UK Government

GHG SMETER Project Final Report

The application and assessment of SMETER methods in determining the change in thermal performance of homes retrofitted through the Green Homes Grant

August 2025

Acknowledgements

The work was undertaken by Dr Frances Hollick, Prof Cliff Elwell, Dr Jessica Few, Dr Clare Hanmer and Dr Jez Wingfield, all of whom contributed to this report. The team are all members of the Buildings Theme in the UCL Energy Institute. We would like to thank the participant households for enabling this research to be undertaken.

Any enquiries regarding this publication should be sent to us at:
EnergyResearch@energysecurity.gov.uk

Contents

Overview of project

This report summarises the findings of the GHG-SMETER project, applying Smart Meter Enabled Thermal Efficiency Rating (SMETER) methods to investigate the performance of properties retrofitted through the Green Homes Grant (GHG) schemes and to evaluate the ability of SMETER methods to characterise these.

Empirically determined building thermal performance, such as that estimated using SMETER methods, is particularly valuable in assessing the impacts of retrofits as it both captures the in-use performance and avoids issues resulting from the performance gap (Wingfield, Bell, Miles-Shenton, South, & Lowe, 2009; Hollick, Gori, & Elwell, 2020). The in-use performance of buildings, as opposed to that measured in an aggregate heat loss (co-heating) test or calculated from U-values, reflects the true energy consumption changes on retrofit and can account for occupant takeback (increasing demand after retrofit) (Deurinck, Saelens, & Roels, 2012) or other similar effects. Using an empirical method also ensures that any degradation of materials or installation issues are captured in the performance estimates. As such, SMETER methods are ideally placed to assess the change on retrofit.

The analysis element of the project ran from June 2021 to April 2024, aiming to investigate the thermal performance of homes retrofitted though both the GHG Vouchers Scheme (GHG-V) and the Local Authority Delivery Scheme (GHG-LAD). The principal analysis was undertaken using data from winter 2021/22 with ongoing data collection enabling further research to be undertaken, including an additional sub-project to help address some discontinuities between SMETER results and those of the benchmark model.

The GHG-SMETER project assessed the energy performance of homes using two distinct approaches: Type A & Type B SMETER methods. Type A methods utilise only smart meter gas and electricity data and weather data, which may be supplemented by contextual data from a building and occupant survey. Type B methods require smart meter gas and electricity data and weather data, with additional temperature measurement inside the home in one or more locations; these measurements may again be supplemented by contextual data from a building and occupant survey. They are not utilised in this project, but there are also Type C SMETERs which require more involved data collection.

The GHG-V scheme was closed to new applications on 31st March 2021. GHG-V retrofits were completed before winter 2021/22 preventing collection of in home data prior to install and these homes were therefore studied using Type A SMETER methods, which require smart meter and weather data only. The Smart Energy Research Lab (SERL) can access smart meter data from the date of participant consent to data access, and for up to 12 months prior to consent (limited by the date of install of the smart meter and requiring continuous occupation by the same household). 2458 households were recruited from the GHG-V scheme to participate in this study.

Type B SMETER methods were studied within a cohort of homes where four temperature sensors were deployed in each property in specified locations (living room, master bedroom, kitchen, hallway (or by thermostat)), recruited from GHG-LAD in early 2022. These homes also consented to share smart meter data as part of this research. 105 homes consented to participate in this work and have internal temperatures measured.

The GHG Scheme and SMETER programme

The Green Homes Grant schemes provided retrofit energy efficiency measures and low carbon heating to homes in England. The GHG-SMETER project piloted the use of SMETER methods to support policy evaluation by assessing the impacts of these retrofit measures using before and after in-use thermal performance estimates. The project has applied multiple SMETER methods to a large sample of properties, to provide insights into not only the thermal performance change upon retrofit, but also into the opportunities, challenges, and limitations of such methods. Both Type A and Type B SMETER methods are utilised in this project. The project builds on insights from the BEIS SMETER Innovation Competition, research literature, and International Energy Agency Energy in Buildings and Communities (IEA EBC) Annex 71 into the potential for in-use metrics to support decarbonisation by providing insights of the real thermal performance of properties at an accuracy suitable to support a range of potential end uses.

SMETER methods assess the in-use thermal performance of a building through estimation of the heat transfer coefficient (HTC). This parameter describes the rate at which heat is lost from a building per degree temperature difference between the internal and external temperatures. An in-use HTC reflects the efficiency of the building envelope, incorporating fabric and ventilation heat losses. The HTC is akin to the heat loss parameter calculated in SAP, multiplied by the floor area, and also the value calculated by a co-heating test (Wingfield, Johnston, Miles-Shenton, & Bell, 2010).

There were two parts of the Green Homes Grant: voucher and local authority delivery (LAD) schemes. Homeowners and landlords could apply for a voucher to cover up to two-thirds of the cost of retrofit or low carbon heating measures, up to a maximum of £5000. The voucher scheme opened in 2020 and was withdrawn in March 2021. The LAD scheme launched in August 2020, and funded retrofits and low carbon heating for low income households in England. The LAD scheme was conducted by several geographically-determined Hubs across England. Households from both schemes were recruited into the GHG-SMETER project.

The Green Homes Grant provided funding for energy saving retrofits split into primary and secondary measures; at least one primary measure had to be installed with each grant awarded in the voucher scheme. The measures are further split between insulation, low carbon heat, windows, doors and draughtproofing, and heating controls and tank insulation. SMETERs are currently designed to assess the performance of the building fabric, therefore changes due to measures for low carbon heat, controls, and improvements to hot water tanks are not accurately assessable by the same metric. However, alternative measures of building performance are possible and are discussed in *SMETER methods* (p24)0. Primary fabric measures that were fundable through the scheme are:

solid wall insulation (internal or external)
cavity wall insulation
under-floor insulation (solid floor, suspended floor)
loft insulation
flat roof insulation
pitched roof insulation
room in roof insulation
insulating a park home

The secondary measures that affect the thermal performance of the building fabric are those associated with windows, doors and draughtproofing:

draughtproofing
double or triple glazing (where replacing single glazing)
secondary glazing (in addition to single glazing)
external energy efficient replacement doors (replacing single glazed doors, or doors installed before 2002)

Implementation of the LAD scheme differed from the Vouchers scheme, where in the former some secondary measures were installed without a primary measure, and solar PV was also available to these households; the LAD scheme is further discussed below (p17 & p21) and the impact of PV panels on in-use HTCs is evaluated in the assessment of the unmetered gains on SMETER reliability (p53).

Research context

The GHG-SMETER project was conceived after announcement of the GHG, aiming to exploit the opportunity offered by the scheme to deploy SMETERs at scale and learn more about their implementation. It is usually difficult and expensive to conduct standalone controlled studies at this scale, so aligning research with schemes like GHG is a way to generate evidence and address research questions quickly and at low-cost. However, this pragmatism sacrifices some of the control that researchers have in standalone studies because priority is naturally given to delivering the planned scheme. Research must be conducted within the constraints of the scheme and this can limit the outcomes of the research.

The GHG-SMETER project was successful, deploying SMETER methods at much greater scale than previous studies and providing valuable new insights on how SMETER methods perform in determining the thermal properties of homes that qualify for funding under such schemes, as well the impact of retrofit through the GHG. However, readers should note that there were some important constraints on the research. Most notably, the conditions and timelines under which data was gathered from homes in this project are not necessarily typical of what providers of SMETER methods would ideally use and there were also some gaps in the information available on the retrofits undertaken. These constraints and their implications are outlined in more detail in the report. It should be noted that these practical constraints and challenges of deploying sensors and collecting representative measurement data is a challenge that remains to be fully addressed.

Nonetheless, this research represents a significant addition to the evidence base on methods for assessing the in-use thermal performance of homes, and what these assessments suggest about the in-use performance of homes and retrofit. The pragmatic application of SMETER methods in this study also provides valuable insight into how well SMETERs can be applied to datasets that have not been collected under the typical conditions for SMETER assessments, which is important for understanding some potential applications, e.g. applying SMETER algorithms to historical or existing datasets.

Aims and research questions

The GHG-SMETER project addressed two core aims: to support the development of SMETER methods and provide insights into the challenges, limitations and opportunities for their application to assess both the thermal performance of properties and the impact of retrofit and to evaluate the change in heat transfer coefficient (HTC) upon retrofit of homes through the Green Homes Grant. The research questions are summarised below.

Table 1: Research questions for the GHG-SMETER project

Impact of the GHG scheme	
	What is the estimated improvement in thermal performance for different fabric retrofit measures in the GHG at a population level?
	At an individual property level what is the impact of different GHG measures on the HTC of homes as determined by SMETERs?
	Are there any statistical relationships between the improvement in HTC with key characteristics of the sample?
	How does the thermal performance of properties compare to a benchmark?
	What are the predicted changes to the energy consumption (standardised, assuming typical occupation and heating patterns) expected due to the change in measured thermal performance for the installation of different types of energy efficiency measures?
SMETER application and development	
	Can the SMETERs applied here be used to provide population-level and building-level insights into the efficacy of retrofit measures to reduce energy use? For what measures is this effective?

> What proportion of homes participating in the GHG may be adequately characterised by SMETERs? Do the evaluation of goodness of fit and other statistical measures suggest the HTC estimate for a home or group of homes is statistically robust? Are there identifiable characteristics of homes that lead to a poor statistical fit?
>
> For what proportion of the sample do SMETERs provide reasonable (within the expected bounds for property type) HTC estimates?
>
> How do HTC estimates from different SMETERs compare over a wide sample? Is it possible to determine the relative effectiveness of individual SMETER products against each other?
>
> How do different SMETERs perform in the estimation of the improvement in performance of different retrofit measures?

Only Type A SMETER methods were used to assess the difference upon retrofit in the GHG-V cohort. However, all SMETER methods in the project were applied to the GHG-LAD sample for comparison where possible.

In-use HTC and the 'true' HTC

The heat transfer coefficient, HTC, is a commonly used term in building physics to refer to the rate at which heat is lost or gained from a building per degree temperature difference between the inside and outside spaces. In the SMETER Innovation Competition project, a definition of HTC was adopted based on BS EN ISO 13789:2017 (British Standards Institute, 2017).

$$HTC = H_{tr} + H_{ve}$$

Where H_{tr} is the transmission heat transfer coefficient (usually measured in WK^{-1}) and H_{ve} is the ventilation heat transfer coefficient (WK^{-1}). This definition of HTC, whilst aligning to the ISO standard for the calculation of heat transfer coefficients (which refers to the modelled heat transfer coefficient rather than an empirically estimated HTC), differs from some definitions of the heat transfer coefficient, such as that used in IEA Annex 71 (Building energy performance assessment based on in-situ measurements), which excludes deliberate ventilation but includes infiltration (IEA EBC, 2021). The inclusion of ventilation is highly relevant to consideration of the energy performance of buildings experienced by occupants but creates an inconvenient challenge for characterising the thermal performance of a changing quantity (due to the variability of ventilation driven by occupant behaviour). Similarly to variability in other factors, and particularly those driven by occupants such as thermostat set point for heating, this can affect both the accuracy of results and their interpretation to provide insights for policy and practice for different purposes.

The transmission heat transfer coefficient is comprised of multiple components:

$$H_{tr} = H_d + H_g + H_U + H_a$$

Where H_d is the direct transmission heat transfer coefficient (through the external fabric of the property), H_g is the transmission heat transfer coefficient through the ground, H_U is the transmission heat transfer coefficient through unconditioned (unheated) spaces and H_a is the transfer heat transfer coefficient to adjacent buildings. All these heat loss paths are implicitly included in empirical testing of the real thermal performance of dwellings unless action is taken to exclude their effect. Similarly, the ventilation heat transfer coefficient can be represented:

$$H_{ve} = \rho_{air} \times c_p \times q_v$$

Where ρ_{air} is the density of air (kgm^{-3}), c_p is the heat capacity of air ($Jkg^{-1}K^{-1}$) and q_v is the air flow rate (m^3s^{-1}). This includes both intentional and unintended ventilation and variation in ventilation will be captured as part of empirical measures to estimate the HTC unless action is taken to remove it.

Whilst the definition of the heat transfer coefficient adopted in this and other work appears simple, it may present challenges for the determination of the HTC empirically and on the use of such values to inform policy and practice. The inclusion of components of heat loss that vary over time, most notably the heat transfer through the ground and ventilation, implies that the HTC is not a fixed value for any home and varies temporally according to the factors that drive those forms of heat loss (the appropriate temperature differences, mechanical ventilation and wind speed). This is addressed in the Standard Assessment Procedure model through assumptions of the value of such losses in different months (BRE, 2014) and since a significant purpose of the SMETER programme is to complement SAP, their inclusion in this work enables appropriate comparison. However, the time-varying nature of HTCs is a challenge to the determination of an in-use HTC, posing questions such as: how big is the seasonal variation in HTC? How big an effect can householders have on the HTC through factors such as ventilation choices? Should the in-use HTC represent the time for which it was estimated or should it be related to an average HTC over a defined period? How do these impact the inclusion of an empirical HTC in SAP-type calculations?

Thermal performance testing vs in-use HTC methods

The new BS EN 17887-2:2024 (*Thermal performance of buildings. In situ testing of completed buildings. Steady-state data analysis for aggregate heat loss test*) (British Standards Institute, 2024) builds on the previous leading test of steady state thermal performance of a building, co-heating (IEA EBC, 2021). In this method a property is vacated for two to four weeks, during which time it is heated to a constant temperature, using stand-alone electric heaters and air is circulated using fans. The thermal performance is derived from the energy use, the measured internal temperatures and the solar gains estimated from measurement of the incident solar flux and the expected solar aperture. Ventilation is typically closed and sealed during the aggregate heat loss/co-heating test, with an expected ventilation heat loss later added on the basis of simple calculation; however, co-heating may be undertaken with background ventilation active (Wingfield, Johnston, Miles-Shenton, & Bell, 2010).

Whilst the co-heating test is able to produce relatively repeatable HTC results (Jack, 2015), it is notable that aspects of the co-heating test method do not align to the estimation of a true in-use HTC as experienced by occupants. The co-heating test does not accurately represent ventilation, a significant component of heat loss. This is not only because ventilation openings are generally sealed during testing, but also due to the use of fans to homogenise temperatures throughout the property. In use interzonal air flows are an important component of the system of ventilation, driven by temperature differences and by the wind; the use of fans prevents such mechanisms of being accurately represented. Heat loss through party walls is challenging to represent. Ideally a process of “guarding” is employed, where the adjacent property (or properties) is also heated to the same constant temperature as the property under test. This ensures that no heat loss between properties is included in the estimate of the HTC, which is heat transfer to the outside. Guarding is practically challenging, requiring the operation of adjacent properties. An alternative to guarding is to measure heat flow into the party element using heat flux plates, take an average of these measurements and use this to represent total heat flow. Apart from the potential variability of heat flow over a party element, this method assumes that any party wall bypass (heat flow into the party wall, then to the outside) is zero, which could be a significant error in specific constructions (Lowe, Wingfield, Bell, & Bell, 2007).

Whilst other test methods to determine the HTC are currently in development or in the early stages of commercial use (e.g. QUBe, ISABELE and PSTAR), these still require a vacant property, their accuracy is yet to be proven over a large range of different building types, and they are subject to the same limitations as aggregate heat loss/co-heating tests, although the reduced duration does reduce the associated cost.

The issues with the co-heating tests discussed here also apply to the new aggregate heat loss test (British Standards Institute, 2024) and suggest that whilst this is a useful method that can provide insight into the thermal performance of the fabric of properties, it does not represent the true thermal performance of an occupied home. SMETERs can therefore be applied to address this issue, and whilst the less controlled indoor environment is expected to lead to larger uncertainty in results, and variation in operation changes the effective HTC (such as ventilation), they can complement aggregate heat loss/co-heating test results, aiming to provide insights into the in-use thermal performance for a large number of homes.

Report outline

The structure of the remainder of this report is given in Table 2, below.

Table 2 Outline of report structure and content.

Section 2	GHG Schemes, Recruitment and Datasets: The GHG schemes, the retrofits installed, recruitment of homes to participate in the study, characteristics of the participant sample and the complementary datasets collected and analysed in this work.

Section 3	SMETER methods: Details of SMETER methods (Type A and Type B), HTC calculation based on Energy Performance Certificate (EPC) input data, and the impacts of data selection on SMETER performance.
Section 4	Plausibility of SMETER results: The ability of the different SMETER methods to estimate a plausible HTC from the data collected in this study.
Section 5	Impact of GHG-retrofit on the in-use HTC: The SMETER-derived change in HTCs on GHG retrofit, and comparison to those predicted by simple a simple EPC-data model. Investigation into the impacts of different retrofit types and property characteristics.
Section 6	Discussion: Outcomes of the project and future work to support the implementation of SMETERs.

GHG Schemes, Recruitment and Datasets

In this section the characteristics of the households in the two GHG schemes are discussed, both those of the full datasets and our samples, with respect to characteristics of the population. It also summarises the recruitment of the samples used in the GHG-SMETER project, the quality assurance process followed for the data and discusses the quality of the datasets.

Statistical Disclosure Control

The Smart Energy Research Laboratory at UCL Energy Institute provided access to smart meter data via a controlled environment in a process compatible with the participant consent and ethics process at UCL. The Statistical Disclosure Control applied by SERL adopts the Office for National Statistics minimum disclosure sample of 10 records (homes) in almost all cases (with exceptions for some derived statistics) (Griffiths, et al., 2019) to ensure adequate anonymisation of records. This limit was required for the GHG-SMETER project and is applied in the reporting of results throughout this report.

Total GHG-V scheme installations

Figure 1 shows the number of installations of each type of fabric primary GHG measure that were successfully installed across the entire scheme. External wall insulation and loft insulation measures were installed most, followed by cavity wall and pitched roof insulation. Recruitment of participants, a sample from this population, is limited by the availability of homes with different measures installed and the proportion of those invited to participate who agree to do so.

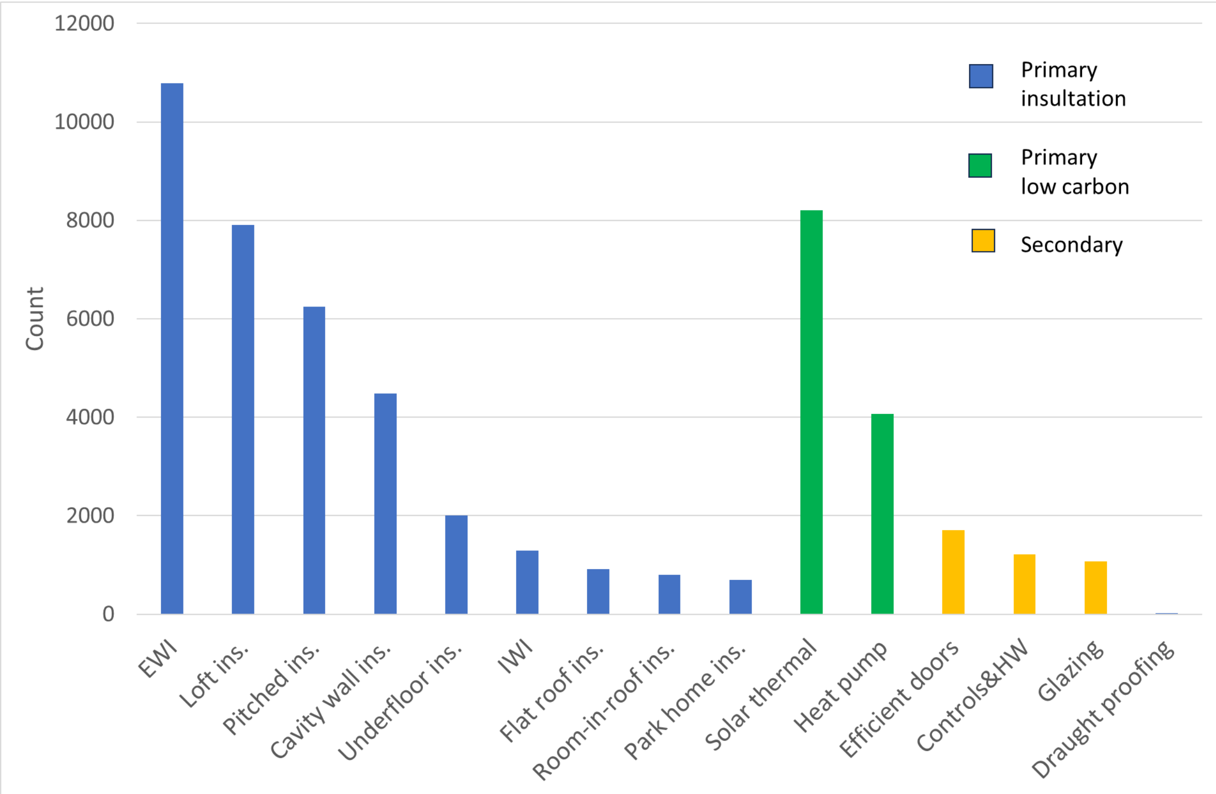


Figure 1 Measures for whole GHG Voucher scheme (45,491 homes with intervention date recorded).

The high number of pitched roof insulation measures is unexpected as this is not a commonly installed measure for a property without habitable loft space; it may suggest that installations have been recorded under this category for room in roof and other pitched roof constructions. However, with no alternative or supplementary data, for the purpose of this report the number of pitched roof insulation measures is assumed to refer to insulation of that construction, in a normal configuration.

