lead to underestimates of the HTC and should be minimised through experimental and analysis techniques – principally by testing during appropriate weather conditions and adopting suitable aggregation intervals, e.g., dawn-to-dawn to allow more time for solar radiation effects to attenuate [38,61]. Stamp et al. [38] suggest that changing the aggregation interval can adjust HTC estimates by as much as 15%. This may particularly apply to heavyweight, highly glazed and low energy buildings [61,62], along with tests conducted in sunny periods. Furthermore, the mix of test weather may be important, particularly alternating overcast and sunny days [16,63], or the presence of successive sunny days which may have a high influence over HTC estimates.

3.3.3. Variations due to varying wind and stack pressures

Variations in wind speed and wind pressures may similarly vary infiltration and heat transfer rates across a test period [20]. External air flows are also known to impact external heat transfer coefficients in test cells [64], although this is likely to be less important in full-scale test houses. Similarly, varying indoor-to-outdoor temperature differences may result in changes to infiltration rates via the stack effect. Higher than normal temperature difference during testing was found to increase the HTC by 3–15 W/K for two and three storey houses respectively [65]. These variations are not normally included in regression models and will increase the dispersion of aggregated data points, notably non-linearly [55,66]. In such cases, these variations do not constitute standard measurement uncertainties but rather reflect the variation in the value of the measured parameter, the HTC.

3.3.4. Moisture effects

The presence of excess moisture in the building fabric may cause additional latent heat loads associated with evaporation, resulting in overestimates of the HTC. Within tests on recently completed dwellings, latent loads have been estimated as accounting for 9% [67,68], 10% [66] or between 2-9% [55] of total heating loads across the course of a test. However, such estimates are likely to themselves contain significant uncertainty and it is likely to be a better strategy to avoid moisture loads rather than attempt to correct for them.

3.3.5. Non-direct heat transfer paths

The linear regression model assumes that heat transfer is directly related to the external air temperature. However, non-direct heat transfer paths may exist between a test building and surrounding elements at a different temperature to the ambient air, e.g., ground, attic or crawl spaces, garages, or radiant sky temperatures [21,37]. Depending on how strongly these may be coupled to the external temperature, they may constitute constant loss or gain terms. In many cases, these coupled loss terms may be small and their inclusion as separate terms is unlikely to yield improved results [37,55]. However, test cases exist where up to 35% of the envelope has been coupled to the ground [55], which may impact the HTC estimation and potentially more importantly comparisons to predicted values [69].

3.3.6. Regression errors

Uncertainties may also be associated with the use of given regression techniques on the experimental data. Statistical checks may be performed to identify appropriate regression techniques and subsequently

Table 1
Definition of uncertainties according to JCGM 100 [23] (unless otherwise stated).

Terminology	Description
Measurand	'Quantity intended to be measured.'
Measurement result	'Set of quantity values being attributed to a measurand together with any other available relevant information.'
Measurement uncertainty	'Non-negative parameter characterising the dispersion of the quantity values being attributed to a measurand, based on the information used.' It includes definitional uncertainties and measurement errors.
Definitional uncertainty	'Component of measurement uncertainty resulting from the finite amount of detail in the definition of a measurand.' It represents the 'practical minimum measurement uncertainty achievable in any measurement of a given measurand' even in case no measurement errors have been introduced by the monitoring process.
Natural Variation	As natural systems change in time and place, so do the parameters of interest [72]. It can be considered a form of definitional uncertainty.
True / reference value	Theoretical true value of the measurand. This can never be known and may itself vary (see natural variation and definitional uncertainty) or be influenced by the model used.
Measurement error	'Measured quantity value minus a reference quantity value.'
Systematic measurement error	'Component of measurement error that in replicate measurements remains constant or varies in a predictable manner.'
Random measurement error	'Component of measurement error that in replicate measurements varies in an unpredictable manner.'
Model uncertainty	'Uncertainty due to imperfections and idealizations made in physical model' [72].
Type A evaluation of uncertainty	'Method of evaluation of uncertainty by the statistical analysis of series of observations.'
Type B evaluation of uncertainty	'Method of evaluation of uncertainty by means other than the statistical analysis of series of observations.'
'Standard uncertainty	Uncertainty of the result of a measurement expressed as a standard deviation.'
'Expanded uncertainty	Quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand.'

validate this choice. Linear regression assumes the independence of data points. However, building dynamics and diurnal patters may mean this does not hold true. Time series plots of residuals can be inspected to see if any clear patterns are present. If patterns occur, it may indicate that the linear model is oversimplified and thus inappropriate, and that more complex (or maybe dynamic) models describing additional physical mechanisms should be adopted. Specifically, tests concerning Gaussian residual hypothesis [70] and the lack of autocorrelation [39] shall be undertaken. Else, linear regression techniques formally imply that the residuals are independent of time (homoscedasticity hypothesis). It is good practice to check if the residuals are homoscedastic by plotting them as a function of time. The magnitude of the randomly fluctuating residuals shall not vary with time.

Further, ordinary linear least squares regression assumes uncertainty is only present in the dependent variable. The presence of uncertainty within the independent variable of regression models may lead to attenuation bias and a tendency to underestimate the HTC. Stamp [55] suggests this is likely to be negligible, or only becomes significant when significant error exists in regression variables, which will then dominate the overall uncertainty. Legendre [71] gives various conditions to be checked to validate the possibility of using ordinary linear least squares regression by comparing normalised uncertainties on both dependent and independent variables.

4. Characterising uncertainties

A variety of uncertainties have been reviewed in the previous section, each with various implications for heat transfer estimates, the reproducibility of measurements, and comparisons to either the estimated HTC from in-situ measurements in other buildings or the HTC from design calculations. Therefore, this section looks to provide a framework to categorise different sources of uncertainty. In Table 2, the type of uncertainty and its impact upon HTC estimates are presented, alongside a description of when such uncertainties are likely to be most significant – i.e. for which building characteristics and under what weather conditions.

The impact of each uncertainty can be categorised in two ways. Firstly, in regards to the difference between the experimentally estimated HTC and the theoretical 'true' value of the HTC (see Table 1 for definitions). This measurement error could then be defined as systematic or random in nature. Systematic uncertainties (listed in Table 2) occur from both model uncertainties (e.g., stored heat, solar measurements, assumed linear heat transfer) and experimental uncertainties (e.g., party wall heat transfer, sensor measurement errors, non-constant and non-uniform internal temperatures). Other sources of uncertainty may simply increase the random error in HTC estimates – i.e. the dispersion of daily data points. Examples here include varying infiltration rates or dynamic external temperatures.

Alternatively, the impact of a source of uncertainty can be assessed in regards to its impact when interpreting the results, either between tests or to design values. In such cases, even if the previous measurement uncertainty is negligible – i.e. the difference between estimated and 'true' values do not significantly diverge – uncertainty may remain when comparing between values. For example, successive tests on the same dwelling should not be expected to yield the same results if they take place under different wind conditions and therefore experience different infiltration rates across their respective test periods. Likewise, a design calculation under average infiltration rates will diverge from a field test under significantly different conditions. Without careful additional measurements and modelling of these effects the two cannot be sufficiently reconciled.