Few secondary measures have been installed, as illustrated by Figure 1, leading to small sample sizes in the participants of this study with associated low statistical power; analysis of secondary measures is consequently not presented in this report.

Total GHG-LAD scheme installations

The measures installed through the GHG-LAD scheme are summarised in Figure 2, showing a different trend to that for the Vouchers. It is notable that with PV installations an allowable measure, these make up the largest proportion of installed measures across both phases. In cases when PV panels are present the energy use within the home is generally unknown in the absence of generation data (not provided through the DCC (Data Communications Company)); unmetered gains are most likely to cause underestimation of the HTC in analysis (more energy is going into the home to heat it than known and it is therefore warmer than without such gains). Homes with PV panels are analysed below (p53)0. The high number of external wall insulations and park home retrofits also contrast to the Voucher scheme. Homes were recruited through the LAD scheme for analysis in GHG-SMETER for in-home monitoring via temperature sensors in addition to smart meter data collected remotely, as outlined below.

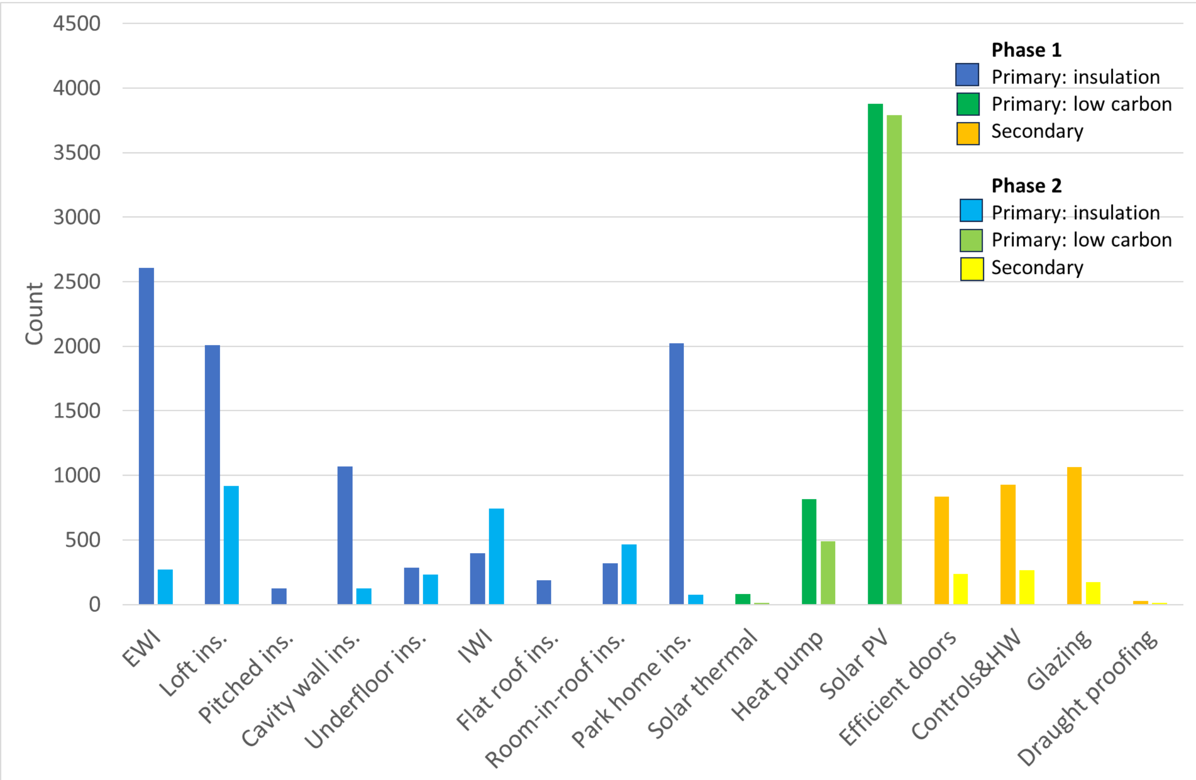


Figure 2 Summarises the number of households with different interventions with either a recorded start or end date of installation in GHG-LAD Phase 1 and Phase 2 as of April 2022 when recruitment for our project was completed.

Recruitment of participants into GHG-SMETER

Participants were recruited into the GHG-SMETER project for two HTC estimation approaches: Type A (remote only) and Type B (in-home) analysis. As discussed below (p25)0, the basic timeseries data for both methods is weather (external temperature, solar incidence) and smart meter data, these enable the energy inputs and the external factors driving heat flow to be included in calculations. Type B (in-home) methods add internal temperature measurement (at a minimum) to characterise the internal conditions and temperature difference driving heat flow. Both approaches may, as here, be supplemented by additional data such as EPC data and participant survey for certain implementations of SMETERs. Methods of SMETER analysis are discussed below (p24), alongside the length of datasets required by SMETER developers during the SMETER project.

Learnings from IEA EBC Annex 71, team expertise and the SMETER Innovation Competition were used to guide the data collection strategy:

- Type A methods generally require relatively long datasets and better results are often obtained with datasets of a year or more (Allinson, et al., 2022). HTC estimation is viable with datasets of three months in the heating season. Analysis of shorter datasets is possible, but increases uncertainty in HTC estimates.
- Type B methods typically require shorter datasets than Type A methods; 1 month is usually sufficient and some methods may provide results with only 2 weeks of data. However, it is desirable to apply both Type B and Type A methods to enable comparison of their performance, therefore the shortest dataset is ideally set by the requirements of the Type A group.

Recruitment for GHG-V

The GHG-V application database from late September 2021, provided by BEIS, was used to identify properties for recruitment. This included Unique Property Refence Numbers (UPRNs), consent for contact for further research and an evaluation ID which could be linked to a separate file recording details of the GHG application and the measures installed. Approximately 21,000 UPRNs were associated with households that consented to contact for further research and who had at least one measure completed. This list was cross-referenced against the Data Communications Company (DCC) register to determine which properties had the required DCC-enrolled smart meters to enable SMETER analysis. This is the same process used to recruit participants to SERL, has been approved by UCL ethics, and is compliant with GDPR and the Smart Energy Code (SERL, 2024).

UPRNs with an Electricity Smart Meter (ESME) that was commissioned at least 90 days before the date of the completed retrofit measure were identified and the addresses associated with these UPRNs were identified using AddressBase. There was no filtering based on the retrofit measure that the participants had installed. This resulted in invitations being sent to participate in the research to 6,862 households.

Recruitment and consent took the following steps:

- Initial contact via a letter posted to a recipient of a GHG measure at a meter's registered address. The first wave of recruitment letters were sent in late October 2021. Recipients were offered a £10 voucher if they agreed to take part in the study.
- Non-responsive recipients were sent a reminder letter 13 days after the first mailing.
- Recipients who were still non-responsive were sent a final reminder 13 days after the second mailing and this included paper/mailing options for sign up and survey completion.

- Participants were recruited by logging on to a web portal using a link and unique code provided in the letter.
- Consent was obtained at this stage through an Ethics Committee approved form.
- Online survey was completed at this point.
- There was a paper/mail option for those unable/unwilling to respond online.
- Once consent was obtained (and any authentication processes completed as required by the Smart Energy Code), collection of smart meter data from the DCC and linking to contextual data (e.g. the survey, EPC data etc.) could begin.

Participant recruitment closed on 13th December 2021, with 2,458 participants signing up to be part of this research, including about 300 participants who signed up by post. The response rate was 35.8%; this high participation rate is expected to result from the subset of the UK population who were motivated to seek a Green Homes Grant Voucher, potentially already being engaged with issues around home energy use, and that all those invited had already consented to be contacted for further research.

Recruitment for GHG-LAD

The GHG-LAD sample were targeted for Type B SMETER techniques requiring internal temperature data. Most Type B techniques ideally require a minimum of a month of temperature data recorded within the heating season, although less can be used if the data is of good quality. In addition to the criteria outlined above several additional filters were applied before homes were recruited from the GHG-LAD scheme. As well as a DCC-enrolled ESME, homes were required to:

- Have a DCC-enrolled Gas Proxy Function (GPF, essentially a gas smart meter). Approximately 75% of homes with an ESME also have a GPF.
- The data provided did not always provide sufficient information for all of the following filters to be applied, but where possible homes were excluded if they had:
    - Anything other than a gas main heating system (e.g. an oil boiler or electric heating)
    - Unsuitable building types such as flats (SMETER techniques work best for homes where the potential impact of heat transfer though party elements is minimised) and park homes (where Building Regulations do not apply and therefore the expected pre- and post- retrofit performance is unknown, and because of the very small number in the sample)

Note that when the information for these cases was not available homes were included to maximise participant numbers.

Recruitment of the GHG-LAD sample took place in two phases using the same steps as the GHG-V sample, in January 2022. At this time, the team received 5,968 records. Of these, the DCC inventory was queried for 1,102 with the rest of the homes unsuitable for a variety of reasons including lack of consent for recontact, missing key data, unwanted property types, no detail of install, etc. The results of the DCC inventory query found 391 households with smart meters who were invited to participate, of these 28 households signed up, with 22 having agreed to have temperature sensors. The DCC inventory was queried for a further 3,098 addresses (without the requirement of a confirmed installation), of which 891 were invited to join the study. Ultimately 173 households were recruited from GHG-LAD, of which 105 consented to internal temperature measurement, out of 1,282 invitations. This gave an overall response rate of 13.5%, of which 82.7% agreed to temperature measurement.

Figure 3 illustrates the retrofit installation status of the Type B sample of 105 homes, with only 44 receiving a GHG measure (some received multiple measures, hence the total of installs in Figure 3 is higher than this). Without timely access to accurate records of homes scheduled to receive a GHG measure installation recruitment resulted in only five of the recruited homes receiving a GHG retrofit during the period of study (the remaining measures were installed prior to temperature sensors), with only two of those receiving a fabric measure. A relatively large (>100 homes) investigation into the impact of the GHG programme on thermal performance wasn't possible with this data, nor into the ability of SMETERs to detect any such change. However, the data enables the investigation of the thermal performance of over 100 homes using both Type A and Type B SMETERs, using smart meter and internal temperature data (where necessary), in addition to contextual survey data and EPC input data from a sub-set of homes. This provides the opportunity to investigate the relative performance of different SMETERs and the performance compared to a white box model of thermal performance based on EPC input data.

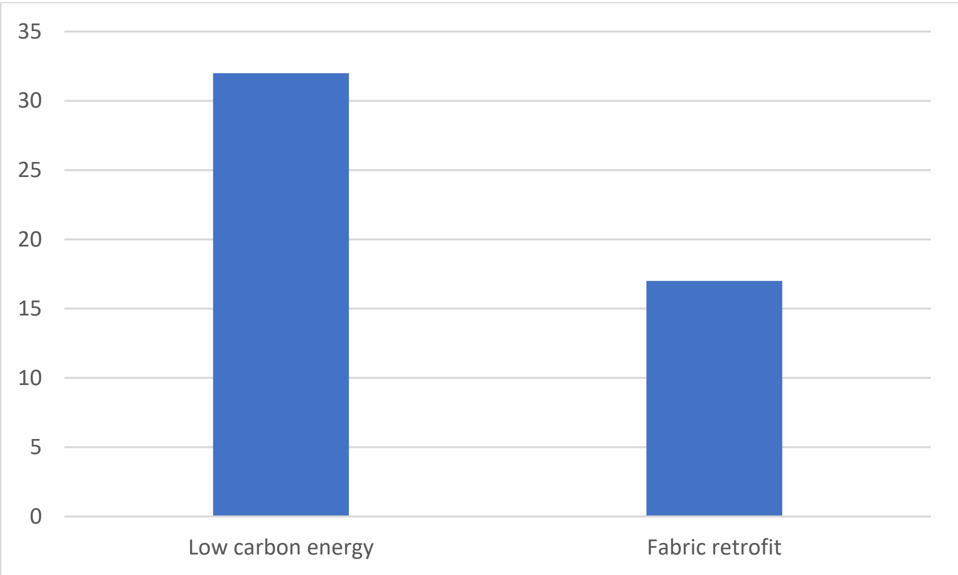


Figure 3 The split of the GHG measures installed in the recruited GHG-LAD sample of homes.

Temperature monitoring of the GHG-LAD households

The 105 households that consented to participate in internal temperature monitoring were posted four UX100-003 HOBO temperature sensors (stated accuracy 0.21°C), each labelled with either kitchen, lounge, hallway, or bedroom and given instructions on how to best place these sensors accordingly. The sensors were installed between January and March 2022, and were then requested to be returned to UCL that summer for the data to be downloaded. Four sensors were returned by almost all participating households, with one returning only three sensors and four households not returning a completed consent form allowing us to use the data.

Sample characteristics

The characteristics of the sample for both the GHG-V and GHG-LAD schemes are discussed in this section, compared to a representative sample, provided by the English Housing Survey (Department for Levelling Up, Housing & Communities, 2021). Since the GHG was a scheme designed to improve the energy performance of homes, it was expected that the sample would be biased towards lower thermal efficiency and EPC band. The following section reviews the characteristics of the GHG-V scheme, followed by those for GHG-LAD.

Characteristics of the GHG-V sample

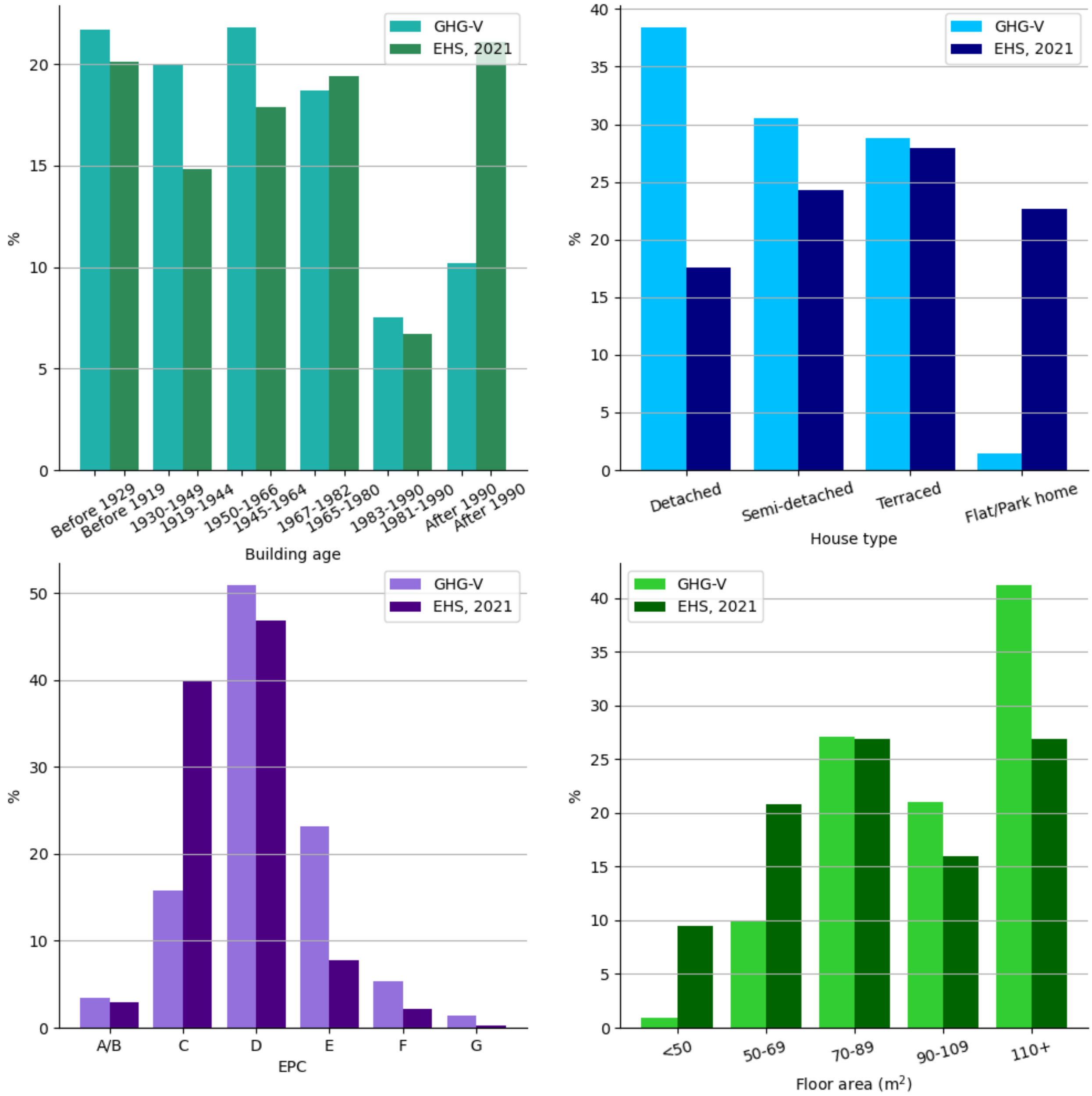


Figure 4 Bar charts of some key characteristics of the GHG-V sample of households. Note the different age bands used for building age in the two samples.

Key characteristics of the GHG-V sample are illustrated by Figure 4. The properties in the sample are skewed older than owner occupied homes in general, which is as expected that older homes are more suited to cost-effective retrofit. There is a high proportion of semi-detached and detached compared to the whole building stock; the impact of party walls on SMETER credibility rates is discussed presented below (p44)0. As expected in the analysis of a retrofit scheme, the sample is weighted towards lower EPC band homes as shown in Figure 4. Finally, the sample has more than 99% owner occupiers (as would be expected from nature of GHG programme). EHS 2021 shows 65.1% of homes owner occupied.

Of the 2455 homes recruited from the GHG-V scheme, 239 were mainly heated through unmetered fuels such as oil, LPG and biomass heating; these homes are not considered suitable for SMETER analysis because the primary heat input into the home is not available via smart meters. 294 homes exported electricity export for at least one half-hour period, most likely due to PV generation either installed as part of the GHG scheme, or already existing. The total energy used in such homes is unknown because smart meters do not record generation (only net import from and export to the grid) and therefore self-usage of generated electricity is unknown. The lack of data on self-usage of PV generation will impact SMETER estimation and this could vary according to the weather conditions experienced and the SMETER method (for example it may filter out high solar days); the impact on results is discussed below (p53).

Figure 5 shows the retrofits installed in the sample recruited through the GHG-V scheme for Type A SMETER analysis and may be compared to Figure 1 which illustrates all GHG-V installations at the point of recruitment. The recruitment of this sample has not led to a sample in which the interventions undertaken are representative of the whole GHG-V scheme, but this does not impact the objectives of the GHG-SMETER project. The sample includes sufficient fabric primary measures for analysis of the most common measures installed over the whole scheme: loft insulation, cavity wall insulation, pitched roof insulation (note comment above, p14,0 on the categorisation of pitched roof and room-in-roof insulation), external wall insulation and under-floor insulation. All of these measures are likely to cause a significant improvement of HTC, dependent on the pre-retrofit performance and area of the element addressed and are suitable for investigation using SMETERs. The data also enables investigation into the impact of solar thermal water heating, which should be minimal on HTC because it should not significantly increase heat gains to homes, and a measure that is expected to have a small impact on thermal performance of the whole house.

The relatively large sample of homes with heat pumps installed through GHG-V presents a challenge and opportunity to the implementation of SMETERs. Heat input from a domestic boiler may be readily assumed based on its manufacturer specification; however, the coefficient of performance of heat pumps has been found to be highly variable (Lowe, et al., 2017) and therefore heat input into homes may not be readily estimated from the electricity demand alone. SMETERs therefore cannot reliably estimate the HTC of homes with heat pumps without a heat meter, but may be used to assess the change in energy requirements of the home on retrofit (see p320).

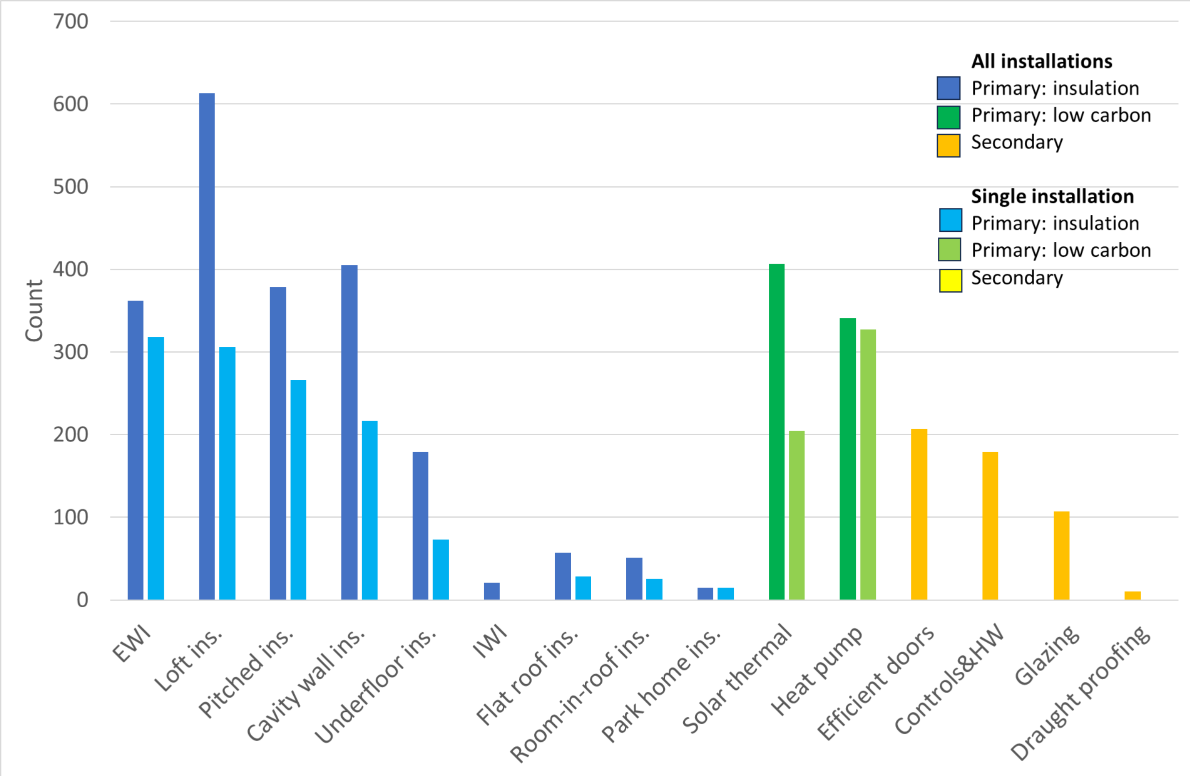


Figure 5 Retrofit measures installed in the properties recruited from the Type A GHG-V sample

Characteristics of the GHG-LAD sample

Of the 105 homes recruited into the GHG-LAD sample, none had reported unmetered main heating, with 10 having some form of unmetered secondary heating. As shown in Figure 2, many received PV as part of the Green Homes Grant and therefore this will affect the results of SMETER analyses of data post-installation (all Type B methods).