It is crucial to recognise that the 'true' HTC is not a constant parameter but rather varies naturally. For example, the HTC is expected to alter with varying infiltration rates, fabric moisture content, long wave sky losses and heat loss to the ground (Table 2). Uusitalo et al. [72] define this type of variation in a measurand as 'natural variation', whilst this can be classed more broadly as 'definitional uncertainty' – i.e. the uncertainty in the definition of the parameter that is trying to be measured (Table 1). In other words, this can be described as the uncertainty in the definition of the HTC when natural variation is not understood. The natural variation does not create a traditional measurement error between the measured and true values. Instead, the result is definitional uncertainty between either tests 1 and 2 performed on the same dwelling or between tests A and B performed on different dwellings when the external environmental conditions vary between tests.

The term 'definitional uncertainty' can be applied to this type of natural variation but also more broadly. For example, uncertainty may exist in the definition of the HTC in terms of the fabric moisture content when a test was conducted or due to degradation of the fabric over time. In a further example, the HTC measured via a co-heating test is defined by a uniform internal temperature, with all external elements therefore exposed to the same internal air temperatures. However, in an occupied dwelling, natural gradients will exist both between and within zones. As such, the two HTCs are definitionally different and cannot

Table 2
Sources of uncertainty, their impacts and when they are likely to occur.

Uncertainty	Description	Type	Impact Upon HTC Estimate	Uncertainty in comparisons to other test / predictions	When
Stored heat	Solar gains stored across 24-hour aggregation intervals.	Model uncertainty (analysis)	Systematic (underestimate)	If systematic error occurs	Sunny, warm periods, heavyweight, highly glazed dwellings. Successive sunny days.
Solar measurement	Imperfect representation of incident solar gains across different surfaces.	Model uncertainty (analysis)	Systematic	If systematic error occurs	E.g., use of horizontal measurement for dwelling with significant south-facing gains.
Dynamic external temperature	Causing dynamic heat flow.	Model uncertainty	Random (increased dispersion of data points)	Not significant in multi-day co-heating tests	High diurnal swings; impact limited to short test periods.
Varying infiltration	Varying infiltration rates from stack and wind pressures.	Model uncertainty (natural variation in true HTC)	Random (increased dispersion of data points)	If test weather differs between tests or from design	Dwellings with high proportion of infiltration losses; periods of high wind/ ΔT.
Sky temperature	Varying long-wave radiation to changing effective sky temperature.	Model uncertainty (natural variation in true HTC)	Random (increased dispersion of data points)	If test weather differs between tests or from design	Varying cloud cover; uninsulated roofs. Influence limited to short test periods and overnight analysis.
Moisture - latent load	Additional heating for latent load.	Model uncertainty (natural variation in true HTC)	Definitional	If building condition differs between tests or from design	Excessive moisture (e.g., new builds with wet finishes).
Moisture - thermal conductivity	Unknown moisture content alters thermal conductivity.	Model uncertainty (natural variation in HTC)	Definitional	If building condition differs between tests or from design	Moisture sensitive constructions (e.g., solid wall, fibrous insulation); new builds with wet finishes.
Party heat transfer	Heat transfer across party walls/ floors.	Experimental uncertainty	Systematic (dependent upon direction of net heat flow)	If systematic error occurs	Apartments, terraced or semi-detached dwellings, particularly where control of adjacent spaces is limited.
Uncoupled temperatures	Heat loss driven by T_{ground} and T_{sky}.	Model uncertainty	Definitional	If uncoupled temperatures vary between test or from design assumptions	Cases with attics, garage, basement, large ground floor area.
Sensor measurement error	Due to sensor error (e.g., calibration).	Experimental uncertainty	Systematic	If systematic error occurs	Inaccurate sensors, influence of radiation, limited number of sensors deployed.
Non-uniform internal temperatures	Representativeness of single average in analysis.	Model uncertainty	Systematic (either direction)	If internal test conditions vary / If temperature weighting different from design	Tight floor plans, high gains/ heat loss zones; insufficient equipment set up.
Operational errors	Due to experimental set up (e.g., elevated T_{i}, fans).	Experimental uncertainty	Systematic	If test conditions (ΔT) differ between tests or to design	T_{set} higher than design T_{i}.
Non-constant internal temperature	Unstable internal temperatures, or deviation from quasi steady-state conditions.	Experimental uncertainty	Systematic (underestimate when warming, overestimate when cooling)	If systematic errors occur	Poor experimental control; solar gains; inclusion of initial warm up period.
Regression errors	Attenuation bias.	Model uncertainty	Systematic (underestimate)	If systematic errors occur	High error in independent variable (Q_{solar} or $Q\Delta T$).

be considered equivalent. No matter the accuracy of the measurement itself, this definitional uncertainty between the two cases will remain.

Table 2 lists when each source of uncertainty is likely to have the most significant impact. It can be seen that heavyweight, highly glazed and low heat loss dwellings are intrinsically more prone to measurement uncertainties. Uncertainties will further increase in sunny test periods and those with high external temperatures. The impact of uncertainties can also vary significantly according to the distribution of weather or more specifically sunny and dull days. An absence of dull days – i.e. those with little solar influence – can increase the influence of solar-related uncertainties. It is for these reasons that both Everett [16] and Lowe and Gibbons [63] recommended that a test should comprise at least two sunny days and two dull days (with no sunny days preceding them). Similarly, sunny days acting as outliers can have significant influence over HTC estimates. A worst-case scenario might be an absence of very dull days along with a pair of successive sunny days, the second of which would be systematically biased by stored heat from the first.

5. Framework for estimating uncertainty

To better evaluate the significance of these uncertainties, and to quantify their contribution to overall uncertainty estimates, a method for estimating uncertainty in the estimated HTC is presented here. This method is then applied to a range of previously conducted tests, with results presented in section 6.2.

5.1. Overall approach to uncertainty estimates

Initially, an approach for estimating measurement uncertainty can be defined by adopting the overall principles defined in JCGM 100 [23] and BSI PD 6461-4 [73], and previously applied to thermal characterisation methods in the PASLINK network [60].

To conduct this uncertainty analysis, first the uncertainty in each input variable needs to be estimated (section 5.1.2). However, it can be suggested that such an approach does not fully incorporate all uncertainties covered in Table 2. As such, the statistical uncertainty – i.e. that determined through the regression analysis – may also be calculated as described in section 5.1.3. Combining this with the previously estimated measurement uncertainty provides an estimated total uncertainty (section 5.1.4).

5.1.1. Process for estimating measurement uncertainty

The general approach is to define the uncertainty (u) in each input variable defined in equation (1) (e.g., $u(T_{\text{i}})$) and create maximum ($T_{\text{i}} + u(T_{\text{i}})$) and minimum ($T_{\text{i}} - u(T_{\text{i}})$) error cases for each variable. Regression analysis as described in section 2 should then be carried out to create maximum and minimum error cases (e.g., $H(T_{\text{i}} + u(T_{\text{i}}))$ and

Table 3
Example estimations of uncertainty in inputs variables.