Figure 6 shows that the homes in the GHG-LAD sample are generally of similar age to those in GHG-V (skewed older than the wider stock), with a smaller proportion of detached homes. The sample consist of a higher proportion of semi-detached and terraced homes and a lower proportion of detached homes compared to the English housing stock (EHS 2021) and a much higher proportion of homes with EPC D or worse. The floor area of homes in the GHG-LAD sample is lower than those in the GHG-V sample, as expected as the “LAD scheme aims to raise the energy efficiency of low-income and low EPC rated homes including those living in the worst quality off-gas grid homes, delivering progress towards: reducing fuel poverty, the phasing out of high carbon fossil fuel heating and the UK's commitment to net zero by 2050” page 4, (Department for Business, Energy & Industrial Strategy, 2020).

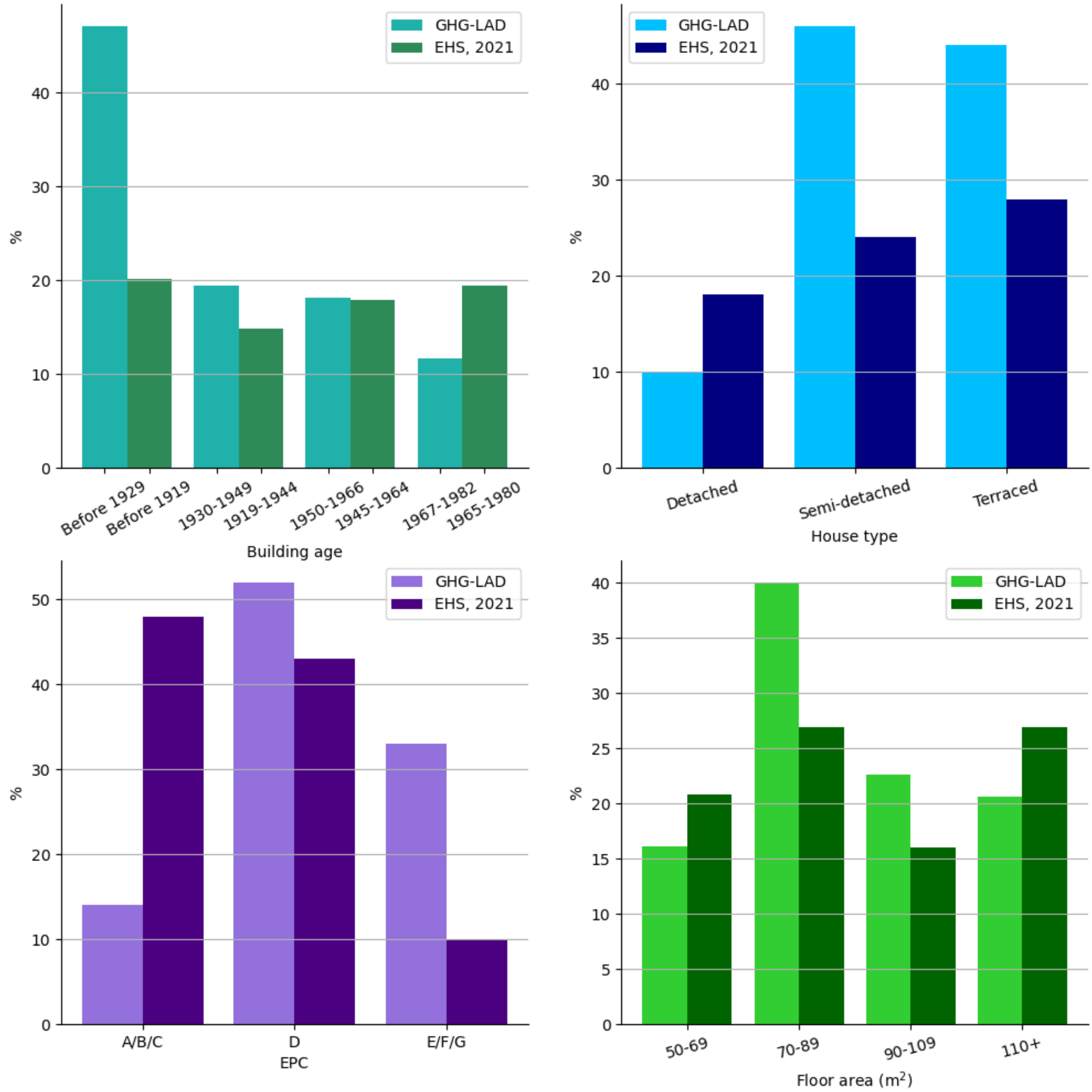


Figure 6 Bar charts of some key characteristics of the GHG-LAD homes in the sample. Note, as for the voucher plot, the different age bands in the data.

GHG-V and GHG-LAD retrofit data

The application for each GHG supported retrofit was required to detail the planned works, creating a database of planned GHG interventions with approval and delivery status to evidence the progress of the scheme. This data was provided by DESNZ to the GHG-SMETER project to enable the impact of the retrofit on HTC to be investigated, with the installation (or lodgement) date and the category of GHG measure recorded.

To enable some estimate of expected change in HTC, a simple model of thermal performance and expected change on retrofit was created based on EPC input data 0(p34). However, this

GHG intervention database did not include accurate details of retrofit in a consistent and documented format. For example, the area treated with insulation was not clearly recorded on the basis of measurements (the potential retrofitted area, amount of insulation purchased and unidentifiable numbers were variously reported), nor were pre-/post- installation U-values recorded. Modelling of the expected change in HTC on retrofit therefore made major assumptions to overcome these data deficiencies as discussed below (p3535). This data issue limits the comparisons that can be drawn between the expected and actual estimated change in HTC on GHG retrofit within this work.

EPC data

The EPC input data held by DLUHC was used to create estimated HTCs before and after an intervention based on SAP assumptions, with 1917 individual EPC records found for the 2568 homes across both GHG-V and GHG-LAD datasets. Where more than one EPC record was present for a household, the most recent was selected. There is no data directory for these EPC records and several formats and naming schemes meaning that not all parameters were able to be identified for all households; this resulted in 1814 EPCs being used to generate pre-intervention HTC values. To compare to the SMETER results per measure, only single fabric measures were modelled using this data and a lodgement date on the EPC was required, producing both a before and after-intervention HTC for 824 households.

Figure 7 shows the age of the most recent EPC record for homes recruited into the sample. As the age of the EPC increases, the likelihood of changes to the property that are not included in the EPC increases; it is likely that a significant number of the homes in the sample have received changes that would be noted on a new EPC. The ability of the EPC data to accurately represent the energy performance of homes in this sample is therefore unknown but is noted as a potential cause of significant uncertainty in modelled HTC. The use of this data to model HTC is discussed below (p34)0.

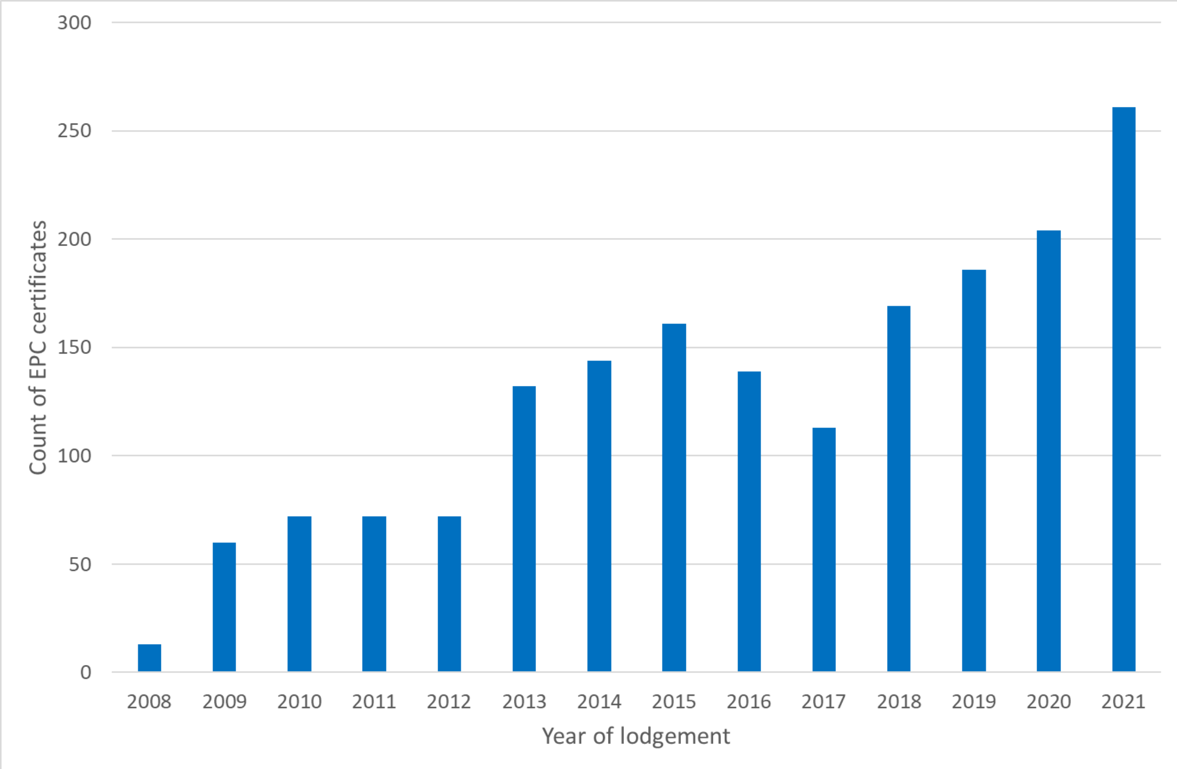


Figure 7 Date of EPC assessments for 1798 GHG-V homes for which Type A PTG HTC and an EPC data HTC were generated

SMETER methods

SMETERs are used to estimate the in-use thermal performance of homes and as such are distinct from methods that either model expected thermal performance or that estimate thermal performance in controlled experiments. The in-use thermal performance characterisation represents the property over the duration of data collection, it includes factors related to the state of conservation of the building, such as damaged or displaced insulation, and factors related to how the building is used by the occupants, such as door and window opening. The in-use thermal performance of a building therefore represents the experienced performance *over the period of study* rather than a theoretical or idealised case. The advantages and disadvantages of in-use thermal performance characterisation compared to modelling or to experimental testing are not the focus of this report. However, some pertinent features are discussed in the following sections.

SMETER methods are discussed in this section in addition to a brief discussion of the metrics used to form this characterisation. A simple building physics-based model of heat loss is also introduced, based on the EPC input data provided for this project. Finally, the intrinsic role of data selection within SMETER methods or application is introduced.

SMETER types and applications

The Department for Energy Security and Net Zero (DESNZ) has termed methods to characterise the thermal performance of homes using smart meter data to be Smart Meter Enabled Thermal Efficiency Ratings (SMETERs), this is typically described by an estimate of the in-use heat transfer coefficient (HTC). Similarly, DESNZ has proposed categories of SMETERs: Type A, requiring only smart meter and other remotely collected data and Type B, requiring additional in home measurements; and Type C, requiring further measurements (the latter are not applied in this project). This terminology is used here and the analysis methods that may be applied are discussed.

Type A methods are attractive due to their ease of application, with low cost, due to the lack of additional sensing within the home, and minimal disruption or engagement required from the occupants. However, without a measure of internal temperature it is challenging to characterise daily variations in energy use and internal conditions due to technology performance and occupant practices. Without other evidence it must generally be assumed that occupants do not adjust the set-point of their heating, radiator controls or programmer schedule, such that the demand for heat is driven by the external conditions alone, with static internal temperature requirements. Analysis of the internal temperatures of UK homes highlights that internal temperatures are often variable between days, in particular between week days and weekend days (Huebner, et al., 2013; Allinson, et al., 2022). Type A methods therefore generally require long datasets to provide sufficient data to average out the impact of some occupant practices. In the SMETER Innovation Competition Type A methods provided a lower accuracy HTC than those using internal temperatures (Allinson, et al., 2022).

Type B methods characterise the internal temperatures of homes, the accuracy of which is limited by the sensors used, and their ability to represent the average temperature of the whole home, dependent upon their location and number. This enables the influences of changes to heating set-point, duration and radiator controls to be incorporated into analysis. It also supports a more detailed characterisation of the impact of solar gains on internal temperatures, and therefore energy use. Type B methods may also account for the rebound effect, or comfort taking, on retrofit (Deurinck, Saelens, & Roels, 2012), whereby improvements in the thermal fabric result in higher internal temperatures in the heating season. Without internal temperature measurements, rebound can be challenging to characterise with Type A methods when the magnitude of rebound varies with external temperature.

The SMETER method employed depends on the constraints (cost, disruption, engagement) and accuracy of the required outputs (e.g. HTC or metered energy saving). Table 3 illustrates some of the potential applications of SMETERs. Insights from the GHG-SMETER project aim to help inform which applications may be addressed using SMETERs and explore the limitations of such methods. Occupant practices are particularly pertinent for the GHG-SMETER project, where changes to the space usage, or heating use, of individual rooms post-retrofit could lead to significant differences in the temperature distribution in homes and be reflected in characterisations of thermal performance derived from them.

Table 3 Applications of the characterisation of the thermal performance of a home.

Application	Comments
Property efficiency rating	Provide information and advice. Need to exclude impact of occupants.
Retrofit planning	Sizing of Heat Pumps Assessment of need and estimation of benefit of retrofit
New build performance	QA, feedback and learning, empower owners Performance gap versus design
Improvement due to retrofit	QA, feedback and learning Difference in HTC may be the key metric
Policy evaluation	National or local scale Assess the focus of future and efficacy of previous policy across building stock
Energy or cost savings	e.g. metered energy savings or pay-as-you-save. No determination of HTC is required, but HTCs can be used in model.

In this project, three UCL methods were applied to the households, along with two methods from participants in the SMETER Innovation Competition. As there are no internal temperature measurements for the GHG-V dataset it was only possible to apply Type A SMETER methods to these households: these were UCL's power temperature gradient (PTG) method and EDF's Deconstruct+. Type B methods applied to the GHG-LAD sample were UCL's MLR (multilinear regression) and Siviour methods used for QA in the SMETER Innovation Competition and Build Test Solutions' (BTS) SmartHTC method. These methods are predominantly physics based, with EDF's Deconstruct+ utilising machine learning.

The results of this work reflect the applied methods, which have been shown to perform well in the SMETER TEST project (Allinson, et al., 2022). They do not preclude the development of methods with improved performance in the future, such as an ability to return results with acceptable accuracy with smaller internal/external temperature differences.

Type A methods

In this section the SMETER methods applied in this project that do not require in-home sensors are outlined.

PTG

The Power Temperature Gradient (PTG) method is one of a family of methods utilising the *energy signature* of a building. The energy signature is a plot of the average (normally daily) power consumption against the external temperature, which can be modelled in a number of ways (Fels, 1986; Summerfield, Lowe, & Oreszczyn, 2010; Chambers & Oreszczyn, 2019). The PTG model takes the core assumption that the power use for heating varies linearly with external temperature until a set point, the balance temperature (T_b), above which no heating is required to achieve thermal comfort. The power use above this point is assumed constant, not dependent upon external temperature, representing a base power consumption when no space heating is used (P_b) (Fels, 1986), as shown in Figure 8.

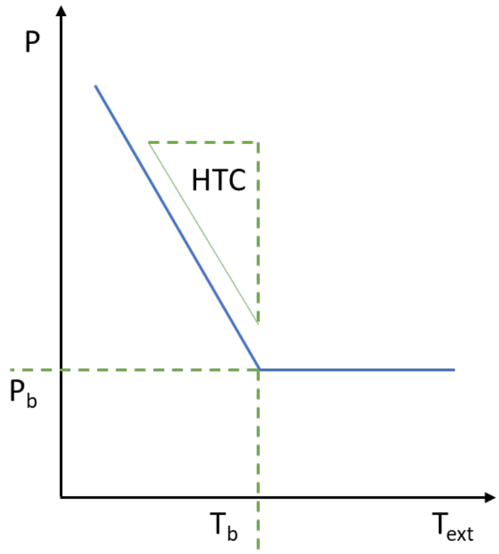


Figure 8 The PTG model, with power (W) on the y axis and external temperature (°C) on the x axis. The gradient of the sloped portion of the data is the HTC, and the flat section outside of the heating season indicates the baseline power P_b. The x-intercept of the changepoint between the two sections is the balance temperature T_b.

The PTG method relies on a range of important assumptions, most notably that free heat gains and occupant behaviour are constant over the averaging period chosen and that a single setpoint temperature is used year-round. As these assumptions are quite large, the application may result in a poor representation of the data by the applied model.

Due to the above limitations of the PTG method it was developed further for application in the GHG-SMETER project. To avoid the assumption of constant solar gains throughout the data analysed, a multilinear approach was adopted, with solar radiation on a third axis. This also enabled the effective aperture through which solar radiation enters a home to be estimated. In addition to this, several filters were applied to the data from each household prior to analysis to remove data that is either due to erroneous measurements or that may be real but does not meet the assumptions of the model, such as unheated days in the heating season. As occasional erroneously high readings are usual within smart meter data, a filter excludes days with power usage higher than a threshold amount, set based on the floor area of the property. A quick, initial analysis was then used to approximate PTG parameters and exclude datapoints too close to the baseline power below the balance temperature, or too far above P_b above T_b; these represent days when heating demand is unexpectedly low within the heating season (e.g. during a vacation) and energy demand is unexpectedly high outside the heating season (e.g. hosting an event with significant catering energy use); whilst such days likely represent real energy use, they are not compatible with the assumptions of the PTG model of a negative correlation between power usage and external temperature in the heating season and no correlation outside the heating season and are excluded from analysis. If a day was excluded through these filters, then the day after was also excluded to remove any residual effects, e.g. extra power usage from reheating the house after a day with no heating in winter. Across all pre- and post-retrofit analyses of the GHG-V households, these filters removed an average of 14.5% of datapoints, with standard deviation 18.6%.

The HTC is given in the PTG model as

$$P = \begin{cases} (T_{\mathrm{bal}} - T_{\mathrm{ext}})\mathrm{HTC} - \mathrm{gS} + P_{\mathrm{base}} & T_{\mathrm{ext}} < T_{\mathrm{bal}} \\ P_{\mathrm{base}} & T_{\mathrm{ext}} \geq T_{\mathrm{bal}} \end{cases}$$

The energy signature of a building is fitted to this equation using Bayesian statistical techniques, enabling both sections of the model to be simultaneously analysed. This provides greater confidence in the balance temperature estimate, and thus the HTC produced, than analysing them separately. A boiler efficiency of 85% (Orr & Summerfield, 2009) was assumed, to enable calculation of the HTC.

Error analysis for UCL-PTG

In line with the approach of the SMETER TEST project for the Innovation Competition, 95% uncertainty intervals for ΔHTC from the UCL-PTG method were calculated (Allinson, et al., 2022). The uncertainty on both the before and after in-use HTC, α_{HTC}, was calculated:

$$\alpha_{HTC} = \sqrt{\alpha_{var}^2 + \alpha_{stat}^2}$$

Where α_{var} is the variance component of the uncertainty, assumed to be 6% based on data from Phase 1 of the TEST project accounting for year-round variation in the HTC, and $\alpha_{stat}$is the statistical component of the uncertainty, calculated from the residuals of the model's fit to the data. For individual PTGs, systematic measurement error is also accounted for, but excluded from the error on the difference.

EDF

EDF applied their SMETER method, Deconstruct+, to the GHG-V sample of households to calculate the HTCs before and after intervention. This is a Type A method, using two months of smart meter, external temperature, and solar data from winter only. Days with high solar radiation are filtered out of analysis and EDF recommends filtering for homes with at least 45 days of data. If insufficient data is available from the current winter, more data is taken from the previous winter if available. More details of the analysis with this method can be found in the Appendix.

Type B methods

The addition of internal temperature measurement negates many of the assumptions necessary for the Type A methods, particularly those concerning the constancy of internal gains and setpoints.

MLR & Siviour

Siviour and Multiple Linear Regression (MLR) are, like PTG, regression methods based on rearranging the steady state heat balance equation and fitting the resulting equation to daily aggregated data. They both, however, require internal temperature as an input in addition to the information required by the PTG method. The two methods use very similar rearrangements of the equation and therefore give similar answers, although their different rearrangements of the heat balance equation lead to different treatment of errors Stamp, 2015).

The Siviour rearrangement of the heat balance equation is given below (adapted from Equation 2.7 in (Stamp, 2015)), where A_{sol} is the effective solar aperture and I_{sol} the solar radiation. Note that the HTC is no longer obtained from the slope of the plot but the intercept instead.

$$\frac{P_{\mathrm{heat}}}{T_{\mathrm{int}} - T_{\mathrm{ext}}} = -A_{\mathrm{sol}} \frac{I_{\mathrm{sol}}}{T_{\mathrm{int}} - T_{\mathrm{ext}}} + \mathrm{HTC}$$

The MLR equation is obtained by multiplying the Siviour equation by $T_{\mathrm{int}} - T_{\mathrm{ext}}$:

$$\mathrm{P}_{\mathrm{heat}} = -A_{\mathrm{sol}}\mathrm{I}_{\mathrm{sol}} + \mathrm{HTC}(T_{\mathrm{int}} - T_{\mathrm{ext}})$$

The MLR equation separates out the two effects of solar heat gain and heat loss through the building fabric by treating them as orthogonal variables. The HTC is now formulated as one of the regression coefficients, i.e. the partial derivative of power with respect to temperature. The same filter for erroneously high values in the smart meter in the PTG method was employed here, as this issue is prevalent in smart meter data. All boilers were again assumed to have a constant efficiency of 85% (Orr & Summerfield, 2009), and adjustments were made based on the number of occupants to account for metabolic gains and hot and cold water losses according to values from SAP (BRE, 2014). The errors in these methods are calculated in an identical manner to PTG (p28), incorporating statistical and measurement errors and an accounting of the natural variability of the HTC.

BTS

The BTS method of analysis is a Type B SMETER and as such was only applied to households from the GHG-LAD dataset with internal temperature measurements. The method applies a steady-state heat loss model and uses smart meter data (or meter readings at least 21 days apart), internal temperature measurements (at least 1 sensor, ideally more), external temperature and solar radiation data, floor area, and approximate location. Optional additional inputs are built form, party wall area, boiler efficiency, window details, and number of occupants. At least 75% of the days in the data period must have a daily average internal-external temperature difference of at least 7^{o}C.

SMETER accuracy, common assumptions and challenges

The ability of a SMETER to accurately characterise the thermal performance of a building depends on its ability to represent real energy demand. This depends on the model/data analysis and data selection that form the SMETER, plus the accuracy of measurements and the characteristics of the home in the state in which it is measured. SMETER accuracy therefore depends on both the activities of occupants (and how they may be incorporated or factored out of analysis) and the physical form of the home and the technologies within it.

SMETER technologies have been shown to produce HTC results in agreement with those from co-heating tests in the SMETER TEST project (Allinson, et al., 2022), with success rates between 70 and 97% across 8 different commercial SMETERs, including 5 SMETERs with over 90% or results found to be in agreement with the co-heating HTCs. These results are based on data from 30 homes of similar archetype in the north of England, with metadata including floorplans and professionally installed monitoring equipment (Allinson, et al., 2022). These SMETERs used varying types and numbers of measurements, with different data selection criteria in order to meet the assumptions of their particular models.

Behaviour/practice-related factors

In addition to the specific assumptions highlighted above, most SMETERs make significant assumptions about how homes are operated, or in order to interpret the results of repeated SMETER analysis on the same home it must be assumed that certain behavioural factors are constant. Whilst high-resolution (<30 minute resolution) sensing can capture most behavioural factors that affect the heat loss and gains of a home, the costs and intrusion to householders become an important consideration. Common un-measured parameters in a home that relate to occupant behaviour include:

- The flow temperature of the heating system, affecting its efficiency
- The open/closed state of internal doors, affecting ventilation rates and internal air flow
- Use of ventilation systems such as trickle vents, windows and extractor fans. Changes in temperature due to window opening events can be identified with some in-room temperature sensors, but measuring all ventilation is challenging
- In-room heating settings, such as thermostatic radiator valve (TRV) settings unless all spaces have internal temperature measurement
- Use of curtains and blinds and their impact on solar gains
- Temperature set point compared to the set point of any neighbouring properties sharing party walls and therefore forming potential heat loss or gain paths.

The impact of these factors is not addressed in this project but they remain an important consideration for the interpretation of results and the repeatability of measurements. They also illustrate why tests of the thermal performance of buildings and building physics models may not capture the lived experience of occupants.

Physical factors

Physical factors also affect SMETER accuracy; whilst a comprehensive discussion of the impact of these is outside the scope of this project, its relation to metered energy demand is discussed here. The ability of SMETERs to accurately characterise homes depends on their ability to account for the energy input into the homes which may be categorised as metered or unmetered. Metered energy use may be considered the smart meter data, unless secondary metering of other sources is present. Unmetered energy is all other flows, including solar gains, unmetered secondary or primary heating (such as a wood stove or oil boiler) and metabolic gains from the occupants.

The higher the metered energy relative to unmetered energy, the easier it is for a SMETER to characterise the thermal performance because fewer assumptions, less advanced models and less data filtering are required. Whilst differences exist between SMETERs, as the energy demand of a home reduces, the impact of errors that don't relate to total heating demand (e.g. data errors) increases. Similarly, as heating energy demand decreases due to better insulation, smaller size or reduced demand (e.g. underheating), the proportion of both unmetered gains and non-heating energy use increases compared to heating demand.

The weather experienced during a test period affects both the effective HTC during that time and the gains. Ventilation is weather-dependent, driven by wind speed, direction and internal-external temperature difference: it is not constant. Ventilation losses, as described above, are an inherent component of HTC and all tests exposed to the weather will therefore experience variable contributions (to a greater or lesser extent) due to changing ventilation rates. A SMETER applied over a limited time can therefore only represent the ventilation over that duration (although if well-characterised it may attempt to extrapolate over other conditions). If a SMETER aims to minimise the impact of ventilation, it similarly will only do so to the efficacy of the algorithms within the period of data collection. As losses through the building fabric decrease (higher efficiency fabric, smaller size), the relative importance of ventilation is expected to increase for in-use HTC calculation. For high-efficiency homes ventilation is expected to be a relatively more important component of heat loss than low-efficiency homes. Precipitation may have a minor effect on heat flow, which is not considered further here.

Solar gains are a major source of heat into many homes but are highly seasonally and weather dependent. As noted above, SMETERs account for solar gains in different ways, but their ability to do so is a key limitation to in-use HTC accuracy. Not only do solar gains change according to the intensity of incoming radiation (affected by cloud cover and season) and its angle of incidence on a home, but also to shading. This shading may be internal (previous section: blinds and curtains) or external such as from adjacent properties and trees; shading is therefore also seasonally dependent (Hollick, Gori, & Elwell, 2020). The ability of a SMETER to account for solar gains is a critical component of its performance and any error in doing so adds to the uncertainty of SMETER HTC estimates.

Metered energy use outside the heated envelope is another source of uncertainty in the heat input into a property. Homes with large energy uses outside the heated area are likely to be subjected to significant bias if this isn't separately metered. For example, homes with an electric car which is not metered on a separate circuit could cause significant overestimation of energy use within the home, as may heating or other energy uses in home offices or workshops outside the heated envelope of the property. It is important to identify such uses in SMETER analysis so that sub-metering may be undertaken, homes excluded from analysis or so that results on such homes are treated with caution. Expected increases in EV ownership as the car fleet decarbonises make this an issue of increasing importance (Department for Transport, 2021).

The SERL survey was completed by all participants in the study (p16) and asked participants to identify any significant energy use outside the heated envelope, such as to charge an EV. Homes with such energy uses are a small proportion of the sample and were excluded from the study, minimising the impact of metered energy use on estimated in-use HTC in this project.

The efficiency of heating plant and relationship between metered energy and space heating

The heat transfer coefficient is a measure of the heat loss per unit temperature of a property (British Standards Institute, 2017). Heat metering enables the heat output of heating plant to be measured (but may exclude some heating gains if it is sited within a property, such as case losses). However, heat metering is expensive and disruptive (Farmer, Johnston, & Miles-Shenton, 2006); the energy demand of a property measured by a smart meter or secondary meter is typically used for SMETER methods. However, this use of energy input rather than heat output necessitates the assumption of heating plant efficiency to estimate the HTC.

The true efficiency of a gas boiler is variable, depending on the specific model installed, its state of maintenance and its manner of operation (Orr & Summerfield, 2009; Hayton, 2009; Bennett, Elwell, & Oreszczyn, 2019). In particular, many boilers' efficiencies will vary with external temperature. Use of a standard, constant, boiler efficiency therefore represents a significant potential source of error for SMETERs applied without heat metering. Further, the highly variable performance of heat pumps due to both their temperature dependence and installation issues (Lowe, et al., 2017) cannot be accounted for without direct measurement of its performance; heat metering is typically required to estimate heat input into homes to estimate the HTC.

Domestic hot water

Energy use associated with the provision of domestic hot water (DHW) can be considerable and its importance increases as the heat loss from the fabric reduces (Department for Energy Security & Net Zero, 2024). As it is rare that DHW is sub-metered with a heat meter, the use of hot water can pose a significant uncertainty in SMETER analysis which introduces a systematic bias according to the assumptions implemented, whilst affecting the most efficient properties worst. The contribution to space heating of DHW may be considered to be 100% of all energy, a fixed proportion, a standard scaled factor or may be characterised outside the heating season and assumed to follow one of these relationships to space heating within the heating season. The challenge of estimating gains from DHW has been widely acknowledged (IEA EBC, 2021) and represents a future area of development for many SMETER methods, particularly with regards to the different heat delivery timescales of combi and system boilers.

An alternative to making assumptions about gains from DHW and estimating the efficiency of heating plant, or heat metering, is the use of an alternative measure that provides an estimate of the dependence of the energy use of a home on external temperature, including all system components such as heating plant. One such metric is the heat power loss coefficient (HPLC) (Chambers & Oreszczyn, 2019), using energy input data (gas and electricity); the HTC represents the HPLC divided by system efficiency. Whilst HPLC does not directly communicate the thermal efficiency of the property, it better characterises the variation of whole home energy use, which may be relevant to occupants and relates better to emissions targets than the HTC.

The factors affecting SMETER performance and applications

The factors discussed above that contribute to SMETER performance are important not only in the application and development of SMETERs to characterise the thermal properties of homes, but also relate to their specific application. For example, when used as here to investigate the impact of the GHG measures on the HTC of homes, factors such as a change in ventilation provision and use, a change in space use (increased or decreased heated area) and internal door use will affect the ability of SMETERs to characterise the observed change in HTC accurately. SMETERs used to evaluate the change in HTC on retrofit are therefore exposed to the most demanding application: the requirement to accurately determine a change in HTC for a home in different conditions, potentially heated and used in different ways before and after retrofit, and potentially subjected to consequential improvements that are not documented in the GHG records.

Factors included in the applied SMETERs

The Type A SMETERs applied in this analysis cannot directly account for any factors requiring additional internal measurements, such as changes to the heated area, set-point or timings, boiler flow temperature, ventilation, or other comfort taking post-retrofit. The UCL-PTG method inherently accounts for changes to the set-point temperature of the heating (internal temperature) through its formulation around a balance temperature . However, it does not account for the remaining issues highlighted above.. Details of the EDF-Deconstruct+ method are confidential and it is not known if it attempts to account for any such issues.

The Type B UCL-Siviour and UCL-MLR methods utilise temperature measurements in four rooms in each home and therefore account for changes to internal temperatures, although it is assumed that these accurately describe the whole space. No account is taken of ventilation, flow temperature, use of blinds and curtains, internal doors or use of energy outside the heated envelope. BTS used the same in-home measured data as UCL methods in this study; the details of their algorithm are not known.

For all of the methods, it was beyond the scope of this study to understand the extent to which factors besides the retrofit measure are contributing to the change in HTC However, the advantage of utilising these in-use methods to assess retrofit is that the real energy demand reduction is measured, taking into account occupant behaviour regarding the retrofit as well as the actual changes to the building fabric or services.

Modelled estimate of HTC based on EPC input data

A building physics model may be constructed to estimate expected HTC based on the physical form and construction details of each home. This approach is the basis of energy and cost calculations in Energy Performance Certificates, which do not directly communicate HTC, and represents the incumbent method of estimating a property's thermal performance.

As part of EPC an assessor undertakes an energy survey of each home to determine its dimensions, construction details, heating, cooling and ventilation systems. This EPC input data was made available to this project, as described above (p22), for all homes within the sample having an EPC assessment. The data was then used to develop a simple model to estimate the thermal performance of the home on the basis of elemental U-values. The model developed for this project was simplified compared to a SAP assessment to utilise U-values, areas, and an annually averaged version of the SAP ventilation heat loss calculation.

The U-values obtained for the elements in the model are those assumed in the most recent version of Reduced Data SAP (RdSAP) version 9.94, specified in Appendix S of SAP 2012 (BRE, 2014). These U-values are largely based on property age, and as such may not be accurate, particularly if buildings have been upgraded since being built. Moreover, standard assumptions regarding the U-values of existing elements may not be accurate, for example, the U-value associated with solid walls was revised down from $2.1Wm^{-2}K^{-1}$ to $1.7Wm^{-2}K^{-1}$ in an update to RdSAP in 2017 following updated empirical evidence Li, et al., 2014).

Recent evidence has identified a systematic difference between the energy use calculated by the EPC model and metered energy use (Few, et al., 2023). This research found that the modelled and metered primary energy use intensity (PEUI) was not statistically significantly different for homes in bands A and B, but for all other bands the modelled PEUI was significantly larger (8% higher for band C and increasing to 48% higher for bands F&G) than metered energy use; the exact causes of the discrepancy remain unclear. However, there are various plausible reasons for the discrepancy such as home upgrades between EPC generation and energy use measurement, limitations in the underlying thermal model, and limitations of the RdSAP model.

As a result of the research findings outlined here, it is prudent to interpret the outputs of the EPC model with a degree of caution if intending to compare these to measured energy use.

For the HTC model based on the EPC input data, the EPC lodgement date was checked against the measure install to ensure it was from before the GHG intervention (fewer than 10 of the EPCs from were from after the install date, and so were excluded from analysis following statistical disclosure control guidance, p13).

Calculating the expected impact of retrofit on HTC

The impact of fabric retrofit on the HTC of properties with EPCs was also calculated using this model, whilst non-fabric measures were not modelled since they do not affect the HTC. In this model it was assumed that all of the specified element in the main part of the building was retrofitted (e.g. the whole loft of the main part of a house, excluding any extensions). If the dwelling had an extension, it was further assumed that this would only be retrofitted if it was in the same age band as the main home, or one age band newer. The retrofitted U-values used were those in Table 4.3 of Part L of the Building Regulations (HM Government, 2021), detailing the minimum U-value of an existing element that should be achieved on retrofit.

The assumption of retrofit details, area and thermal properties is a potentially large source of uncertainty in calculating the expected impact of retrofit on the HTC of homes in this study. As it is assumed in this model that the whole element receives retrofit, rather than only parts of it (e.g. all external walls rather than just two of three), this assumption may lead to a systematic over-prediction of energy savings associated with GHG retrofit. The calculated change in HTC on retrofit can only be considered an approximate guide and not of sufficient robustness to enable any performance gap between expected and actual energy use to be investigated. In addition, as this method is based entirely on the insulation levels of the building fabric, any effects of occupant behavioural changes post-retrofit will not be accounted for in the results.

Data selection and SMETER results

Data selection is an integral part of any SMETER method and can have significant impact on the results obtained. Consequently, although all of the SMETER methods included in this project were provided the same datasets for analysis, each method applies its own data filtering according to the requirements of that method; HTCs were produced on the basis of the analysis of the sub-set of data filtered according to each method.

The PTG method employed by UCL excludes cold days with low power usage since either internal temperatures will be lower than generally experienced in the home, or unmetered heat gains are used to maintain thermal conditions. For example, the use of set-back temperatures during vacations should be identified using this method and excluded from analysis; this is not exclusion of erroneous data, rather, is the exclusion of data from analysis that does not meet the assumptions of the model. Similarly, days with unusually high energy use compared to that expected given the external temperature were excluded from this method (such days may indicate high energy use outside the heated envelope, or the increase of thermal set-point to accommodate different needs for a period of time).

The EDF-Deconstruct+ Type A method is proprietary and as such it is not possible or appropriate to discuss all data filters employed to identify appropriate analysis periods. However, we highlight two notable differences from PTG. Firstly, only data from the winter months, ideally December and January, is used to try and ensure a large internal-external temperature difference. This is expected to reduce the impact of unmetered gains (such as metabolic gains); it increases the expected heating use and therefore reduces the impact of effects that do not scale with external temperature. Reducing the data collection period may also increase the likelihood that occupants operate their homes consistently and maintain a constant thermostat set-point. Secondly, days with solar gain, an unmetered gain, above a threshold amount are omitted as contrasted to the PTG method which directly incorporates solar gains into the model.

In home (Type B) methods utilise internal temperature measurement, at a minimum, in addition to the data available to Type A methods. This enables the analysis of data with different heating system operation (such as thermostat set-point and heated period), although some methods may still reject such periods to increase their accuracy. Unmetered gains remain an issue and algorithms may be developed to either exclude days with specific characteristics from analysis, or to account for them within the HTC analysis model, as outlined above (p24). The data filtering of BTS SmartHTC is an integral part of the method and is proprietary and thus not discussed in this report, however requires a certain number of days with valid data and an internal-external temperature difference greater than 7^{o}C.

The implications for SMETERs from internal temperature data

The internal temperature data from the 105 homes included in the GHG-LAD sample were studied to provide insights into the operation of heating and likely impact on the application of SMETERs. As noted above, the temperature sensors were installed between January and March 2022, and returned that summer, with the consequence that less heating season data was available than would be ideal. The 105 homes monitored were drawn from the GHG-LAD scheme, targeting households more likely to be in fuel poverty with criteria set by the local authorities such as a combined gross household income of no more than £30,000 (Department for Business, Energy & Industrial Strategy, 2020). This is not a representative sample of UK homes, most notably being biased towards lower household income and lower thermal performance. However, insights from this sample provide valuable insights into the potential variation in performance of SMETERs in these types of homes; this sample is directly relevant to the evaluation of the thermal performance of homes retrofitted under similar schemes, such as ECO and therefore the evaluation of the policy impact of thermal efficiency measures.

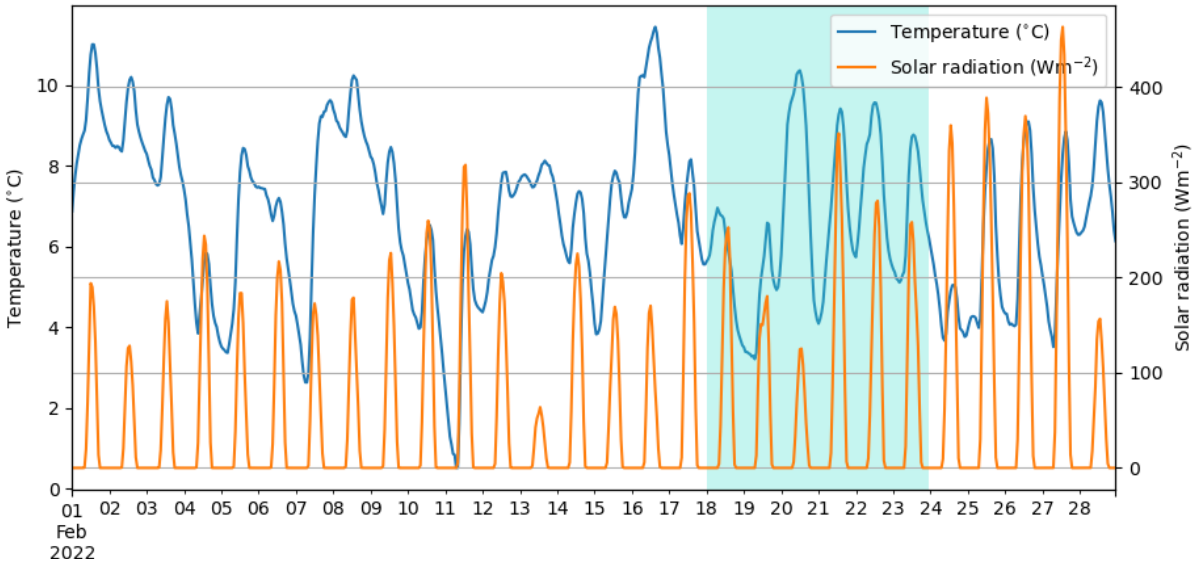


Figure 9 The average weather conditions in Great Britain during February 2022. The period highlighted in light blue is when the internal temperatures displayed in this section were taken; the weather in these 6 days appears typical of the rest of the month, and sufficiently cold for space heating to be required.