Input	Error Source	Type	Uncertainty calculation
T_{i}	Calibration	B	$u(T_{\mathrm{i}})_{\mathrm{calib}} = \frac{u_{\mathrm{sensor}}}{\sqrt{n_s}}$
T_{i}	Spatial Variation	A	$u(T_{\mathrm{i}})_{\mathrm{spatial}} = \begin{cases} \sqrt{\frac{n}{n-2}}\,\frac{\sigma(T_{\mathrm{i}})}{\sqrt{n_s}} & \text{if } n < 30 \\ \frac{\sigma(T_{\mathrm{i}})}{\sqrt{n_s}} & \text{if } n > 30 \end{cases}$
T_{e}	Calibration	B	$u(T_{\mathrm{e}})_{\mathrm{calib}} = u_{\mathrm{sensor}}$
P_{h}	Calibration	B	$u(P_{\mathrm{h}})_{\mathrm{calib}} = \frac{u_{\mathrm{sensor}}}{\sqrt{n_s}}$
P_{h}	Party Wall	A	$u(P_{\mathrm{h}})_{\mathrm{pwall}} = \begin{cases} \sum_j^n (q_j\ A_j) & * \\ \left(\left(\frac{u(q_j)}{q_j}\right)^2 + \left(\frac{u(A_j)}{A_j}\right)^2\right)^{\frac{1}{2}} & ** \end{cases}$
			$= \begin{cases} \sum_j^n (U_j\ A_j \dot{\Delta} T_j) & * \\ \left(\left(\frac{u(U_j)}{U_j}\right)^2 + \left(\frac{u(A_j)}{A_j}\right)^2 \left(\frac{u(\Delta T_j)}{\Delta T_j}\right)^2\right)^{\frac{1}{2}} & ** \end{cases}$
q_{sw}	Calibration	B	$u(q_{\mathrm{sw}})_{\mathrm{calib}} = \frac{u_{\mathrm{sensor}}\, q_{\mathrm{sw}}}{\sqrt{3}}$

* = if q uncorrected.
** = if q corrected for party wall heat transfer.

$H(T_{\mathrm{i}} - u(T_{\mathrm{i}}))$. The impact upon estimates of H can be defined by the sensitivity coefficient, e.g., $c(T_{\mathrm{i}})$.

$$c(T_{\mathrm{i}}) = \frac{H(T_{\mathrm{i}} + u(T_{\mathrm{i}})) - H(T_{\mathrm{i}} - u(T_{\mathrm{i}}))}{2\,u(T_{\mathrm{i}})}. \tag{4}$$

Uncertainty from all inputs can then be combined to give the overall measurement uncertainty in H:

$$u(H)_{\mathrm{meas}} = \sqrt{[c(T_{\mathrm{i}})\,u(T_{\mathrm{i}})]^2 + [c(T_{\mathrm{e}})\,u(T_{\mathrm{e}})]^2 + [c(P_{\mathrm{h}})\,u(P_{\mathrm{h}})]^2 + [c(q_{\mathrm{sw}})\,u(q_{\mathrm{sw}})]^2} \tag{5}$$

5.1.2. Estimating uncertainty in input variables

The process described in section 5.1.1 requires the estimation of the uncertainties in the input variables (P_{h}, q_{sw}, T_{i}, T_{e}) to incorporate the different uncertainties in each (previously identified in section 3). In some cases, this can take the form of Type A uncertainty analysis (Table 1), based upon statistical analysis of measurements. Alternatively, some form of expert knowledge or past experience may be required to define estimated uncertainties, taking the form of Type B uncertainty analysis. This may include manufacturer specifications, calibration uncertainties, or assumptions based upon expert knowledge. The complexities of this latter approach are discussed in section 7.2. Example estimations of uncertainty in input variables are provided in Table 3; a uniform distribution is assumed for $u(q_{\mathrm{sw}})_{\mathrm{calib}}$ in this example.

The different error sources for each input variable may be combined in quadrature, given they are independent and uncorrelated, to give the total uncertainty in each input:

$$u(T_{\mathrm{i}}) = \sqrt{u^2(T_{\mathrm{i}})_{\mathrm{spatial}} + u^2(T_{\mathrm{i}})_{\mathrm{calib}}}\,. \tag{6}$$

5.1.3. Statistical uncertainty

Uncertainty may also be estimated as an output of the chosen regression model. Thébault and Bouchié [6] provide details of how the applicability of the vertical ordinary least squares approach may be assessed and provide uncertainty estimates for identified parameters (i.e. the slope and the intercept) depending on the variance of the residuals. For Siviour analysis, statistical uncertainty may be estimated assuming a vertical ordinary least squares approach. Dependency may be reformulated using the Pearson coefficient r^2 and variances on both dependent and independent variables ($P_{\mathrm{h}}/\Delta T$ and $q_{\mathrm{sw}}/\Delta T$). The resulting standard uncertainty is:

$$u(H)_{\mathrm{stat}} = \sqrt{\mathrm{Var}(H)}, \tag{7}$$

where:

$$\mathrm{Var}(H) = \mathrm{Var}(A_{\mathrm{sw}})\,\frac{\sum_{i=1}^n X_i^2}{n}\,. \tag{8}$$

In the case of Siviour analysis, $\frac{\sum_{i=1}^n X_i^2}{n}$ represents the sum of the squared independent variable $\frac{q_{\mathrm{sw}}}{\Delta T}$ across n data points, divided by the number of data points. The variance in the solar aperture is then calculated as:

$$\mathrm{Var}(A_{\mathrm{sw}}) = \frac{\mathrm{Var}(\frac{P_{\mathrm{h}}}{\Delta T})(1 - r^2)}{\mathrm{Var}(\frac{q_{\mathrm{sw}}}{\Delta T})(n-2)}, \tag{9}$$

with r^2 defined by:

$$r^2 = \frac{\mathrm{Cov}(\frac{P_{\mathrm{h}}}{\Delta T}, \frac{q_{\mathrm{sw}}}{\Delta T})^2}{\mathrm{Var}(\frac{P_{\mathrm{h}}}{\Delta T})\mathrm{Var}(\frac{q_{\mathrm{sw}}}{\Delta T})}\,. \tag{10}$$

5.1.4. Total derived uncertainty

The residuals seen in data points, on which the statistical uncertainty is estimated, result from various sources of uncertainty. This includes, for example, some random sensor errors that have already been accounted for within measurement uncertainty estimate. However, statistical uncertainty alone would not account for an outdoor temperature sensor that had a large offset. Conversely, estimates of measurement uncertainty as described in sections 5.1.1 and 5.1.2 may not be able to incorporate all uncertainties reviewed in section 3, or to account for the number, fit and distribution of data points seen in Figs. 1 and 2. Therefore, neither the statistical uncertainty nor the estimated measurement uncertainty fully capture all expected uncertainties. To avoid double counting for some sources of uncertainty, within this paper the statistical and measurement uncertainties are combined in quadrature:

$$u_{\mathrm{tot}} = \sqrt{u_{\mathrm{meas}}^2 + u_{\mathrm{stat}}^2}\,. \tag{11}$$

The estimates should be then expressed as expanded uncertainty across an increased confidence interval (where k depends on the desired confidence level:

$$U_{\mathrm{tot}} = k\,u_{\mathrm{tot}}. \tag{12}$$

Here 95% confidence intervals (approximately $k = 2$, assuming the errors are normally distributed) are judged reasonable and adopted.