The living room temperature from 9 GHG-LAD dwellings from 18th-23rd February is shown in Figure 10, with the corresponding weather data from this period in Figure 9. This time period was selected as it is prior to the invasion of Ukraine and any associated media coverage of effects on energy costs, and also highly likely to be within the heating season of dwellings. There is a very wide range of temperatures and profiles, with considerable deviation from those assumed in SAP (BRE, 2014). Dwelling 1015 has a strikingly low temperature suggesting a lack of space heating input resulting in a free-floating temperature responding to external conditions and solar gains; other data for this dwelling indicate that the dwelling was occupied during this time. Dwelling 1011 has very little variation in temperature, perhaps indicating a constant setpoint and good levels of insulation, and dwelling 1020 appears to have a single daily heating period with a setpoint of around 20^{o}C. The other dwellings plotted in general have no clear repeating heating pattern or heating setpoint. EFUS 2021 (Department for Business, Energy & Industrial Strategy, 2021) found that 28% of households had no regular heating pattern, which was more common amongst non-centrally heated homes. EFUS reports the most common heating pattern to be a short period in the morning and a longer one in the evening, similar to that assumed for weekdays in SAP, however with little difference between weekday and weekend patterns (Department for Business, Energy & Industrial Strategy, 2021).

Internal temperature profiles can be affected by a number of factors including the impact of internal and solar gains, changes in occupancy, use of TRVs/smart thermostats, occupant ventilation practices and any operational limitations of the heating system, heat emitters or heating controls. How well the measured profiles represent the dwelling will also depend on the placement of the temperature sensors; participants were issued instructions on the placement however whether these were followed is impossible to verify. Despite the large variation shown in Figure 10 the average internal temperature across all 9 dwellings during this time is 18.9°C, which is close to what would be expected from a SAP calculation for the daily mean temperature of the living area of a typical UK dwelling in the heating season, assuming a space heating temperature set point of 21°C and typical parameters of the housing stock for insulation and thermal mass. It is slightly lower than the median setpoint found in EFUS 2021 (Department for Business, Energy & Industrial Strategy, 2021) of 20°C.

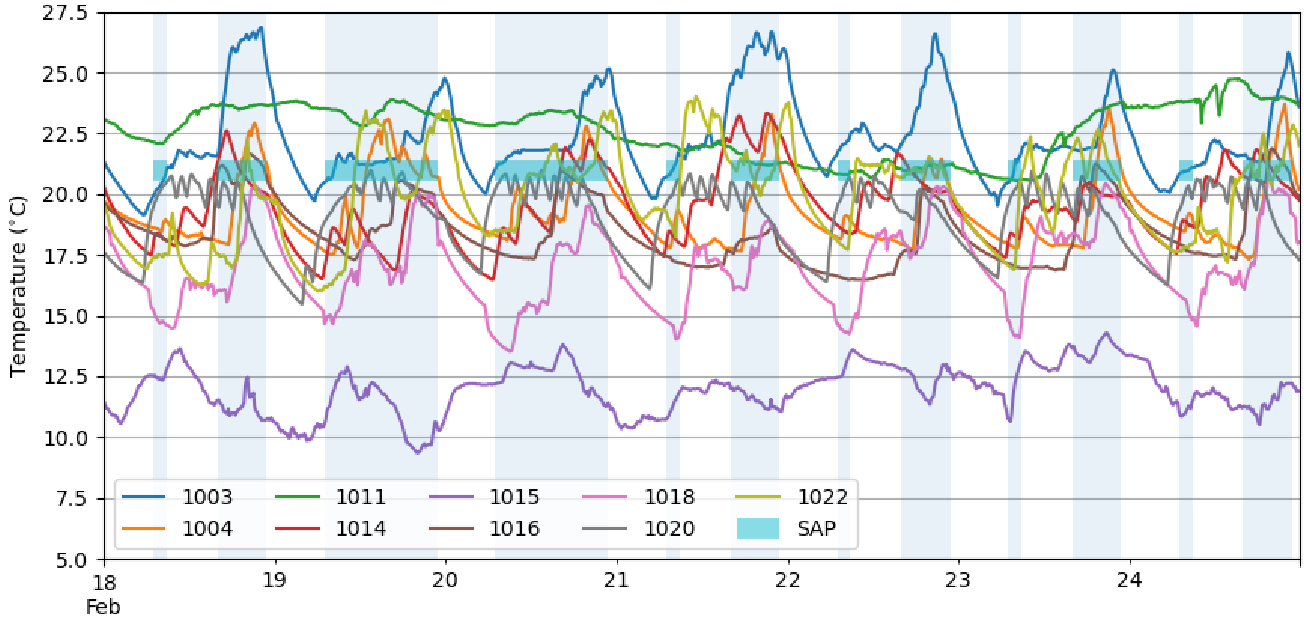


Figure 10 The living room temperature in 9 households in the third week of February, with the setpoint and heating profile assumed in SAP also shown.

The majority of homes in the GHG-LAD sample (~60%) do not have a regular or consistent heating pattern (the same timings and temperature setpoints each day); such lack of regular heating pattern is particularly challenging to Type A methods that assume consistent heating. Approximately 15% are underheated and/or have consistently low temperatures. Such homes are challenging to analyse due to the low internal-external temperature gradient; such homes also generally have low heating energy use. In both cases this increases the expected uncertainty in HTC estimation since the error in temperature and energy measurement is larger relative to the absolute difference in temperatures measured.

Potentially more importantly, unmetered gains are a larger proportion of total energy input for underheated properties than well-heated properties. Unmetered gains result in a lower HTC estimation than the true HTC because the home is warmer than that which would be measured on the basis of metered energy alone. 15% of the homes are underheated, with significantly lower temperatures than experienced in most homes, so the likelihood of significant heat flow through party elements is relatively high. An underheated property with party walls is more likely to be cooler than its neighbour(s), and experience potentially significant unmetered gains through party elements, depending on the construction details. Similarly, any home may be next to an underheated home that does not form part of this sample and may therefore experience potentially significant party wall losses. The impact of these issues is to respectively bias in-use HTC estimates low or high respectively. More extensive monitoring including neighbouring properties is required to understand the potential impact of this issue, which will be related to the thermal properties of the party wall.

Finally, the vast majority of occupants turned off their heating for the last time in April or May, whilst around 10% turned off their heating earlier in the year; SAP assumes a heating season to the end of May, whilst EFUS (Department for Business, Energy & Industrial Strategy, 2021) found the most common to be until the end of April.

The measured temperature data of a home is critical to Type B SMETER accuracy as the HTC is the rate of heat loss per degree temperature difference between the interior and exterior. In analysing internal temperature data it must be assumed that this represents the whole home or can be scaled appropriately to do so, which may be achieved in different ways. For example, a volume-weighted average may be taken as the average temperature of the home, alternatively rooms of different types may be assumed to have the same temperature, or set differences in temperature between rooms may be assumed.

Ideally, the temperature in every room of a home would be measured to minimise error in the calculated representative internal temperature of the home for SMETER analysis. However, the cost and disruption to householders of a SMETER increase with the number of rooms monitored due to the cost of equipment, technician or occupant time. It is therefore desirable to reduce the number of sensors required to provide an adequate representative measure of internal temperature in each home; in this study sensors were in the living room, kitchen, a bedroom, and hallway of each home. These locations were chosen to best represent the main living space (living room), temperature rises due to potentially significant gains from cooking in the kitchen, the transitional space, hallway, which may be heated differently from occupied spaces and a bedroom to represent sleeping spaces (and potentially the first floor of the home). Whilst more sensors would be desirable, these locations were selected to balance cost and disruption against accuracy.

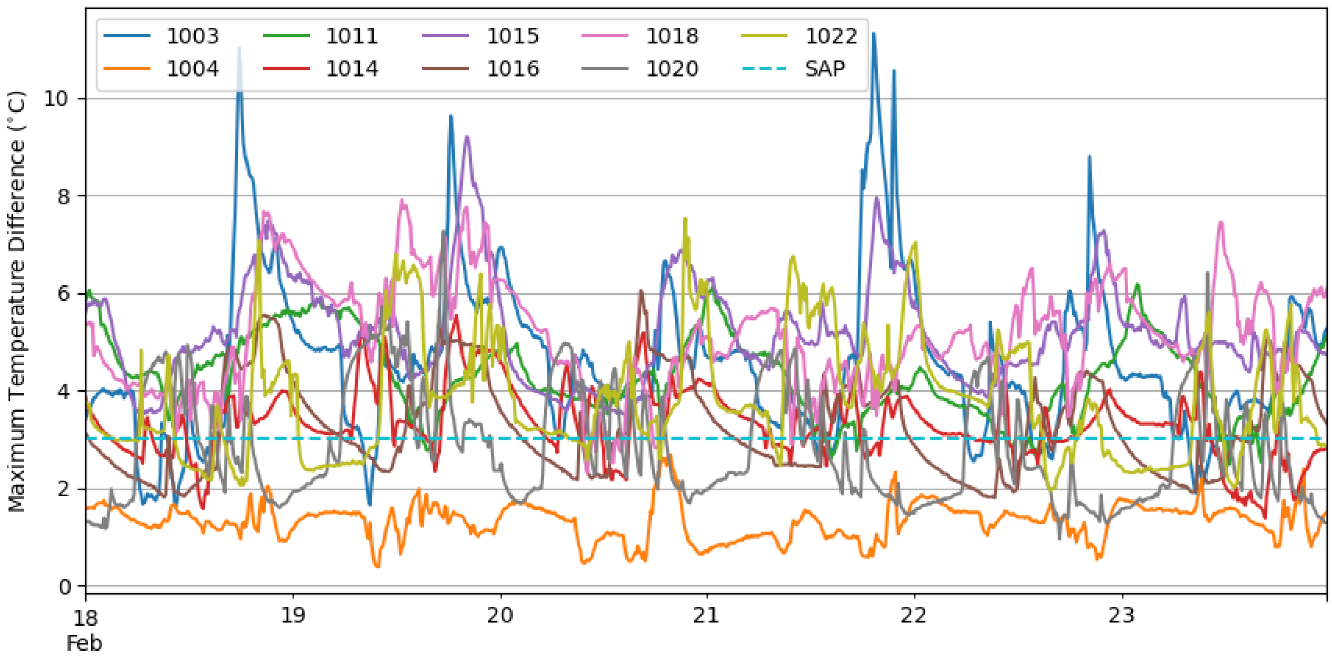


Figure 11 The maximum temperature difference between the 4 sensors in each home during the same week in February, with the dotted light blue indicating the maximum assumed in SAP.

In Figure 11 the maximum temperature difference measured between rooms at each timestep in the same 9 homes for the same 6 days is plotted, with a dashed line indicating the maximum as assumed in SAP (BRE, 2014). EFUS (Department for Business, Energy & Industrial Strategy, 2021) found that bedrooms were around 1°C cooler than living rooms on average. The majority of the homes in Figure 11 exceed these temperature differences, with an average of 3.7°C. There are some particularly sharp peaks in the difference for 1003, which appear likely to be due to heat gains from cooking as they occur in the evening between the kitchen and other spaces. The observed high variability of internal temperature in the sample homes may be attributed to a range of causes, such as the use of TRVs, manual control of the heating, the presence of unheated rooms, incorrect sensor placement and high gains in certain spaces. The impact of these issues on the HTC derived by a SMETER is case and method specific but would include increased uncertainty and increased likelihood of inaccurate estimates as key assumptions are broken. Similarly, differences in space use and temperatures achieved are critical for the comparison of expected energy use from models such as SAP, subject to assumed temperature differences, compared to that measured in real homes, where it depends on the actual conditions achieved.

Summary of SMETER methods

This section has summarised the methods used by SMETERs to estimate the thermal performance of properties. It started with a discussion of the most appropriate metric to measure the thermal performance, highlighting that the heat transfer coefficient is a property of the fabric and therefore an attractive measure to use, but that this can be hard to identify with sufficient accuracy without extensive heat metering to identify the performance of boilers, heat pumps and the use of domestic hot water, which must otherwise be assumed in SMETER models. A characterisation of the energy signature of the property, describing how energy use changes with external temperature, is an alternative to the HTC and can provide useful insight to householders. Further, energy uses outside the heated envelope can cause significant error, such as heating a workshop, office or charging an EV and sub-metering of such energy use is required to accommodate the analysis of such homes with SMETERs.

The methods used for Type A and Type B SMETERs were then reviewed. This project uses the Type A methods of UCL-PTG and EDF-Deconstruct+ in addition to the Type B methods of UCL-MLR, UCL-Siviour and BTS SmartHTC. The core assumptions and data requirements of the UCL methods were reviewed.

A simple building physics based model of thermal performance was presented, based on EPC input data, to represent an expected HTC according to SAP-type methods. The modelled change in HTC on retrofit was noted to be uncertain due to the poor data quality of the GHG-V retrofit database, requiring the treated areas and products to be assumed in modelling. The change in HTC on retrofit predicted from the EPC-data Model should therefore be considered only a guide and not suitable to identify any potential performance gap.

Finally, the impact of data selection on in-use HTC estimation was explored and the temperature profiles of homes in the GHG-LAD sample discussed. The results highlight that data selection is an important component of SMETERs and automated HTC analysis must be developed to identify suitable data to represent a repeatable and reliable HTC. Internal temperatures in the LAD homes were found to deviate significantly from assumptions in SAP, and often low energy use and temperatures were noted. These issues also affect SMETER reliability and uncertainty and the high prevalence of underheating in the LAD homes likely reduces the accuracy of SMETERs for these properties and may increase the number of homes for which SMETERs are unable to produce a reliable result.

The following section presents an investigation into the reliability of SMETERs, applied to the data presented in *GHG Schemes, Recruitment and Datasets* (p13) and using the methods presented in *SMETER methods* (p24).

Reliability of SMETERs

This section presents results of this study into the reliability of SMETER methods in characterising the thermal performance of the GHG data, the learnings of which are then carried forward to the next section which investigates the impact of retrofit on the in-use HTC of homes retrofitted through the GHG.

The reliability of SMETERs may be conceptually divided into separate components: their fit to the data; their self-consistency or repeatability; and how they compare to each other and some ground truth. Discussion of the in-use compared to a tested HTC (p10) and factors affecting SMETER accuracy (p24), many of which also apply to testing methods, highlight the challenge of establishing an appropriate “ground truth” to compare SMETER in-use HTC results. Additional testing was not possible in this project, such as the new aggregate heat loss test (British Standards Institute, 2024), due to the high cost and disruption since it requires vacant properties. This section explores these different components of SMETER reliability as it may be investigated within the data availability of the GHG-SMETER project, without the provision of a comparator tested under controlled conditions.

SMETERs cannot always produce in-use HTC estimates that meet basic internal robustness and consistency checks. This reliability is explored through an example using UCL SMETER methods, applying internal tests that can be applied to determine if a SMETER result should be considered sufficiently internally robust to be reported, or should be considered a null result. The “plausibility rates” of SMETERs with regards to the GHG datasets are then explored, where success is considered to be results passing the SMETER internal validity process, and how this relates to characteristics of homes and heating systems. The use of plausibility rates enables investigation into an important component of ways it is possible to characterise the reliability of a SMETER; it does not indicate any true accuracy compared to a “ground truth”, known, HTC. SMETER estimates of in-use HTC are then compared to HTC estimated using the EPC-data Model.

SMETER result self-robustness

Any quantity that is obtained indirectly through the application of a measurement model, which derives this quantity from the measured quantities, is subject to uncertainty associated with how accurately that model accurately represents the system (Joint Committee for Guides in Metrology, 2012). These uncertainties rely on both the abstraction of reality into a model representation and assumptions the modeller makes during the modelling process (such as the discretisation of continuous data (Gori & Elwell, 2018). In the worst case, such uncertainties mean that the model does not represent the system appropriately to estimate the measurand. Such uncertainty is particularly significant for SMETERs, where in addition to the random and systematic errors that are associated with measurement, models of homes often include implicit assumptions about the operation of that building by occupants. Various tests can be applied to SMETER analysis and outputs to investigate whether the SMETER provides an acceptable fit to the building performance, which inherently also address the random errors of the analysis. The SMETER methods utilised in this project mostly use physics-based models, however there are SMETERs, including EDF's Deconstruct+, that make use of machine learning techniques. For these SMETERs, although no physical relationships are assumed between the variables, different assumptions, dependent on the algorithms used, will still apply.

The models within SMETER methods are subject to varying assumptions, particularly Type A methods where the lack of internal temperatures increases assumptions about home operation, and as such cannot accurately characterise all homes. Not all homes are operated according to the inherent assumptions of each SMETER method, and there can be erroneous and missing points in smart meter data, particularly gas (Webborn, et al., 2021). Without a ground truth value for the HTC for each home, such as a aggregate heat loss test result (although we note the expected difference between in-use HTC and aggregate heat loss HTC as discussed above, p100), whether a SMETER result should be considered sufficiently robust to be a reasonably accurate characterisation of the performance of the home needs to be determined through other methods. In this section the application of various tests for the three UCL methods (PTG, MLR, and Siviour) are presented with the difference they make to the results. The final methods used for determining the "plausible" results of these methods are discussed.

A first useful cross-check for a SMETER result is physical plausibility. The true HTC distribution of HTCs in UK homes is unknown. However, HTCs below 50WK^{-1} are very rare and are associated with the highest performance standards such as Passivhaus (Johnston, Sidall, Ottinger, Peper, & Feist, 2020); it is very unlikely that homes recruited in the GHG scheme, designed to improve the energy performance of the stock, already perform to this level and therefore homes with an HTC estimate of lower than 50WK^{-1} were excluded from further analysis as they are most likely an erroneous estimate (this could result from homes having irregular heating patterns, frequently being unoccupied, or considerable unmetered heat gains such as those through party walls among other real-world effects). Similarly, few homes have HTCs greater than 1000WK^{-1}, being associated with very large and highly inefficient properties; as no such large homes were identified through EPC or survey data, homes with an estimated HTC above 1000WK^{-1} were therefore excluded from further analysis. This filter on the value of the estimated HTC was applied to all three UCL methods in this project, with further filters applied to PTG.

The Type A PTG method assumes a linear relationship between household power consumption and external temperature, and as such homes without this negative correlation present in their data will not be effectively analysed with this method. A lack of this relationship in the dataset could indicate a household which does not have a regular daily heating schedule and setpoints; this was tested on data where the average daily external temperature was less than 8^{o}C, to ensure all days were ones where space heating would be used. This was not applied to the Type B methods as regression against the temperature difference rather than the external temperature alone allows for more variation. Goodness of fit tests were then applied to the PTG results. For each PTG result the R^2 value and the coefficient of variation of the root mean square error (CV(RMSE)) were calculated. Due to the formulation of the MLR and Siviour methods R^2 and CV(RMSE) are inappropriate tests for these results; R^2 is the percentage of variation in the dependent variable that is explained by the independent variable(s) (Statology, 2024), which is clearly inappropriate for Siviour, and for MLR excludes heat gains which are integrally assumed in PTG in the balance temperature. Similarly, CV(RMSE) describes the coefficient of variation of the predicted input series relative to the actual input series (Sharma, 2024) so will also be affected by the above.

The appropriate thresholds selected for statistical tests is somewhat arbitrary here and in wider application. The thresholds applied here were adopted according to a combination of statistical norms and empirically according to the results. The R^2>0.6 threshold was determined empirically from the results using PTG, considered to strike a reasonable balance between excluding failed analyses whilst maintaining a sufficient sample size. The CV(RMSE)<50% limit was determined both based on typical use and empirically from the data (Sharma, 2024).

The criteria for a plausible SMETER result from the UCL methods in this project is this summarised as follows:

PTG	R^2>0.6; CV(RMSE)<50%; negative correlation below 8^{o}C; 50<HTC<1000WK^{-1}
MLR/Siviour	50<HTC<1000WK^{-1}

For UCL-PTG, the R^2 filter removed 39.5% of results, CV(RMSE) 6.5%, negative correlation 30.2%, and HTC range 2.4%; households would commonly not meet multiple thresholds, resulting in 48.9% of all results (i.e. 2 per house, representing the pre and post install analyses) being excluded. The HTC range filter removed 31.8% and 32.7% of results for MLR and Siviour respectively; the lack of heating season data for some homes would result in especially low HTCs being estimated (and in practice it would generally not be recommended to conduct a SMETER HTC estimate over such periods). These are the numbers across all results before and after install, and including those from households with unmetered energy consumption that are excluded in the final results; they therefore represent performance in a sub optimal scenario for the application of SMETER methods.

The plausibility criteria for a SMETER method is a crucial and integral aspect of the particular method; such criteria may be adapted according to the evolution of the method and the purpose of the analysis. For example, if high certainty in results is required, more stringent criteria may be applied. The criteria selected for UCL SMETER analysis in this project are relatively relaxed, aiming to include as many properties as possible, whilst excluding those where the SMETER model performs poorly; changing these criteria would have some influence on the results reported.

What affects the plausibility rate of a SMETER?

As discussed above, a SMETER measurement may not always successfully characterise a home. In this section the effect of different characteristics of the households and datasets on the plausibility rate of SMETERs is explored. It aims to provide insight into the most suitable homes for SMETER application and into the research agenda required to analyse more homes with SMETERs. The plausibility rate is taken to be the proportion of homes for which a SMETER result passes its own internal robustness tests and is therefore heavily influenced by the intended purpose of the SMETER. In the following discussion of plausibility rates, all results from data with any unmetered energy (non-electric or gas heating, heat pumps, PV, solar thermal, etc.) have been excluded in line with physical principles. Discussion of the impact of such factors on SMETER results is provided in later sections.