6. Application of uncertainty estimates to existing tests

6.1. Description of data set

To evaluate the range of expected uncertainties with co-heating tests and the key components, the approaches to uncertainty estimates detailed in section 5.1 are applied to 14 co-heating tests previously conducted by the authors. Some details of specific cases or larger datasets can be found in previous publications [4,20,55,69,74,75], although anonymity is preserved in some cases here. Relevant summary details can be found in Table 4. The case studies do not constitute a representative sample but, as it can be seen in the table, they provide an overview of a range of different buildings and tests carried out by the authors. Siviour analysis is used, with 24-hour aggregation periods. Previous work has suggested little difference between Siviour and MLR approaches and little benefit in adopting longer aggregation intervals [55]. Therefore, only the Siviour analysis will be reported in the following, for clarity and conciseness.

6.2. Results

Table 5 summarises the estimated HTC for each test, alongside the estimated statistical, measurement and total uncertainty. Both measurement uncertainty and statistical uncertainty are seen to vary significantly between tests. The expanded measurement uncertainty (at 95% confidence intervals, $k = 2$) U_{meas} varies between 1.7 and 23.8 W/K, with an average value of 9.6 W/K or 4.8% of the measured value. Similarly, the statistically estimated uncertainty, U_{stat} varies between 1.0 and 14.7 W/K, with an average value of 7.0 W/K or 6.4%. Added in

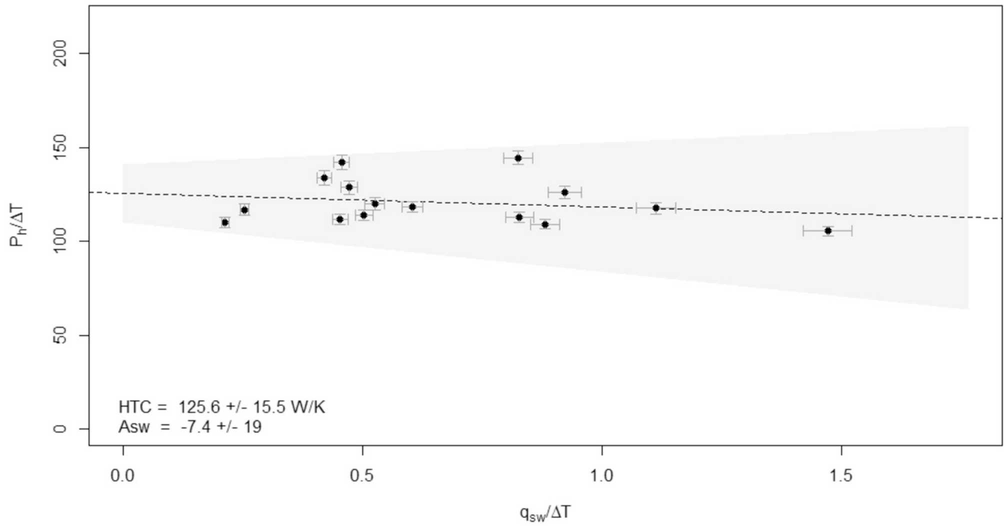

Fig. 1. Siviour analysis and error estimates from the D4 Test. Large statistical uncertainty is related to the horizontal solar measurement.

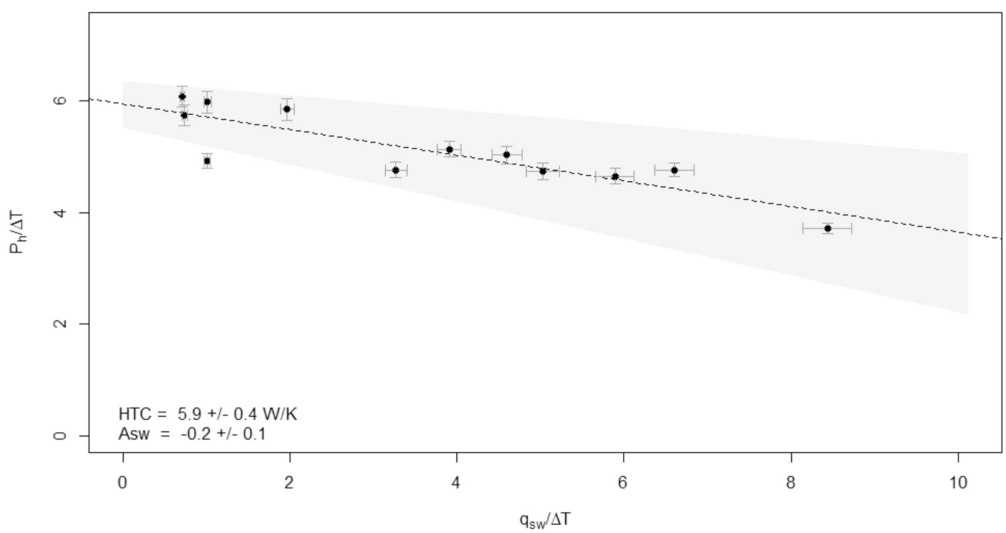

Fig. 2. Siviour analysis of D1 test, where south-facing vertical solar measurements, combined with a mix in dull any sunny days, results in a lower overall uncertainty.

Table 4
Details of buildings and tests used for uncertainty analysis.

ID	Type	Duration	When	Notes
D1	Detached	12 days	Feb	[30]
D2	Detached	15 days	Jan	-
S1a	Semi	9 days	Jan	-
S1b	Semi	12 days	Mar	Repeat of S1a after cavity insulation
D3	Detached	5 days	Jan	Passivhaus [55]
D4	Detached	15 days	Dec	Horizontal solar measurement
T1	End-terrace	34 days	Nov/Dec	Passivhaus bungalow, adjacent to T2 [75]
T2	Mid-terrace	31 days	Nov/Dec	Passivhaus bungalow, adjacent to T1 [75]
T3	End-terrace	13 days	Jan	Passivhaus [75]
S2	Semi-detached	56 days	Oct/Nov/Dec	Adjacent to S3 [75]
S3	Semi-detached	56 days	Oct/Nov/Dec	Adjacent to S2 [75]
D5	Detached	19 days	Feb	Bungalow [75]
S4	Semi-detached	13 days	Mar/Apr	[75]
D6	Detached	11 days	Dec/Jan	Horizontal solar measurement [75]

Table 5
Estimated HTC (using the Siviour analysis) and associated uncertainty estimates (at $k = 2$).

Case	HTC (W/K)	U_{meas} (W/K)	U_{stat} (W/K)	U_{tot} (W/K)	U_{tot} (%)	Predicted (W/K)
D1	71.3	3.8	4.4	5.8	8.2	68.4
DER	227.1	9.6	13.4	16.4	7.2	203.4
S1a	249.8	15.5	14.3	21.1	8.5	83.4
S1b	147.9	8.3	4.2	9.3	6.3	83.4
D3	56.2	3.0	14.7	15.0	26.7	64.6
D4	125.6	6.6	14.0	15.5	12.4	83.8
T1	47.4	2.1	2.6	3.4	7.1	43.4
T2	39.0	1.7	2.6	3.1	8.0	36.6
T3	47.2	1.9	1.0	2.2	4.6	39.6
S2	128.1	5.8	2.9	6.4	5.0	95.1
S3	116.6	5.1	2.9	5.9	5.0	92.6
D5	222.4	9.4	7.2	11.8	5.3	134.9
S4	142.9	7.0	5.8	9.1	6.4	113.3
D6	144.7	8.2	7.3	16.8	11.6	135.0
Average		6.1	7.0	10.1	8.7	

quadrature (according to equation (11)), one of these terms often dominates U_{tot}. The value of U_{tot} is again seen to vary significantly, between

Table 6
Measurement uncertainty in each input variable, along with the respective contribution to uncertainty. Uncertainties and contributions are stated in their expanded form ($k = 2$).