EDF and BTS have their own internal metrics for determining whether a result should be considered plausible and reported, or if data from a household is suitable for analysis with their method. EDF reported results to UCL for 54.6% of the GHG-V sample with no unmetered energy use, and BTS for 69.2% of the GHG-LAD sample with no unmetered energy use. The plausibility ratings for the UCL methods are now explored

Table 4 shows the categories investigated with regards to their effect on plausibility rating of the UCL SMETER methods using the GHG-V and GHG-LAD data for PTG and MLR/Siviour respectively; more categories are grouped for some MLR/Siviour results compared to UCL-PTG due to the smaller sample size of GHG-LAD requiring some aggregation to provide sub-categorisations of at least ten homes.

Table 4 The categories investigated for the plausibility rates of the UCL methods.

Building Characteristic	Categories	
	PTG	**MLR/Siviour**

Building type	Detached, semi-detached, mid-terrace, end-terrace	Detached/semi-detached, mid-terrace, end-terrace
EPC rating	A/B/C, D, E, F, G	C/D, E/F/G
Building age	Pre-1900, 1900-29, 1930-49, 1950-66, 1967-75, 1976-82, 1983-90, 1991-95, 1996-2002, Post-2002	Pre-1929, 1930-49, 1950-66, 1967-90
Floor area (m^2)	<50, 50-100, 101-150, 151-200, >200	50-100, 101-150
Number of occupants	1, 2, 3, 4, 5, 6, 7 and above	1, 2, 3 and above
Financial comfort (self-reported)	Living comfortably, Doing alright, Just about getting by, Finding it quite difficult	Comfortable (first two categories of PTG), Not comfortable (remaining two categories of PTG)
Regularity of heating	N/A (assessed by internal temperature measurements)	Regular, Changing setpoint, Very irregular
Dataset characteristic	**Categories**	
	PTG	**MLR/Siviour**
Percentage of daily average external temperatures below 15.5°C	<10, 10-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100	N/A
Average daily internal-external temperature difference (°C)	N/A	<7, 7-8, 8-9, >9
Range of daily average power (electricity + gas) (W)	<1000, 1000-2000, 2001-3000, 3001-4000, 4001-5000, 5001-6000, 6001-7000, 7001-8000, 8001-9000, 9001-10000	<1700, 1700-3400, >3400
Dataset length (days)	<30, 30-60, 61-90, 91-120, 121-150, 151-180, 181-210, 211-240, 241-270, 271-300	<80, 80-95, 96-110, >110

The results of this analysis are shown in Table 5 below, and summarised in Figure 12. Monotonically increasing or decreasing trends were not observed for each sub-category of household characteristics; however, the table provides a trend summary that reflects the results of the chi-squared test on proportions. Whilst it is possible that effects or characteristics exist that may explain this non-monotonic trend (potentially related to the factors listed in this table relating to the particular sub-sample created for each characteristic), this is not explored further in this work; a larger sample size would be required to enable further sub-categorisation to enable more detailed investigation into such issues.

Table 5 Results of the exploration of plausibility rate of the three UCL SMETER methods (plausibility criteria are given on the previous page) . MLR and Siviour are listed together as they produced plausible results for the same houses with fully metered energy use. The p-values are generated from a chi-square test on proportions, where the null hypothesis is that all plausibility rates are equal; a p-value of less than 0.05 indicates that there is a statistically significant difference between the plausibility rates for the different options in that category, whilst the higher the p-value the less likely that category is to have an effect on the plausibility rate. The small sample sizes in the GHG-LAD dataset analysed with the Type B methods limit the statistical power of this analysis and strongly contribute to all p values appearing high.

	PTG (Type A)			**MLR/Siviour (Type B)**		
	Highest plausibility rate	**Statistically significant plausibility changes?**	**Trend summary based on chi-squared**	**Highest plausibility rate**	**Statistically significant plausibility changes?**	**Trend summary based on chi-squared**
Building type	Detached 70.1%	$p < 0.05$	Fewer party walls is better	Mid-terrace 77.8%	$p = 0.3$	
EPC rating	G 75%	$p = 0.09$	Lower rating is better	C/D 66.7%	$p = 0.6$	
Building age	1967-1975 73.2%	$p < 0.05$	Older homes tend to have higher plausibility rates	1930-1949 81.8%	$p = 0.2$	Older homes tend to have higher plausibility rates

Floor area	200+m^2 75.9%	$p < 0.05$	Larger is better	101-150m^2 75%	$p = 0.7$	Larger is better
Number of occupants	4 71.1%	$p < 0.05$	More occupants tend to result in higher plausibility rates, up to a point	1 75%	$p = 0.9$	
Financial comfort	Living comfortably 72.7%	$p < 0.05$	More financially comfortable is better	Comfortable 75%	$p = 0.5$	More financially comfortable is better
Temperature difference	-	-		7-8^{o}C 78.6%	$p = 0.3$	Minimum daily average ΔT 7^{o}C
Proportion < 15.5^{o}C	81 – 90% 85.6%	$p < 0.05$	More data below, but not all, is better	-	-	
Regularity of heating	-	-		Regular 71.4%	$p = 1.0$	Regular heating schedule is better
Total power range (elec + gas)	8001-9000W 87.8%	$p < 0.05$	Large power range is better, up to a point	1701-3400W 100%	$p = 0.3$	Large power range is better, up to a point
Dataset length	271 – 300 days 93.5 %	$p < 0.05$	Longer is better	81 – 95 days 70.8%	$p = 0.8$	

The small sample size of the GHG-LAD homes (105 before sub-categorisation) limits the size of sub-categories in this analysis, which is likely a factor in the statistical significance results for Type B methods in Table 5. However, they do not conflict with results from the larger GHG-V sample and therefore this discussion focuses on the interpretation of the GHG-V cohort's results. The overarching finding is that the larger the heating energy use, the signal, compared to the noise, the greater the chance of a method returning an internally robust in-use HTC estimate. This "noise" relates to a wide range of factors in real homes where energy use can vary considerably according to behavioural and physical factors (Wingfield, Bell, Miles-Shenton, South, & Lowe, 2009; Porritt, 2012; Ritosa, Saelens, & Roels, 2023; Bennett, Elwell, Lowe, & Oreszczyn, 2016). Table 5 shows that this increased plausibility rate often relates to characteristics associated with an expected increase in energy demand on building physics grounds: larger homes; older homes and lower EPC homes.

It has previously been proposed that party walls may be a source of unknown heat flow, with an unidentified temperature gradient and thermal resistance, and may therefore pose challenges for SMETER characterisation (Lowe, Wingfield, Bell, & Bell, 2007). The results here cannot determine the accuracy of in-use HTC estimates compared to a well-characterised benchmark and how the different thermal conditions experienced at either side of a party element may influence the in-use HTC estimate. However, as shown in Table 5, the chi-squared test indicates a statistically significant relationship between increased number of party elements (inferred from the property type) and SMETER results failing internal robustness tests. This may be caused by the same mechanisms as the heat transfer through party walls or could relate to other factors, such as home size. Increased occupancy increases SMETER plausibility rates in this case up to four occupants, and it decreases thereafter; the causes of this trend are not clear, but likely relate to how home use changes as the number of occupants increases and may also correspond to other factors such as home size. For both party walls and occupancy, a larger sample is required to further subdivide the sample and investigate these issues in detail. Self-reported financial comfort positively correlates to SMETER plausibility rates, which could be an indicator of increased energy consumption with financial comfort, however there is little evidence for this.

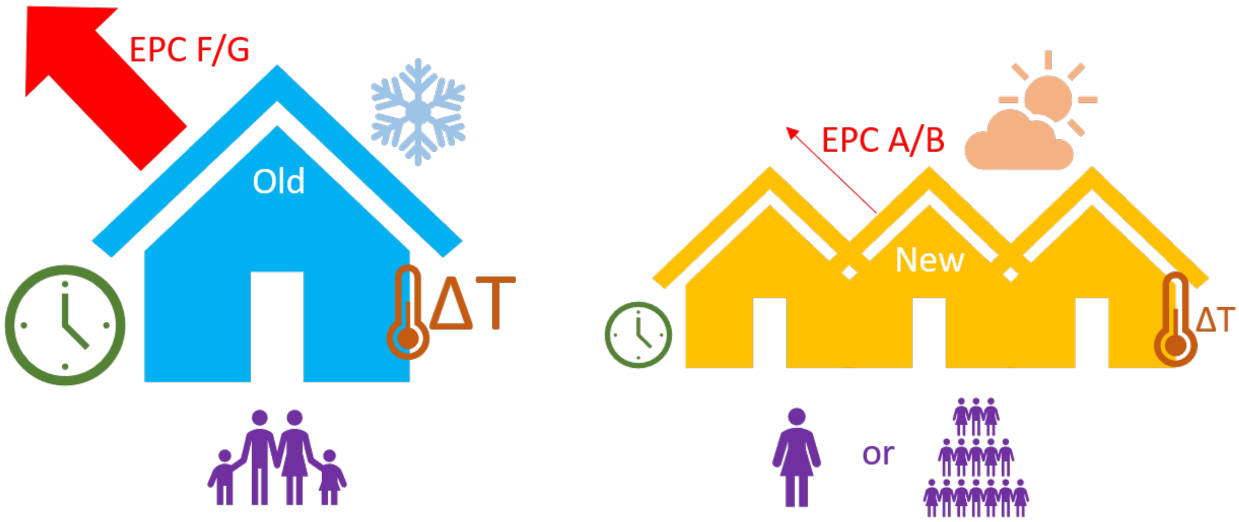


Figure 12 The homes in this analysis with the highest (left) and lowest (right) number of plausible SMETER results, as indicated by the results in Table 5. The ideal home has fewer party walls, is poorly performing, older, with around 4 occupants. It is monitored for a long time with cold external temperatures and large internal-external temperature differences in the data.

Whilst no statistically significant results are observed for Type B methods, which may be caused by the small sample size compared to the Type A sample, we also note that as expected regular heating schedules (determined by visual inspection of internal temperature profiles) and large temperature differences tend to lead to better SMETER characterisation of properties.

The most significant factor in determining plausibility rates for the Type A UCL-PTG is the length of the dataset, demonstrating that a minimum of 9 months of data is highly desirable to improve the likelihood of receiving an internally robust in-use HTC. No clear trend was observed for the impact of data length on Type B SMETER plausibility rates. Previous work (IEA EBC, 2021) and physical interpretation of the SMETER methods applied here combined with plausibility results for other variables, suggest that beyond a minimum data collection length, data quality is more important than data length for Type B SMETERs (for example the need for a large internal-external temperature gradient).This contrasts to the UCL-PTG method due to the Type A's use of non-heating season data to set base energy usage. That longer datasets are better for Type A is echoed in the results for the power range and the proportion of data below 15.5^{o}C for PTG, with the result for the latter highlighting the value of data from outside of the heating season in identifying the slope of the relationship between external temperature and power use, representing the HTC.

The results of this section provide indications of how SMETERs may perform for different homes, primarily using UCL-PTG as an example. Further insights will emerge from future research, but ensuring long datasets are recorded wherever possible and recording energy use over a range of conditions, particularly for Type A methods, maximises the chances of a range of SMETERs to successfully be applied. Further, this study suggests that the greater the heat loss, the higher the chance of SMETER success.

SMETERs and homes with unmetered energy use

As mentioned above, in line with physical principles, homes with any unmetered energy use are not generally considered suitable for analysis with only smart meter data. The large GHG-V sample enables investigation into some of the impacts of unmetered energy inputs into homes and, in particular, whether the impact of retrofit on thermal performance may be determined for these homes. This section discusses the impacts of unmetered gains to homes on the ability of Type A SMETERs to return internally robust results and the change in performance associated with retrofit; due to the presence of unmetered energy use it is not possible to estimate an HTC for these homes, where HTC reflects the efficiency of the building fabric. As mentioned above, in previous work the HPLC (heat power loss coefficient) has been introduced, which is the HTC divided by the heating system efficiency (Chambers & Oreszczyn, 2019; Hollick, Gori, & Elwell, 2020). This report introduces the grid power loss coefficient (GPLC) reflecting the amount of additional power from the grid required per degree increase in internal-external temperature difference. Results are presented for unmetered primary and/or secondary heating, heat pumps, solar thermal, and PV panels.

Plausibility rates for homes with unmetered energy use

The plausibility rates for homes with unmetered primary heating, secondary heating, heat pumps, solar thermal and PV panels are shown in Table 6, alongside results for the cohort of homes that do not contain such technologies. The results clearly show that, as expected, households with unmetered main heating have significantly lower SMETER plausibility rates than those with metered heating. This also aligns with current SMETER practice, which generally recommends against using SMETERs in such scenarios. It may be surprising that homes with unmetered primary heating can be characterised by SMETERs at all, but these “plausibility rates” relate to all metered energy use, including that associated with operating most unmetered main fuel appliances (e.g. oil and LPG boilers) including circulation pumps and other electrical demands. Unmetered secondary heating use has a lesser but still significant effect on the plausibility rate. This is likely because secondary heating methods are often used sporadically.

The presence of solar PV, solar thermal water heating, or heat pumps do not have a significant detrimental effect on the plausibility rates of the SMETERs tested; homes with heat pumps in fact more often produce a plausible result than those without (although the difference is not statistically significant). This is likely due to the more consistent nature of the heating schedule recommended for heat pumps in comparison to gas boilers, and could potentially also suggest that COP is fairly consistent. All these results of “plausibility rates” must be taken in the context of their measure of the ability of a SMETER to return a result passing internal consistency checks, rather than an indication of the accuracy of any result. This is explored in the following sections.

Table 6 The plausibility rates for homes with and without unmetered energy use. No homes in the GHG-LAD sample analysed with the Type B methods had heat pumps or solar thermal water heating.

	PTG		MLR		Siviour	
	Plausibility %	Significant difference?	Plausibility %	Significant difference?	Plausibility %	Significant difference?
Heat pump	53.7	$p = 0.3$				
No heat pump	50.8					
Solar PV	49.3	$p = 0.4$	75.0	$p = 0.6$	75.0	$p = 0.5$
No PV	51.3		67.0		66.0	
Solar thermal	48.9	$p = 0.3$				
No solar thermal	51.3					
Unmetered main heating	41.5	$p < 0.05$	45.4	$p = 0.09$	45.4	$p = 0.1$
No unmetered main heating	53.5		70.4		69.4	
Unmetered secondary heating	42.0	$p < 0.05$	54.5	$p = 0.3$	54.5	$p = 0.4$

No unmetered secondary heating	52.5		69.4		68.4	

SMETER analysis of homes with unmetered heating energy use

This section discusses the impact of both primary and secondary unmetered energy use on GPLCs derived by SMETER analysis. Around a quarter of the GHG-V sample with at least one plausible SMETER result have unmetered main or secondary heating systems, whilst the number was far lower for the GHG-LAD sample, preventing reporting on the application of Type B SMETERs due to statistical disclosure requirements. This section therefore focuses on the ability of Type A SMETERs to assess change in performance on retrofit in homes with unmetered energy use, using only results which have met the plausibility criteria outlined above, and for which there are plausible before and after estimates from a SMETER method.

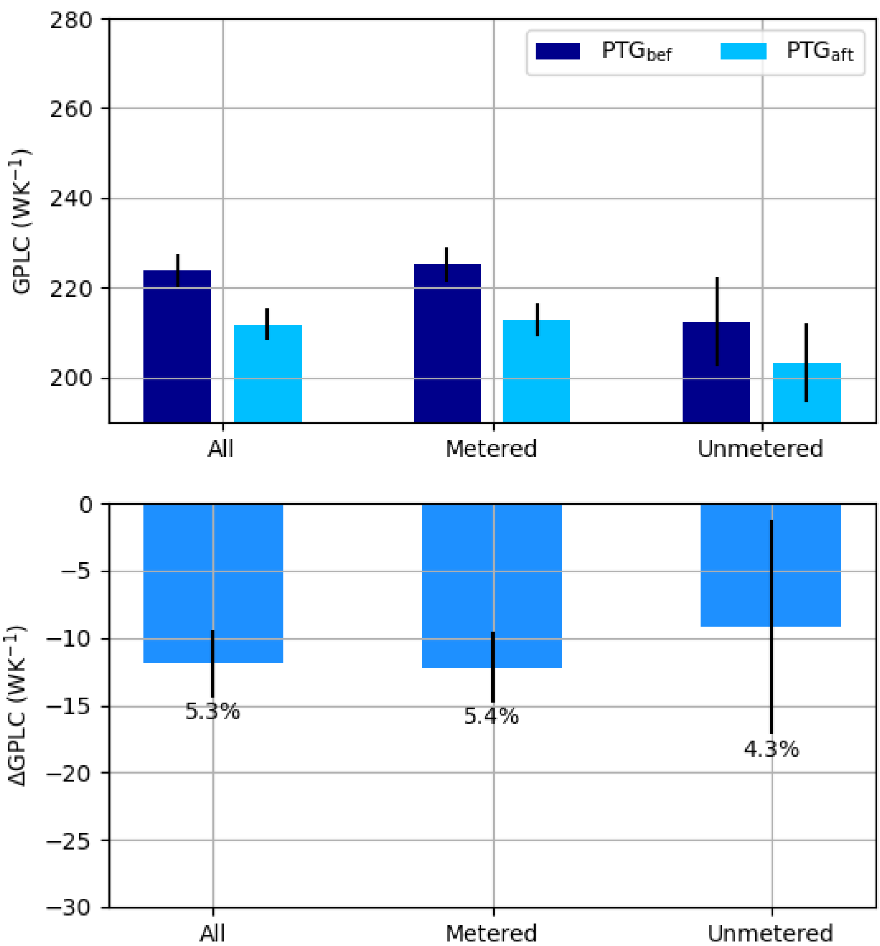


Figure 13. Top panel: absolute in-use GPLC (grid power loss coefficient) and bottom panel: mean difference in HTC pre-/post- retrofit for the UCL Type A SMETER method, split according to whether the primary heating fuel was metered. UCL-PTG reports on 662 homes

(588 metered, 74 unmetered). The y-axis on the top plot is reduced to more clearly demonstrate the differences.

Figure 13 (top) shows the mean GPLC calculated using the UCL-PTG, in addition to the standard error, for homes with and without unmetered primary heating. The change in GPLC on retrofit is shown with its standard error in Figure 13 (bottom). The presence of unmetered primary heating does not seem to have a large effect on the mean estimated GPLC values, however the standard error on the mean does increase for the unmetered heating homes. The error on the mean difference in GPLC for homes pre- and post- retrofit where there is unmetered primary heating is of similar magnitude to the difference itself; within this sample, the UCL-PTG method cannot be used to identify the change in performance on retrofit in homes with unmetered primary heating.

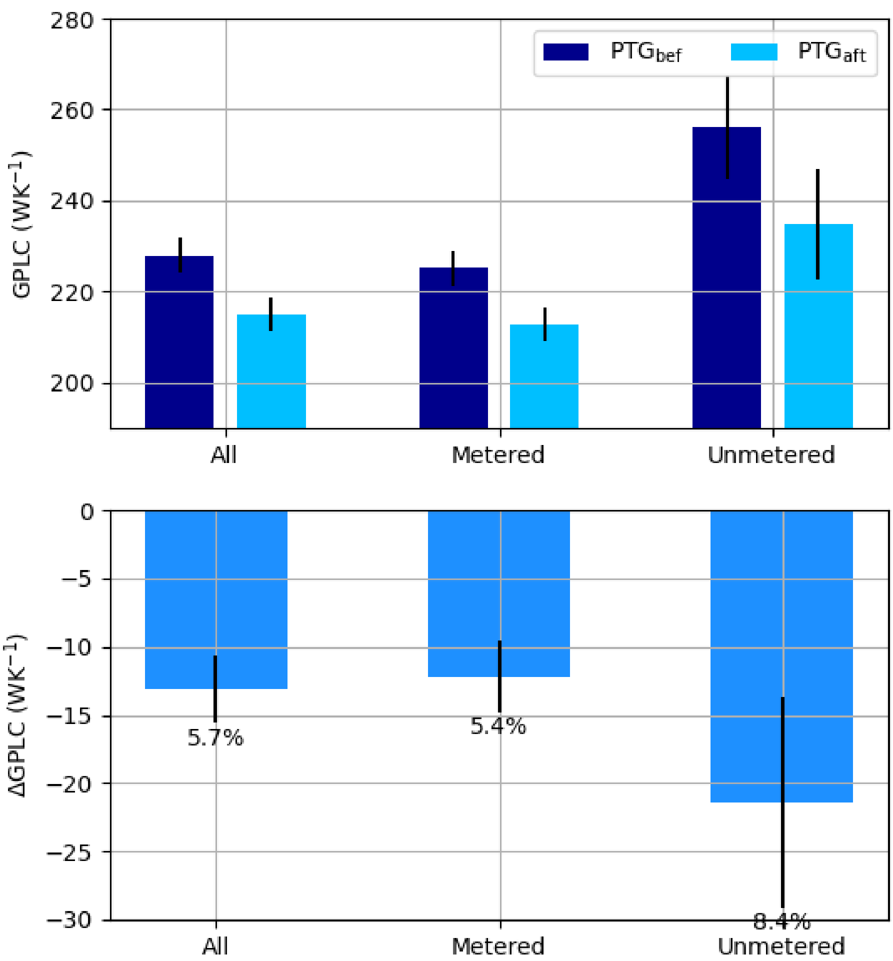


Figure 14 Top panel: absolute in-use GPLC and bottom panel: mean difference in HTC pre-/post- retrofit for the UCL Type A SMETER method, split according to whether the secondary heating fuel was metered (when the main heating was metered). UCL-PTG reports on 649 homes (588 metered, 761 unmetered).The y-axis on the top plot is reduced to more clearly demonstrate the differences.