Case	$U(T_i)$ (°C)	$U(T_e)$ (°C)	$U(P_h)$ (W)	$U(q_{sw})$ (W/m^2)	$c.U(T_i)$ (W/K)	$c.U(T_e)$ (W/K)	$c.U(P_h)$ (W/K)	$c.U(q_{sw})$ (W/K)
D1	0.4	0.7	33.1	4.4	-1.8 -2.6%	2.8 3.9%	2.0 2.8%	0 0%
D2	0.3	0.6	87.1	5.7	-3.6 -1.6%	6.9 3.0%	5.7 2.5%	0 0%
S1a	0.7	0.7	208.0	1.2	-9.6 -4.0%	7.4 3.1%	9.2 3.9%	0 0%
S1b	0.5	0.7	123.8	2.3	-3.2 -2.3%	4.5 3.2%	6.4 4.6%	0 0%
D3	0.4	0.7	29.6	0.7	-1.1 -1.9%	2.2 3.9%	1.7 3.0%	0 0%
D4	0.5	0.7	58.5	0.8	-3.4 -2.7%	4.8 3.8%	3.1 2.5%	0 0%
T1	0.2	0.6	23.4	1.8	-0.5 -1.1%	1.6 3.3%	1.3 2.8%	0 0%
T2	0.2	0.6	20.8	1.8	-0.3 -0.8%	1.2 3.2%	1.1 2.9%	0 0%
T3	0.2	0.6	35.2	0.6	-0.4 -0.9%	1.1 2.4%	1.5 3.2%	0 0%
S2	0.3	0.6	66.1	1.5	-1.7 -1.4%	4.2 3.3%	3.5 2.7%	0 0%
S3	0.2	0.6	60.5	1.6	-1.3 -1.1%	3.7 3.2%	3.2 2.7%	0 0%
D5	0.2	0.6	117.9	1.6	-2.4 -1.1%	6.7 3.0%	6.0 2.7%	0 0%
S4	0.2	0.6	141.2	2.6	-1.2 -0.9%	3.5 2.4%	5.9 4.1%	0 0%
D6	0.2	0.6	141.2	2.6	-2.0 -1.4%	3.5 2.4%	5.9 4.1%	0 0%

2.2 and 21.1 W/K, with an average value of 10.1 W/K or 8.7% of the estimated HTC.

The size of uncertainty estimates should be placed in the context of the discrepancies they are trying to detect. The average uncertainty estimate of 13% compares to an average difference between measured and predicted HTCs of 36%, with 10 predicted HTCs sitting outside of the uncertainty bands of the estimated HTC and four showing good agreement and sitting within. This will be discussed in more detail in Section 7.1.

These results demonstrate the significant variability in uncertainty estimates associated with a given test. Underlying reasons for this can initially be explored by examining components of measurement uncertainty. In Table 6, the measurement uncertainty in each input variable is reported alongside the associated contribution to uncertainty (equation (4)). It is worth noting that the contribution to uncertainty from q_{sw} is always zero, as measurement uncertainty as applied via equation (4) does not impact HTC estimates with q_{sw} included as an independent regression variable. Uncertainties associated with global solar irradiance, including stored solar gains and model uncertainties associated with the measurements relationship to solar gains are however significant. Difficulties in characterising them are covered in section 7.2.

Particular cases may be picked out. For example, S1a has a contribution to uncertainty from T_i of -9.6 W/K (or 4.0% of the estimated HTC). In this case, the accuracy of sensors and number of sensors deployed is similar to other tests. Therefore, this is largely due to poorer internal air mixing. The average standard deviation of internal air temperatures was reported as 1.8 °C across the test period, compared to an average of 0.7 °C from all tests. Here, poor experimental control is being penalised in the uncertainty estimate.

Amongst tests with party walls (S1a, S1b, S2, S3, T1, T2) contributions to uncertainty from party walls can be higher, with additional uncertainty associated with uncontrolled heat loss/gain across these party walls or floors. However, only a small increase in the contribution of uncertainty from P_h is seen in these cases. This indicates that with careful experimental control, in carefully guarded, semi-detached dwellings this party wall heat transfer may not be as significant as other contributions to uncertainty.

For some of the cases reported in Table 5 the estimated statistical uncertainty is significantly higher than the measurement uncertainty (e.g., D3, D4, D6). D3, at just 5 days, is one of the shortest tests. This fact, combined with mainly dull days (maximum daily q_{sw} = 14.6 W/m^2, compared to average maximum of 86.9 W/m^2) leads to significantly higher statistical uncertainty. In other cases (e.g., D4, D6), the solar radiation measurement was made in a horizontal plane. This, along with potentially leading to an underestimate of the HTC, also tends to lead to a higher statistical uncertainty (see Fig. 1). In some cases, statistical uncertainty is very low (e.g., D1). In this case a wide range in solar radiation and external temperature were captured over 12 days, resulting in lower overall uncertainty (Fig. 2).

7. Discussion

7.1. Comparisons of predicted and measured values

In most cases, a predicted value, based upon a suitable calculation methodology (e.g., the Standard Assessment Procedure, SAP [76]; the Passive House Planning Package, PHPP [77]), will be compared with the value estimated from the measurements. This comparison forms the basis of any estimated building fabric performance gap [4]. However, fundamental differences between these two values may exist, suggesting that such a comparison is not equivalent.

Firstly, given the natural variation in the HTC and the relatively short window in which testing takes place, environmental conditions during the test period may not be equivalent to those used in the calculated value (e.g., wind speeds and infiltration losses, temperature gradients and stack losses, ground and apparent sky temperatures). Additionally, re-calculating the HTC to match the conditions experienced during the measurement is not straightforward and requires greater knowledge of the building. For example, a test may be conducted under windier conditions than those used in the calculation of a predicted HTC. However, without fully understanding the relationship between the building's HTC and wind speed, it remains challenging to try and re-align the measured and predicted values to the same boundary conditions.

On top of this, the calculated value may itself contain significant uncertainties, errors or modelling assumptions. This is often termed the 'prediction gap'. For example, audits of SAP assessments have found considerable proportions of incorrect inputs for U-values and thermal bridges [31], or that assumed thermophysical values mis-represent actual performance [3]. Product substitution and on-site as-built difference may also be missed from calculations. Modelling of heat transfer through thermal bridges, ventilated spaces (e.g., attics) and infiltration may then be simplified and represent significant uncertainty in the calculated value. On-site observations and measurements might lead to more 'informed' predictions [4], providing a more meaningful comparison; although it is likely that these issues will remain to some degree.

It is important to separate out the underlying reasons for differences between measured and predicted values. This difference may relate to a) natural variation in the HTC; b) definitional differences between the predicted HTC and the measured building (e.g. predictions are made under different wind conditions to the measurement); c) uncertainties in the predicted value; or d) differences associated with on-site performance. It is only the latter that is associated with a 'fabric performance gap', meaning that care is required in interpreting results.