Figure 14 (top) shows the mean GPLC before and after retrofit for homes with and without unmetered secondary heating only (i.e. their primary heating fuel is metered), with the change in Figure 14 (bottom). As for primary heating, Figure 13, the error is larger for homes with unmetered secondary heating (noting that the smaller sample size may detract from the certainty of the distribution of results), however the effect is smaller in the case of secondary heating. This may be due to the way in which many secondary heating systems are used as the survey noting presence of secondary heating does not record the frequency or duration of its use; secondary heating may or may not be used to make a significant contribution to the heating of a home in which it is installed.

These results imply that caution should be taken in applying Type A SMETERs to homes with unmetered primary or secondary heating, but that differences in performance can be identified in homes with unmetered secondary heating provided the primary heating fuel is metered.

PV panels and SMETERs

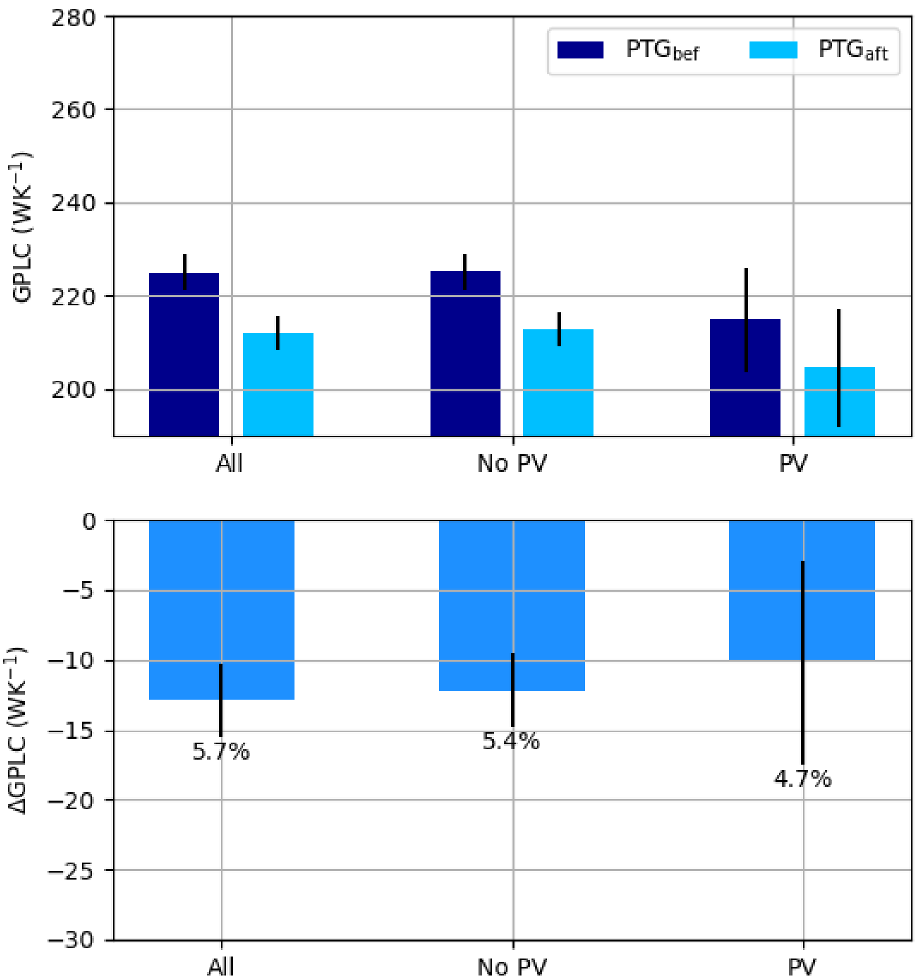


Figure 15 Top panel: absolute in-use GPLC and bottom panel: mean ΔGPLC for the UCL Type A SMETER method in homes where the primary heating fuel was metered, split according to whether the homes had photovoltaic panels. UCL-PTG reports on 651 homes (588 without PV, 63 with PV).

Figure 15 shows the impact of PV panels on performance estimated through the UCL Type A method on the GHG-V sample. Self-consumption of generated electricity from PV panels is not recorded and therefore represents an unmetered gain, potentially raising internal temperature and biasing in-use HTC estimates low. However, there is no clear evidence of any bias in this sample due to the unmetered self-use of PV generated electricity. In both cases the low numbers of homes with PV leads to large error bars. Gains from PV are relatively low in winter and UCL-PTG accounts for solar gains in the estimation of GPLC, which is likely to also at least partially address PV generation; it is likely that these factors minimise the impact of PV on in-use HTC estimates. Further research for a dataset that includes both generation data as well as smart meter data (which provides import/export readings) is required to explore this issue in detail, the impact will depend on the size of PV array, household energy use pattern and thermal efficiency of the home (such gains represent a higher proportion of total energy input to more efficient properties). The results in this study suggest that the presence of PV panels may not necessarily prevent the evaluation of a change in performance on a group of properties.

Heat pumps and solar thermal and SMETERs

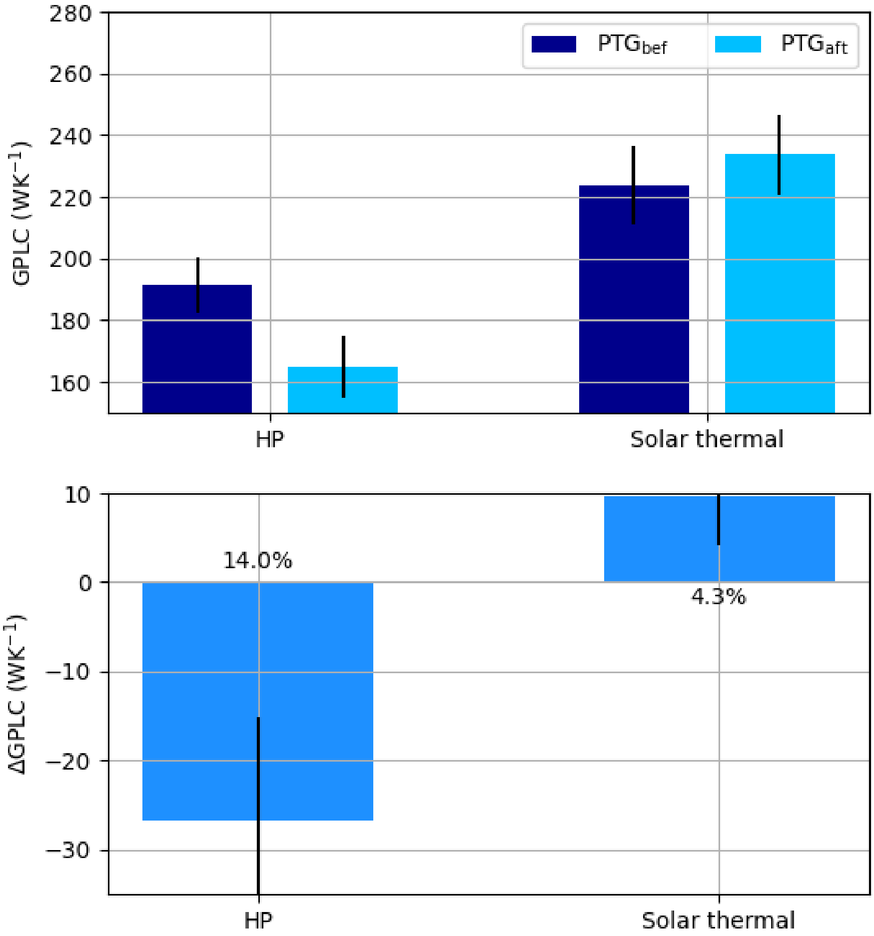


Figure 16 Top panel: absolute in-use GPLC and bottom panel: mean difference in GPLC for the UCLType A SMETER method in homes where the only intervention was the installation

of either a heat pump or solar thermal water heating. The UCL-PTG method reports on 68 homes with heat pumps, and 67 homes with solar thermal installations.

Changing from traditional space heating methods to a heat pump is expected to dramatically reduce the total energy consumption of households; with a coefficient of performance of 3, heating energy use drops to 1/3 of its starting value when heating demand matches. Installation of a heat pump generally leads to longer heated periods (Lowe, et al., 2017), leading to a consequent increase in heating energy demand, which is commonly cited at approximately 20%, but appears to be only weakly supported by empirical evidence (Watson, Lomas, & Buswell, 2021). Despite an increase in energy use associated with moving from intermittent to constant heating, installation of a heat pump is expected to significantly reduce the power required to maintain the setpoint temperatures in homes. However, whilst reductions in GPLC are observed on fitting a heat pump, they do not reduce as much as expected, as shown in Figure 16. This outcome is unexpected and potentially indicative of issues with the application of Type A SMETERs in the characterisation of homes upon retrofit of a heat pump. Since the primary difference, in terms of the temperature variation and energy input into the home, is the change from high power intermittent heating with a boiler to lower power more constant heating with a heat pump, this suggests potential issues with the application of Type A SMETERs when the scheduling of heating is changed.

The installation of solar thermal panels is not expected to change the thermal fabric of the home, although they could affect SMETERs if their output varies with external temperature (as broadly expected). Figure 16 shows no clear trend in the impact of solar thermal on GPLC; it is likely that the changes in GPLC are related to the repeatability of the methods rather than any significant change in GPLC, although some workmanship issues could also affect results. Similarly to PV panels, solar thermal water heating has relatively low power output in winter and is therefore expected to only have a small effect on SMETERs, although further research is required to confirm this.

Plausibility rates for homes in the sample for different SMETER methods

The SMETER methods employed in this project did not return results for all households in the samples analysed with them and had differing success criteria as described above. There are also issues that will apply to all SMETER methods: for example, data availability for gas smart meters is known to be lower than that for electricity meters (Webborn, et al., 2021), and some of the homes have unmetered energy consumption and this issue will apply to all SMETER techniques. Only homes with metered heating are discussed in this section.

The plausibility rates result from a combination of the quality, duration, how the home is operated and conditions during the collection of the data and the SMETER methods themselves. These issues affect the plausibility rates reported in Table 7, noting that those for Type B SMETERs were impacted by the constraints of data collection within this project. As discussed above (p18) the internal temperature measurements from some of the GHG-LAD homes do not begin until March 2022, and as such there is limited heating season data for some of the sample analysed with Type B SMETERs. Accordingly, there were fewer cold days and days with low levels of solar radiation within the dataset, both factors being expected to decrease plausibility rates. These issues, combined with the low energy use within some properties (and accordingly low internal temperatures), are a major factor in GHG-SMETER plausibility rates falling below those reported in the SMETER TEST project (Allinson, et al., 2022) and highlights the importance of planning the data collection to achieving a successful evaluation.

Table 7 The numbers of households with results returned by the UCL SMETER methods in this project. Homes which did not receive an intervention are included in the 'before' category for GHG-LAD, analysed by the MLR and Siviour methods. The plausibility rates reflect the quality of the data available in combination with the SMETER methods.

		Available houses	Houses analysed with metered heating	Houses with metered heating meeting plausibility criteria	Plausibility rate % (no unmetered)
PTG	Pre	2455	1022	665	
	Post	2455	942	663	
	Total	4910	1964	1328	54.0%
MLR	Pre	67	51	38	
	Post	44	11	5	
	Total	111	56	43	66.2%
Siviour	Pre	67	51	38	
	Post	44	11	5	
	Total	111	56	43	66.2%

As discussed above (p19)0, the sample of households in this project is not nationally representative, but the large GHG-V cohort provides an indication of the potential proportion of homes with a smart meter for which Type A SMETER methods may be able to return an internally robust result. The GHG-LAD sample is much smaller and significantly different from the national stock.

This simple comparison of the number of results does not indicate the accuracy of those results and their ability to provide useful insight into the thermal performance of the studied homes; the distribution of in-use HTCs compared to the expected HTC is discussed in the next section.

Comparison of HTCs between SMETER methods

No “ground truth” in-use HTC is known for the properties in this project and therefore SMETER methods can only be cross-compared and compared to the HTC expected on the basis of the UCL method using EPC input data, where those records exist. Such cross-comparison is discussed in this section, noting that concerns about the performance gap between expected and actual performance (Wingfield, Bell, Miles-Shenton, South, & Lowe, 2009) have been a significant driver of SMETER development. The source of this performance gap is both practice oriented and physical (Wingfield, Bell, Miles-Shenton, South, & Lowe, 2009), and whilst SMETERs aim to characterise the fabric of the home and minimise the impact of occupants on the in-use HTC, certain aspects of occupant behaviour, for example ventilation practices, will affect the in-use HTC (see in-use vs true HTC, p10). On both behavioural and physical grounds, the distributions of SMETER HTC and EPC-data Model HTCs may not match, and on an individual home level such variance is expected.

This section compares both the distributions of in-use HTCs to each other and to the EPC-data Model, and statistical measures of agreement. The normalised mean bias error (NMBE) between the Type A SMETERs and the EPC model is discussed and compared to the coefficient of variation of the root mean squared error (CV(RMSE)) as an indication of potential bias in results compared to random error (both statistical measures were also applied in the TEST Project (Allinson, et al., 2022).

Type A SMETER in-use HTCs compared

The GHG-V dataset enables insights to be drawn from the cross-comparison of two SMETER methods, PTG and EDF-Deconstruct+, and to the HTC derived from EPC input data and a simple element-based model (EPC-data Model). Figure 17 shows the distribution of in-use HTC results before interventions for the 323 homes that have results for both SMETER methods alongside EPC data to enable a model result to be derived. It shows that the EPC-data Model produces a wide range of HTC estimates; both SMETER methods estimate HTC over a more compressed range. The range of results from EDF-Deconstruct+ is significantly more compressed than the EPC-data Model and UCL-PTG. With no ground truth HTC for the homes in this project (it would require detailed experimental investigation using invasive methods such as aggregate heat loss tests), it isn’t possible to determine which range of HTCs is correct.

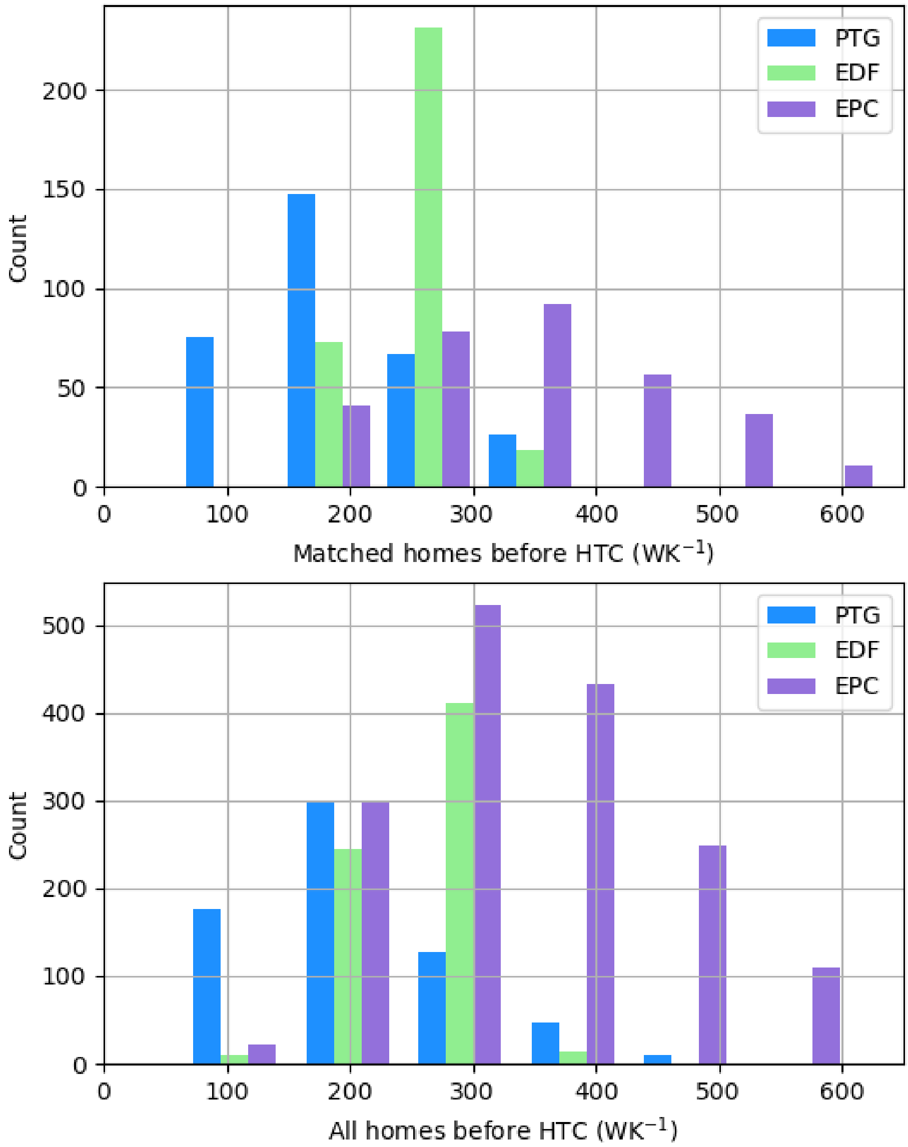


Figure 17 (Top) Distribution of before retrofit HTC results for 323 GHG-V homes which have a credible before HTC result for all 3 methods. (Bottom) Distribution of before retrofit HTC results for all GHG-V homes with a credible HTC result from any method. Bars with N<10 have been suppressed.

The difference between each SMETER method and the EPC-data Model is illustrated by Figure 18, showing that both methods tend to produce lower in-use HTC estimates than the calculated HTC based on EPC data, with both showing negatively skewed distributions; on average PTG results are 32% lower than those based on the EPC-data Model, whilst those for EDF-Deconstruct are 30% lower. Both EDF-Deconstruct+ and PTG produce mean HTCs close to those previously estimated in the literature for the average in the UK stock, also using energy signature (Type A) methods (Summerfield, Lowe, & Oreszczyn, 2010), whereas the EPC-data Model produces notably higher HTC estimates. Whilst we note that this GHG-V sample is not representative of the UK stock (p19), we also note that this finding aligns to research highlighting that EPCs generally overestimate energy use for band C-G homes (Few, et al., 2023).

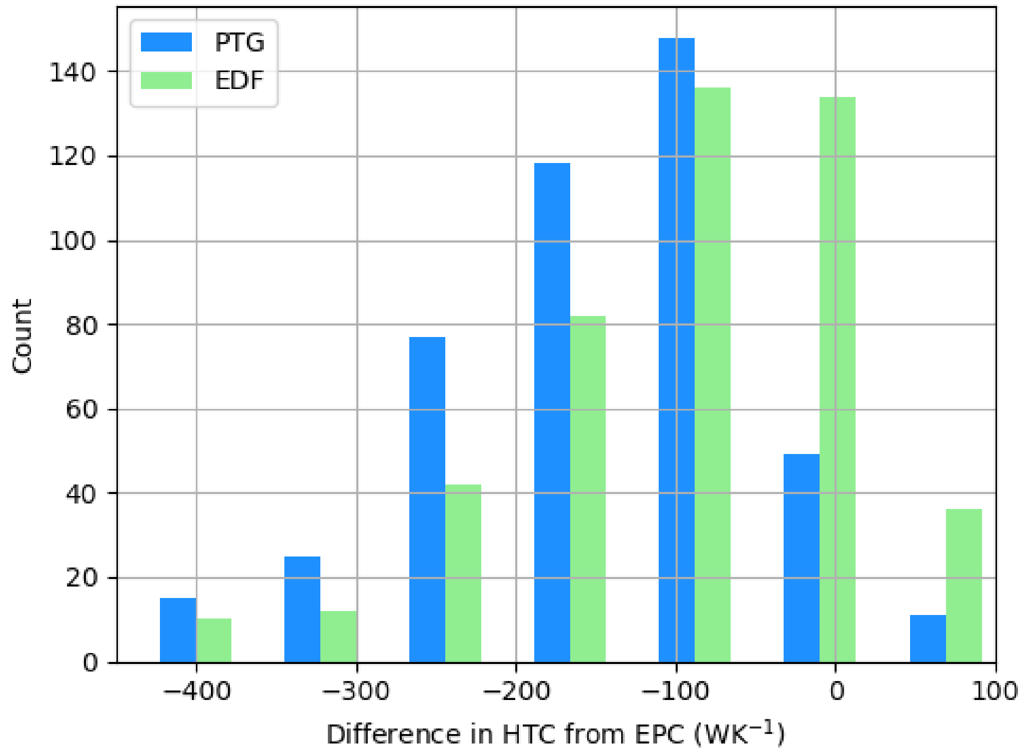


Figure 18 Distribution of the difference between the SMETER in-use HTC and the EPC-data Model HTC for the GHG-V sample, for all credible pre-retrofit estimates. Bars with N<10 have been suppressed.

The relationship between the characteristics of properties and the difference between SMETER in-use HTC and EPC-data Model HTC has been explored for the GHG-V sample. As discussed above, in this analysis the EPC-data Model is taken as a baseline for comparison and the NMBE and CV(RSME) are applied to enable a comparison between these methods but do not represent actual errors as no true value to the HTC is established. Table 8 presents the results for UCL-PTG, whilst Table 9 presents those for EDF-Deconstruct+. For UCL-PTG the general trend in both NMBE and CV(RMSE) follows building physics principles, here increasing as floor area increases and EPC band falls. Similarly, as the number of party elements increases NMBE and CV(RMSE) both increase, as found above (p44). No clear trend in error metrics is apparent in the impact of the age of properties.