Whilst full consideration of uncertainties in predicted HTCs is outside the scope of the paper, it is important to note that any assessment of the fabric performance gap is a function of both the uncertainties in the predicted and measured values. There is little effect in developing a highly accurate test when the predictions used for comparisons remain so uncertain themselves.

7.2. Excluded uncertainties

The approach detailed in section 5 does not capture all uncertainties that may be present in a given HTC estimate. Some sources of uncertainty in Table 6 remain challenging to quantify and doing so would require significant additional measurements or more complex models. An approach could be to further adopt Type B uncertainty analysis, assigning default uncertainties from expert knowledge or previous experience - an approach taken in the ISO 9869-1 [34] standard. This could incorporate uncertainties from excess latent loads, stored solar heat gains and solar measurement errors into uncertainty budgets. However, given the current lack of evidence, these assumed uncertainties may be hard to establish and justify. Furthermore, links between uncertainty and experimental technique, building characteristics and weather conditions are likely to mean that this 'one-size fits all' approach may be limited.

7.3. Relation to dynamic test methods

Dynamic methods may offer advantages over static co-heating tests, particularly over test length. However, many of the uncertainties described for static co-heating tests will equally apply. In both types of method there are uncertainties with sensor accuracy, temperature distributions and party walls. Uncertainties related to stored heat should be reduced due to the dynamic approach - although some previous studies have stated the importance of preceding weather conditions before short-term tests [59]. The uncertainty and bias related to model identification should be estimated, for example following the methodologies in [6,39]. Additionally, some definitional uncertainties (e.g., moisture content, infiltration rate, sky losses) will remain, although they may show more variation given the overall shorter measurement period.

Another disadvantage associated with some of the dynamic tests is that an indication of the true HTC of the building is required in order to be able to adequately size the space heating load required to undertake the test. If the space heating load is not sized appropriately, then the test may be invalid.

7.4. Relation to measurements in occupied dwellings with smart meters

In-use HTC estimates, via smart meters (sometimes known as SMETER measurements [46]), might allow the estimation of HTC on a far wider scale than dedicated co-heating tests. However, many uncertainties described here will apply equally (or to an even greater degree) in such tests and the in-use approach may introduce further uncertainties. As in-use methods offer less experimental control (e.g., due to occupancy) and dedicated local measurements (e.g., to reduce intrusiveness), it may be expected that uncertainties related to sensor accuracy, solar gains and temperature uniformity all increase. Additionally, definitional uncertainties over the state of the dwelling will remain (e.g., moisture content and latent loads) and in some cases will be much larger (e.g., from varying and unknown ventilation rates). Finally, given typical internal temperature distributions (e.g., affected by stratification and ventilation practices), heat loss through the fabric is weighted in a different way to the well-mixed co-heating test. For reasons such as these, it is important to note that an in-use and a co-heating derived HTC are fundamentally different, and should not be compared.

8. Conclusion

HTC estimates via co-heating tests have provided key evidence of a fabric performance gap. However, there has been little research into the associated uncertainties. This has limited the application of the co-heating method and leaves results to date stated without the context of their respective uncertainty. This paper has addressed these two issues, firstly by reviewing and characterising uncertainties. Secondly, by devising and applying a method for estimating uncertainty to 14 tests, revealing typical ranges of uncertainty and significant sources of uncertainty in steady-state HTC estimates.

Estimated total uncertainties ranged between 2.2-21.1 W/K (or 4.6-26.5%), with a mean value of 10 W/K (9%). To put this into context, the average difference between measured HTC has been found to be 60% times greater than the predicted HTC [74]. It would therefore appear that the current co-heating method may suitably distinguish underperforming building fabrics and give a suitable estimate of the size of the fabric performance gap. Comparisons between different dwellings or retrofit improvements where the expected difference is similar to the expected uncertainty of the test (approx. 10 W/K) may however be difficult to quantitatively measure.

The natural variation of the HTC (due for instance to background air infiltration rates that may vary day-to-day because of wind velocity) and inaccuracy of design calculations may mean there is little benefit in more accurate measurements. There is likely little use in measurement methods that can estimate the HTC to within 2% if the true value of the HTC is itself varying by a greater amount or that the predicted value can only be known to the nearest 10%.

Results suggest that unsuitable weather conditions (e.g., too warm or sunny or an insufficient mix in weather), challenging buildings (e.g., highly glazed, heavyweight, significant party walls) and poor experimental technique (e.g., non-uniform internal temperatures, low accuracy for external temperature measurements) can all form significant difficulties within given tests. Given suitable buildings are tested under suitable weather conditions and by experienced technicians, HTC estimates through steady-state co-heating are likely to continue to provide a useful tool in assessing and understanding real-world building fabric performance.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Virginia Gori reports financial support was provided by Engineering and Physical Sciences Research Council. Remi Bouchie reports financial support was provided by PROFEEL French National Funding Programme.

David Johnston reports financial support was provided by Technology Strategy Board (Innovate UK).

Data availability

The authors do not have permission to share data.

Acknowledgements

This research was made possible by support from the EPSRC UCL Doctoral Prize Fellowship [grant number EP/N509577/1] and the SEREINE Project from PROFEEL French National Funding Programme. The authors gratefully acknowledge the funding provided by the Technology Strategy Board (now Innovate UK) as part of its Building Performance Evaluation Programme. This funding was used to undertake the in-depth performance tests on a number of the case study dwellings identified within this paper.

References

[1] S. Doran, Field investigations of the thermal performance of construction elements as built, Technical Report, Building Research Establishment, 2000.
[2] P. Baker, Technical Paper 10: U-values and traditional buildings-in situ measurements and their comparisons to calculated values, Technical Report, Historic Scotland, Glasgow Caledonian University, Glasgow, 2011.
[3] F. Li, A. Smith, P. Biddulph, I.G. Hamilton, R. Lowe, A. Mavrogianni, E. Oikonomou, R. Raslan, S. Stamp, A. Stone, A. Summerfield, D. Veitch, V. Gori, T. Oreszczyn, Solid-wall U-values: heat flux measurements compared with standard assumptions, Build. Res. Inf. 43 (2015) 238–252, https://doi.org/10.1080/09613218.2014.967977.
[4] D. Johnston, D. Miles-Shenton, D. Farmer, Quantifying the domestic building fabric 'performance gap', Build. Serv. Eng. Res. Technol. 36 (2015) 614–627, https://doi.org/10.1177/0143624415570344.
[5] R. Bouchie, F. Alzetto, A. Brun, S. Boisson, S. Thebault, Short methodologies for in-situ assessment of the intrinsic thermal performance of the building envelope, in: Proceedings of the Sustainable Places, Nice 2014, 2014, pp. 1459–1477.
[6] S. Thébault, R. Bouchié, Refinement of the ISABELE method regarding uncertainty quantification and thermal dynamics modelling, Energy Build. 178 (2018) 182–205, https://doi.org/10.1016/j.enbuild.2018.08.047.
[7] K. Subbarao, Primary and secondary terms-analysis and renormalization: a unified approach to building and energy simulations and short-term testing—a summary, Technical Report SERI/TR-254-3347, Solar Energy Research Institute, Colorado, USA, 1988.
[8] K. Subbarao, J. Burch, C. Hancock, A. Lekov, J. Balcomb, Short-Term Energy Monitoring (STEM): Application of the PSTAR method to a residence in Fredericksburg, Virginia, Technical Report, Solar Energy Research Inst., Golden, CO (USA), 1988.
[9] K. Subbarao, J. Burch, J. Balcomb, Short-term energy monitoring summary of results from four houses, ASHRAE Trans. 96 (1) (1990) 1459–1477.
[10] E. Mangematin, G. Pandraud, D. Roux, Quick measurements of energy efficiency of buildings, C. R. Phys. 13 (2012) 383–390, https://doi.org/10.1016/j.crhy.2012.04.001.
[11] J. Wingfield, D. Johnston, D. Miles-Shenton, M. Bell, Whole House Heat Loss Test Method (Coheating), [Internet] Leeds, Leeds Metropolitan University, Technical Report, Leeds Metropolitan University, 2010.
[12] D. Johnston, D. Miles-Shenton, D. Farmer, J. Wingfield, Whole house heat loss test method (Coheating), Technical Report, Leeds Metropolitan University, 2013.
[13] R. Socolow, The twin rivers program on energy conservation in housing: highlights and conclusions, Energy Build. 1 (1978) 207–242.
[14] R.C. Sonderegger, P.E. Condon, M.P. Modera, In-situ measurements of residential energy performance using electric co-heating, in: ASHRAE Transactions, vol. 86, 1979 (I), 1979.
[15] J. Siviour, Experimental thermal calibration of houses, in: Conference on Comparative Experimentation of Low Energy Houses, 1981.
[16] R. Everett, Rapid thermal calibration of houses, Technical Report, Open University Energy Research Group, Milton Keynes, 1985.
[17] TSB, Building Performance Evaluation, Domestic Buildings - Guidance for Project Execution, Technical Report, The Technology Strategy Board, Swindon, 2010.
[18] GHA, GHA Monitoring Programme 2009-2011: Technical Report - Results from Phase 1: Post-construction testing of a sample of highly sustainable new homes, Technical Report, Good Homes Alliance, 2011.
[19] J. Palmer, D. Godoy-Shimizu, A. Tillson, I. Mawditt, Building performance evaluation programme: Findings from domestic projects - Making reality match design, Technical Report, Innovate UK, 2016.
[20] D. Miles-Shenton, J. Wingfield, R. Sutton, M. Bell, Temple Avenue Project – Summary Report, Technical Report, Leeds Metropolitan University, 2011.
[21] E. Latif, M. Lawrence, A. Shea, P. Walker, In situ assessment of the fabric and energy performance of five conventional and non-conventional wall systems using comparative coheating tests, Build. Environ. 109 (2016) 68–81, https://doi.org/10.1016/j.buildenv.2016.09.017.
[22] D. Glew, D. Johnston, D. Miles-Shenton, F. Thomas, Retrofitting suspended timber ground-floors, comparing aggregated and disaggregated evaluation methods, Build. Res. Inf. (2019), Publisher: Routledge.
[23] J.C.G.M. 100, Evaluation of measurement data - Guide to the expression of uncertainty in measurement, Technical Report, Joint, Committee for Guides in Metrology, 2008.
[24] Archimetrics Ltd, The SPAB Building Performance Survey - Final Report, Technical Report, The Society for the Protection of Ancient Buildings (SPBA), London, UK, 2019.
[25] F. Seguro, Insights from Social Housing Projects - Building Performance Evaluation Meta-Analysis, Technical Report, Innovate UK, 2015.
[26] A. Rudd, D. Bergey, Ventilation System Effectiveness and Tested Indoor Air Quality Impacts, Technical Report, U.S. Department of Energy, 2014.
[27] N. Brelih, O. Seppänen, Ventilation rates and IAQ in European standards and national regulations, in: The Proceedings of the 32nd AIVC Conference and 1st TightVent Conference in Brussels, Brussels, 2011, pp. 12–13.
[28] R. Jack, D. Loveday, D. Allinson, K. Lomas, First evidence for the reliability of building co-heating tests, Build. Res. Inf. 46 (2018) 383–401, https://doi.org/10.1080/09613218.2017.1299523.
[29] V. Gori, C. Elwell, Estimation of thermophysical properties from in-situ measurements in all seasons: quantifying and reducing errors using dynamic grey-box methods, Energy Build. 167 (2018) 290–300, https://doi.org/10.1016/j.enbuild.2018.02.048.
[30] D. Butler, A. Dengel, NHBC review of co-heating test methodologies, Technical Report, National House Building Council Foundation, 2013.
[31] ZCH, Closing the gap between design & as-built performance - Evidence review report, Technical Report, Zero Carbon Hub, 2014.
[32] S. Stamp, H. Altamirano-Medina, R. Lowe, Assessing the relationship between measurement length and accuracy within steady state co-heating tests, Buildings 7 (2017) 98, https://doi.org/10.3390/buildings7040098.
[33] A. Law, J. Wong, Application of the Co-heating test in Australian climates, Energy Build. 224 (2020) 110153, https://doi.org/10.1016/j.enbuild.2020.110153.
[34] ISO 9869-1, Thermal insulation – Building elements – in-situ measurement of thermal resistance and thermal transmittance. Part 1: Heat flow meter method, 2014.
[35] ISO 9869-2, Thermal insulation – Building elements – in-situ measurement of thermal resistance and thermal transmittance. Part 2: Infrared method for frame structure dwelling, 2018.
[36] EN ISO 13789, Thermal performance of buildings. Transmission and ventilation heat transfer coefficients. Calculation method, 2017.
[37] G. Bauwens, S. Roels, Co-heating test: a state-of-the-art, Energy Build. 82 (2014) 163–172, https://doi.org/10.1016/j.enbuild.2014.04.039.
[38] S. Stamp, H. Altamirano-Medina, R. Lowe, Measuring and accounting for solar gains in steady state whole building heat loss measurements, Energy Build. 153 (2017) 168–178, https://doi.org/10.1016/j.enbuild.2017.06.063.
[39] H. Madsen, P. Bacher, G. Bauwens, A.-H. Deconinck, G. Reynders, S. Roels, E. Himpe, G. Lethé, Thermal performance characterization using time series data – iea ebc annex 58 guidelines, 2015.
[40] G. Pandraud, R. Fitton, Qub: Validation of a rapid energy diagnosis method for buildings, Free Paper presented at IEA–ECBCS–Annex, 2013.
[41] F.P. Hollick, V. Gori, C.A. Elwell, Thermal performance of occupied homes: a dynamic grey-box method accounting for solar gains, Energy Build. 208 (2020) 109669.
[42] D. Farmer, D. Johnston, D. Miles-Shenton, Obtaining the heat loss coefficient of a dwelling using its heating system (integrated coheating), Energy Build. 117 (2016) 1–10, https://doi.org/10.1016/j.enbuild.2016.02.013.
[43] J. Chambers, T. Oreszczyn, Deconstruct: a scalable method of as-built heat power loss coefficient inference for UK dwellings using smart meter data, Energy Build. 183 (2019) 443–453, https://doi.org/10.1016/j.enbuild.2018.11.016.
[44] A. Rasooli, L. Itard, Automated in-situ determination of buildings' global thermo-physical characteristics and air change rates through inverse modelling of smart meter and air temperature data, Energy Build. 229 (2020) 110484, https://doi.org/10.1016/j.enbuild.2020.110484.
[45] R. Jack, Crossway Performance Monitoring, Technical Report, Built Test Solutions, 2021.
[46] A. Allinson, B. Roberts, K. Lomas, D. Loveday, C. Gorse, A. Hardy, F. Thomas, D. Miles-Shenton, D. Johnston, D. Glew, K. Rakhshanbabanari, F. Fylan, D. Farmer, C. Elwell, J. Crawley, F. Hollick, J. Wingfield, G. Roberts, L. Reevell, Technical Evaluation of SMETER Technologies (TEST) Project, Technical Report, UK Department for Business, Energy and Industrial Strategy (BEIS), 2022.
[47] R. Jack, Crossway performance monitoring, https://www.buildtestsolutions.com/case-studies/crossway-performance-monitoring, 2022.
[48] F. Alzetto, D. Farmer, R. Fitton, T. Hughes, W. Swan, Comparison of whole house heat loss test methods under controlled conditions in six distinct retrofit scenarios, Energy Build. 168 (2018) 35–41, https://doi.org/10.1016/j.enbuild.2018.03.024.
[49] A. Marshall, R. Fitton, W. Swan, D. Farmer, D. Johnston, M. Benjaber, Y. Ji, Domestic building fabric performance: closing the gap between the in situ measured and modelled performance, Energy Build. 150 (2017) 307–317.

[50] B. Lloyd, T. Bishop, M. Callau, Retrofit alternatives for State Houses in Cold Regions of New Zealand, Technical Report 2, University of Otago, 2007.

[51] J. Andrews, Electric coheating as a means to test duct efficiency: a review and analysis of the literature, Technical Report, Brookhaven National Laboratory, 1995.

[52] R. Judkoff, J. Balcomb, C. Hancock, G. Barker, K. Subbarao, Side-by-side thermal tests of modular offices: a validation study of the STEM method, Technical Report, National Renewable Energy Laboratory, 2000.

[53] F. Alzetto, D. Gossard, G. Pandruad, Mesure rapide du coefficient de perte thermique des bâtiments, in: Congrès Sciences et Technique Ecobat, 2014, pp. 19–20, Library Catalog, docplayer.fr.

[54] R. Bouchie, F. Alzetto, A. Brun, C. Weeks, M. Preece, M. Waseem Ahmad, M. Sisinni, Methodologies for the assessment of intrinsic energy performance of buildings' envelop, Technical Report D1.2, Centre Scientifique et Technique du Bâtiment (CSTB), 2015.

[55] S. Stamp, Assessing uncertainty in co-heating tests: Calibrating a whole building steady state heat loss measurement method, Doctoral thesis, University College London, 2016.

[56] V. De Meulenaer, J. Van der Veken, G. Verbeeck, H. Hens, Comparison of measurements and simulations of a "passive house", in: Ninth International IBPSA Conference, Montreal, Canada, 2005, p. 8.

[57] G. Bauwens, S. Roels, On site thermal performance characterization of building envelopes: how important are heat exchanges with neighbouring zones, Energy Proc. 132 (2017) 339–344, https://doi.org/10.1016/j.egypro.2017.09.744.

[58] D. Farmer, C. Gorse, W. Swan, R. Fitton, M. Brooke-Peat, D. Miles-Shenton, D. Johnston, Measuring thermal performance in steady-state conditions at each stage of a full fabric retrofit to a solid wall dwelling, Energy Build. 156 (2017) 404–414, https://doi.org/10.1016/j.enbuild.2017.09.086.

[59] M. Liu, D. Claridge, Improving Building Energy Performance Through Continuous Commissioning, Technical Report, Energy Systems Laboratory, Texas A & M, Texas, US, 1995.

[60] P. Baker, H. van Dijk, PASLINK and dynamic outdoor testing of building components, Build. Environ. 43 (2008) 143–151, https://doi.org/10.1016/j.buildenv.2006.10.009.

[61] D. Alexander, H. Jenkins, The validity and reliability of co-heating tests made on highly insulated dwellings, Energy Proc. 78 (2015) 1732–1737, https://doi.org/10.1016/j.egypro.2015.11.282.

[62] O. Guerra-Santin, C. Tweed, H. Jenkins, S. Jiang, Monitoring the performance of low energy dwellings: two UK case studies, Energy Build. 64 (2013) 32–40, https://doi.org/10.1016/j.enbuild.2013.04.002.

[63] R. Lowe, C. Gibbons, Passive solar houses: availability of weather suitable for calibration in the UK, Build. Serv. Eng. Res. Technol. 9 (1988) 127–132, https://doi.org/10.1177/014362448800900306.

[64] P. Baker, IQ-test—improving quality in testing and evaluation of solar and thermal characteristics of building components, Energy Build. 36 (2004) 435–441, https://doi.org/10.1016/j.enbuild.2004.01.046.

[65] R. Lowe, J. Wingfield, M. Bell, J. Bell, Evidence for heat losses via party wall cavities in masonry construction, Build. Serv. Eng. Res. Technol. 28 (2007) 161–181, https://doi.org/10.1177/0143624407077196.

[66] G. Bauwens, In Situ Testing of a Building's Overall Heat Loss Coefficient - Embedding Quasi-stationary and Dynamic Tests in a Building Physical and Statistical Framework, Doctoral thesis, KU Leuven, 2015.

[67] BBRI, Study and monitoring of the PLEIADE dwelling - Final report, Technical Report, Belgian Building Research Institute, 1997.

[68] M. Santamouris, Energy Performance of Residential Buildings: A Practical Guide for Energy Rating and Efficiency, Earthscan, 2005.

[69] D. Johnston, M. Siddall, The building fabric thermal performance of passivhaus dwellings—does it do what it says on the tin?, Sustainability 8 (2016) 97, https://doi.org/10.3390/su8010097.

[70] ISO 5479, Statistical interpretation of data – tests for departure from the normal distribution, 1997.

[71] P. Legendre, Model II Regression User's Guide, R edition, 2018.

[72] L. Uusitalo, A. Lehikoinen, I. Helle, K. Myrberg, An overview of methods to evaluate uncertainty of deterministic models in decision support, Environ. Model. Softw. 63 (2015) 24–31, https://doi.org/10.1016/j.envsoft.2014.09.017.

[73] BSI PD 6461-4, General metrology. Practical guide to measurement uncertainty, 2004.

[74] D. Johnston, D. Farmer, D. Miles-Shenton, Quantifying the aggregate thermal performance of UK holiday homes, Build. Serv. Eng. Res. Technol. 38 (2017) 209–225, https://doi.org/10.1177/0143624416681614.

[75] D. Johnston, D. Farmer, M. Brooke-Peat, D. Miles-Shenton, Bridging the domestic building fabric performance gap, Build. Res. Inf. 44 (2016) 147–159.

[76] BRE, SAP 2012 the government's standard assessment procedure for energy rating of dwellings 2012 edition, Technical Report, Department of Energy and Climate Change, 2014.

[77] W. Feist, Z. Bastian, W. Ebel, E. Gollwitzer, J. Grove-Smith, O. Kah, B. Kaufmann, B. Krick, R. Pfluger, J. Schnieders, Passive House Planning Package Version 9, the Energy Balance and Design Tool for Efficient Buildings and Retrofits, Technical Report, Passive House Institute, 2015.