Table 8 Relation between property characteristics and in-use HTC results for UCL-PTG compared to HTC calculated using the EPC-data Model.

Variable	Category	N	Mean difference (WK^{-1})	N within confidence intervals	N above CI	N Below CI	NMBE (%)	CV(RMSE) (%)
EPC rating	A, B or C	61	86.2 (32.7%)	30	0	31	48.6	71.2
EPC rating	D	268	141.3 (41.4%)	77	0	191	70.6	83.1

EPC rating	E	88	202.4 (46.4%)	17	0	71	86.5	100.8
EPC rating	F or G	24	215.7 (47.5%)	6	0	18	90.4	104.5
House type	Detached	141	161.7 (40.0%)	37	0	104	66.7	80.9
House type	Semi-detached	165	147.3 (42.0%)	54	0	111	72.5	90.3
House type	Terraced	108	140.6 (45.4%)	31	0	77	83.0	98.3
Floor area (m^2)	50 to 100	242	121.7 (41.6%)	85	0	157	71.1	85.1
Floor area (m^2)	100 to 150	151	174.9 (43.3%)	32	0	119	76.3	87.2
Floor area (m^2)	150 to 200	33	259.0 (46.4%)	5	0	28	86.7	102.1
Floor area (m^2)	More than 200	10	157.2 (26.8%)	4	0	6	36.6	53.7
House age	Before 1900	28	210.9 (47.9%)	6	0	22	92.0	108.7
House age	1900 - 1929	49	202.4 (47.7%)	9	0	40	91.3	106.3
House age	1930 - 1949	76	142.1 (38.9%)	26	0	50	63.7	77.2
House age	1950 - 1975	153	149.5 (42.4%)	41	0	112	73.5	88.2
House age	1976 - 1990	58	110.8 (34.3%)	25	0	33	52.2	65.8
House age	1991 - 2002	39	128.7 (42.2%)	12	0	27	73.1	88.5

In comparison of the EDF-Deconstruct+ method compared to the EPC-data Model baseline, as shown in Table 9, results mostly agree with UCL-PTG, albeit with lower CV(RMSE) and RMSE in all cases, reflecting the compressed distribution of in-use HTCs presented above. NMBE and CV(RMSE) both increase with increasing floor area and as EPC band drops. By contrast to UCL-PTG, this SMETER returns lower NMBE and CV(RMSE) as the number of party walls increases, indicating a closer agreement to the EPC-data Model; the origin of this effect cannot be determined but may be a combination of multiple effects, such as co-variance with floor area.

Table 9 Relation between property characteristics and in-use HTC results for EDF-Deconstruct+ compared to HTC calculated using the EPC-data Model.

Variable	Category	N	Mean difference (W/K)	N within confidence intervals	N above CI	N Below CI	NMBE (%)	CV(RMSE) (%)
EPC rating	A, B or C	68	32.9 (12.1%)	38	9	21	13.8	41.5
EPC rating	D	279	88.8 (26.4%)	105	6	168	35.8	51.4
EPC rating	E	88	160.6 (37.5%)	19	0	69	59.9	75.9
EPC rating	F or G	20	182.1 (41.4%)	2	0	18	70.6	84.3
House type	Detached	144	131.0 (32.7%)	39	1	104	48.5	63.2
House type	Semi-detached	166	98.2 (28.3%)	64	4	98	39.5	59.5
House type	Terraced	114	69.8 (22.8%)	46	8	60	29.5	50.7
Floor area (m^2)	50 to 100	264	56.4 (19.2%)	135	11	118	23.8	40.1
Floor area (m^2)	100 to 150	146	132.2 (32.9%)	25	2	119	49.0	58.8
Floor area (m^2)	150 to 200	33	263.0 (48.5%)	1	0	32	94.1	108.2
House age	Before 1900	26	155.1 (36.7%)	6	2	18	58.0	84.5
House age	1900 - 1929	51	148.1 (36.6%)	15	0	36	57.6	76.1
House age	1930 - 1949	93	118.0 (31.5%)	24	1	68	45.9	60.2
House age	1950 - 1975	153	93.4 (27.2%)	58	2	93	37.3	53.4
House age	1976 - 1990	54	83.5 (25.3%)	20	2	32	34.0	48.6
House age	1991 - 2002	35	54.6 (18.7%)	18	2	15	23.0	45.1

The results of this analysis highlight how the disparity between EPC-data Model HTCs and SMETER in-use HTCs relate to building characteristics. However, they do not indicate which method is “best” in the absence of ground truth. The results may align to previous work that has found the performance gap appears to increase as HTC increases (Few, et al., 2023) but requires further research, ideally with a larger sample to enable further sub-division of the sample and draw out clearer insights.

Dataset lengths for Type B SMETERs

As discussed above with the results in Table 5, the length of the available dataset has a significant effect on the ability of Type A SMETER methods to produce plausible HTC estimates. Although the finding was not significant for the Type B methods due to the small sample size, it follows that there will also be a minimum length requirement for these methods, the duration of which will be case and method dependent. To investigate this further, the GHG-LAD sample homes were repeatedly analysed using the UCL MLR and Siviour methods with an increasing dataset length. The results of these are shown in Figure 19 and Figure 20. For Siviour there appears to be little change in the difference from the ‘final’ HTC beyond using 21 days, whereas for MLR this limit appears slightly longer at 35 days. Likewise, the error in the Siviour results does not seem to continue to decrease beyond using 21 days of data; for MLR there does seem to be a slight continuing trend of reduced error with increasing dataset beyond 35 days, but it is largely settled at this point.

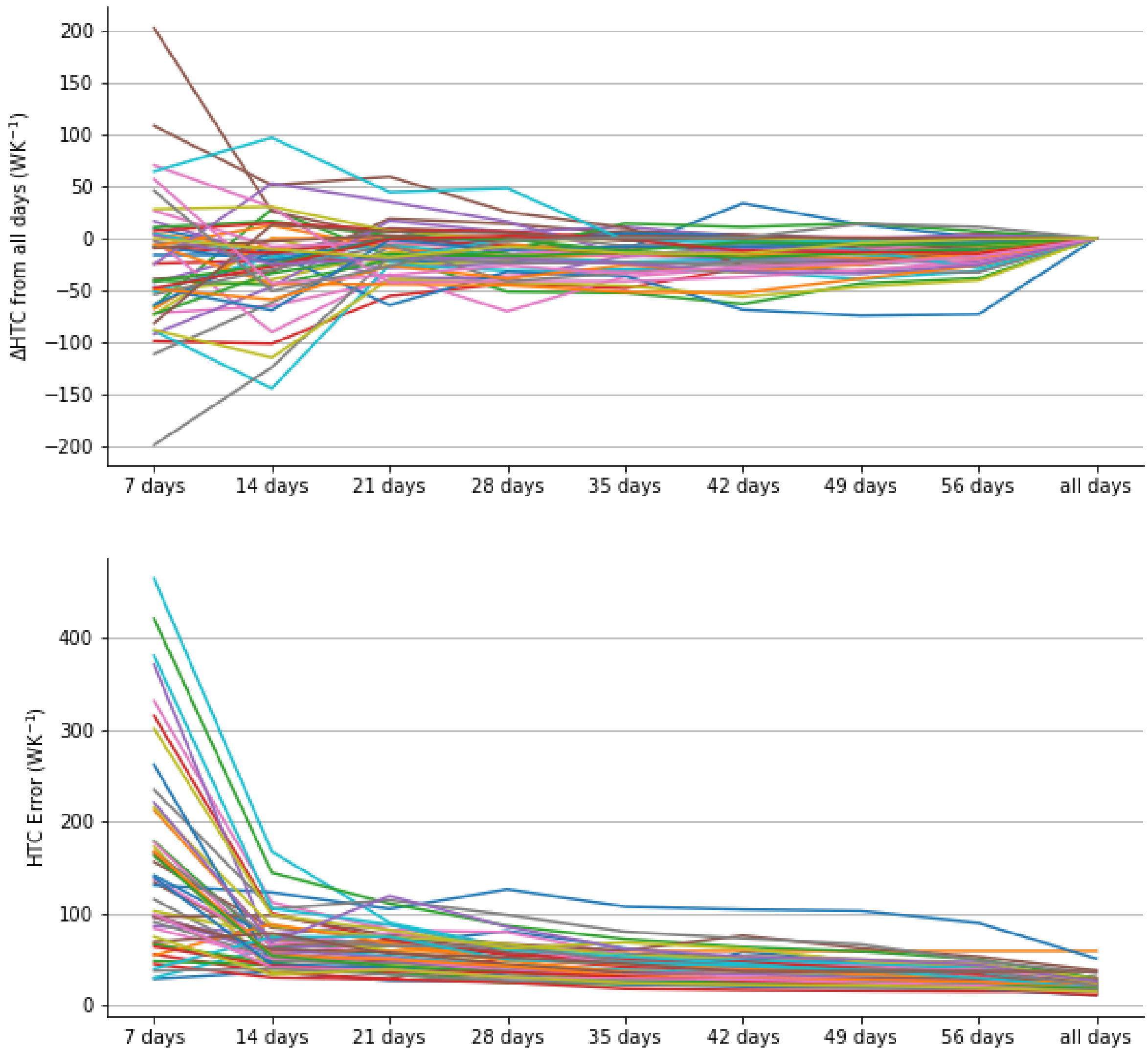


Figure 19 The change in the difference from the HTC calculated using the MLR method with the full dataset compared to that calculated with different dataset lengths for the GHG-LAD sample (top), and the change in the error on the HTC estimate calculated in the same way (bottom).

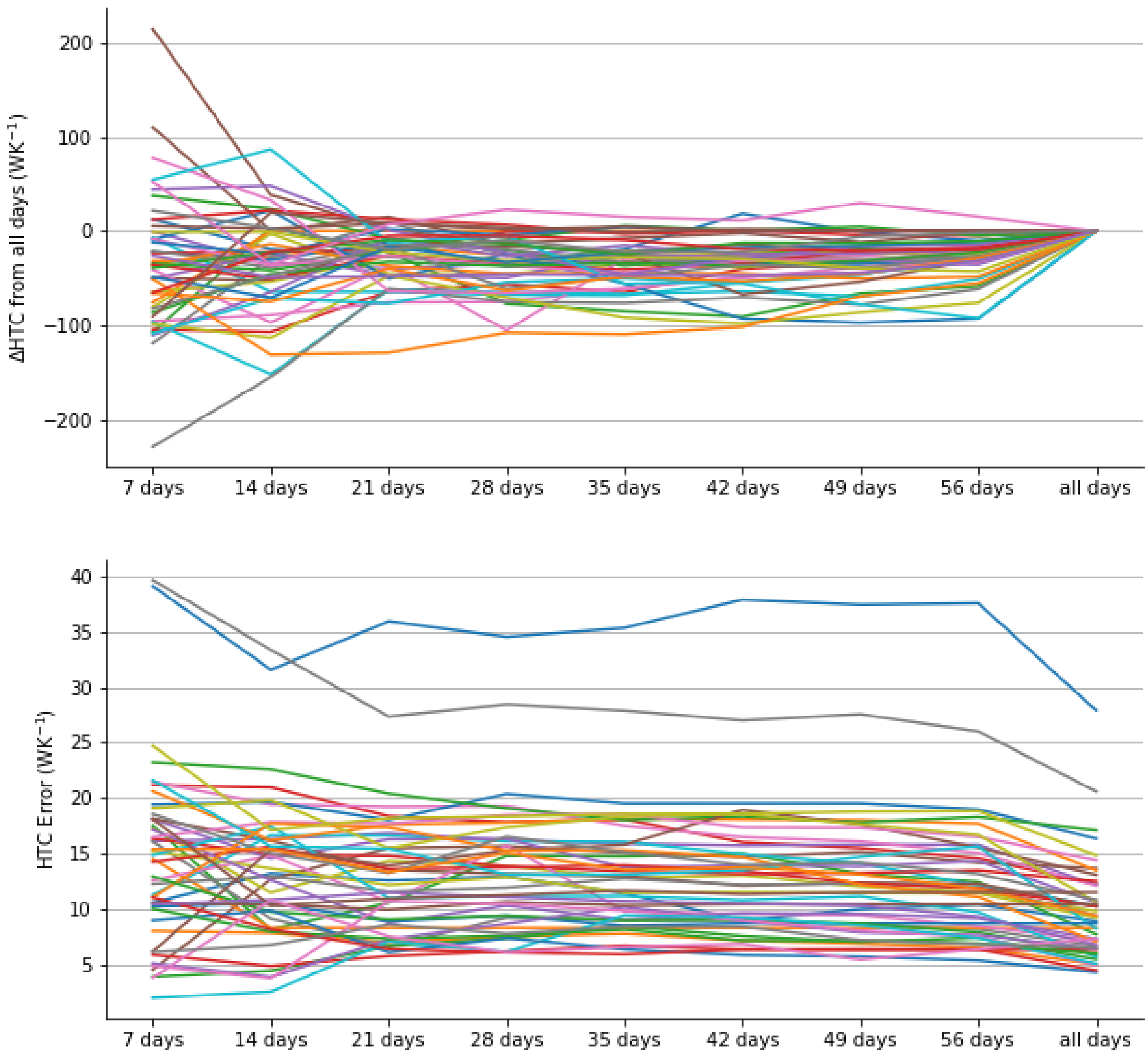


Figure 20 The change in the difference from the HTC calculated using the Siviour method with the full dataset compared to that calculated with different dataset lengths for the GHG-LAD sample (top), and the change in the error on the HTC estimate calculated in the same way (bottom).

Summary and discussion of SMETER reliability

The reliability of SMETER Type A and Type B in-use HTC results has been investigated in the GHG-SMETER project through investigation into SMETER internal robustness and the impact of household characteristics on this, the impact of unmetered energy flows on SMETERs, comparison of SMETER in-use HTCs with each other and a modelled comparator and the repeatability of SMETER results. However, what we mean by "SMETER reliability" is key and in part reflects the intended use of such methods.

The internal "plausibility" of SMETERs is used here to refer to the ability of the SMETER to return a result judged to be internally robust, by whatever methods that SMETER applies and may be simply characterised by a "plausibility rate": the number of homes for which it does so compared to those for which results fail internal checking or have insufficient data. The plausibility rate, and internal reliability or robustness, of SMETERs is therefore dependent on a combination of the quality, duration, and conditions during the collection of the data. It is also strongly dependent on the choices made in classifying what data is acceptable and identifying acceptable results, for example a statistical test passed at a certain level. Such choices are not informed by theory and are instead based on conventions, assumptions or practical considerations. For example, as applied in this work, UCL-PTG applies inclusive rather than stringent tests, aiming to maximise the sample size of homes with SMETER results. Increasing the stringency of tests would decrease the number of homes with a result, but increase the robustness of the sample of homes passing the tests (a randomly selected home is more likely to have an internally robust in-use HTC) and could result in clearer research outcomes in absolute in-use HTCs or changes to HTCs. This comes at the cost of a potential reduced statistical power over the whole sample due a reduction in sample size; in this research the value of large samples to determining the impact of retrofit on in-use HTC is explored below (p8282).

The impact of household characteristics on plausibility rates of the UCL-PTG SMETER was investigated through the GHG-V sample; this large sample enabled statistically significant trends to be identified with a Chi-squared test whereas the small samples of the GHG-LAD scheme did not return statistically significant results for the UCL-MLR and Siviour methods despite the lower uncertainty estimates on individual results. Focusing on the Type A sample, the longer the dataset the better the chance of returning a successful SMETER result with UCL-PTG, up to 9 months which is also reflected in other data characteristics such as the range of energy use recorded and the presence of data within and outside the heating season. This duration of data capture for UCL PTG exceeds the length of the heating season (and uses non-heating season data to evaluate baseload energy usage), albeit at low cost and without disruption as no in-home monitoring is required. This contrasts with the typical requirements for Type B methods, which generally focus within the heating season, ideally with a large internal-external temperature difference. In the SMETER TEST project (Allinson, et al., 2022), dataset lengths from 7 to 74 days were reported for Type B methods, with professionally installed equipment, compared to 2-3 months duration in this work with self-installed sensors.

The Type A sample also indicated that when higher heat loss is expected on building physics principles (larger, older homes and lower EPC bands), SMETER methods were more likely to produce plausible results than those with lower expected heat loss. This is expected since the purpose of the SMETER is to determine the thermal performance of homes: the larger the metered energy use (signal) and the heat loss through the fabric, the smaller the impact of unmetered gains and of variable factors such as ventilation (noise).

Unmetered energy use, studied using the GHG-V sample and Type A methods, was found to decrease SMETER plausibility rates for both unmetered primary and secondary heating. This is expected as the use of unmetered heating may or may not be directly related to metered energy use. For example, for an oil-fired boiler the metered energy use for its associated electrical demands, such as the circulation pump, increases as the boiler provides more heat to a home but may not do so linearly, providing a weakened relationship between metered and unmetered energy demand. By contrast, metered energy demand may be decreased by the use of unmetered secondary heating, such as a wood stove, but how it does so depends on the regularity of the use of the stove, and its heat output, and therefore is expected to reduce plausibility rates. Results also show that unmetered heat potentially biases estimated in-use HTC low (the home needs less metered heating) and consequently ΔHTC on retrofit is low (as this is a relative measure).

The impact of PV panels on plausibility rates and estimated in-use HTCs for the UCL-PTG Type A method is less clear, despite the potential for significant self-consumption. It is notable that this method attempts to address the potential bias of solar heat gains into homes, via a regressive relationship. Such methods are likely to, at least in part, capture or minimise the impact of PV generation and self-consumption on in-use HTC estimates. We also note that PV generation is low in the winter months when heating use is highest, plus many homes do not self-consume a high proportion of their total demand at this time (Webborn, et al., 2021).

The total SMETER plausibility rate for the GHG-V sample with Type A method was similar for both UCL-PTG and EDF-Deconstruct+, around ~54% with exclusion of homes with unmetered heating. Similarly, the results from all Type B methods (UCL-MLR, UCL-Siviour and BTS SmartHTC) were plausible for approximately 67% of the GHG-LAD sample. The addition of internal temperature measurement is associated with a higher plausibility rate, as expected because variations in heating demand and internal conditions can be accounted for. This rate would potentially be higher were planning and implementation of the study to enable longer data collection, particularly during the coldest months.

The plausibility rates reported here are significantly lower than those observed in the SMETER TEST project, where “success rates” comparing SMETER results to the co-heating test HTC fell in the range 70-97%, using self-reported 95% confidence intervals and a standard confidence interval for co-heating. There are numerous reasons why this could be the case including several which are already understood, including differences in the timing and duration of the measurements. . However, it is also likely driven by challenges that emerge when generating in-use HTC results across large samples, such as the use of self-installed equipment and the automation of data selection and analysis. Firstly, the impact of self-installed equipment here compared to professionally installed equipment in SMETER TEST cannot be determined, but the incidence of inappropriately installed sensors may be considerably higher. Repeated visits to homes by professionals in SMETER TEST, and strong engagement of the landlord (Halton Housing) may also have helped ensure that homes and heating systems were operating as intended, whilst we have no knowledge of this for the homes in the GHG-V and GHG-LAD schemes. Secondly, automatic data selection and model choice are required to manage the large datasets of this project compared to the 30 homes in SMETER TEST. The UCL Type A and Type B methods applied here do not use advanced data filtering to identify the most suitable data for analysis. However, data filtering and selection (for example to best identify when model assumptions are met) is inherent to a SMETER method; development of such components of the methods is expected to improve the ability of the UCL-MLR and UCL-Siviour analyses to return internally robust results. These are challenging to implement over a portfolio of properties where baseline and heating power use is highly variable, as reflected in individual energy use patterns.

The in-use HTCs estimated in this work were compared to the HTC calculated by the EPC-data Model; in-use HTCs estimated by UCL-PTG and EDF-Deconstruct+ were found to be significantly lower than those predicted based on the simple building physics model. This could be caused by a disparity between EPC data derived HTCs and in-use HTCs; recent work has identified a significant performance gap between energy use measured and that predicted in EPCs, with increasing gap as the EPC band drops (Few, et al., 2023). Further research is required to better characterise and diagnose this issue.

In-use HTC is variable due to both physical effects, such as changes in ventilation rate as the internal to external temperature gradient changes, changes to technologies within the property (e.g. issues with the boiler or ventilation) and practices of the occupants. A true estimation of the in-use HTC experienced should therefore vary accordingly and may do so significantly if, for example, previously underheated rooms become heated or if ventilation is increased or decreased according to changes in the perceived needs of the household. This changing in-use HTC estimate may provide occupants with the information required to understand the impact of their actions on the thermal performance of their home, but presents challenges for determining the inherent performance of the fabric or impact of retrofit. Development of SMETERs and test protocols is required to better understand whether these have an impact on HTC estimation and determine how an in-use HTC may be estimated that is more reflective of the fabric thermal performance, minimising the impact of services and occupants.

The large number of homes within the GHG-V scheme recruited to this study enables a range of insights to be drawn and challenges identified; however, further research is required to clarify and further investigate findings. The reliability of SMETER results is method and dataset specific and whilst general findings of this work have wider implications for the field, they cannot be directly transferred to new situations.

Impact of GHG-retrofit on in-use HTC

The impact of fabric retrofit on in-use HTC, delivered under the Green Homes Grant Voucher Scheme, is discussed in this section. This change in in-use HTC is compared to the change in HTC expected on the basis of a model utilising EPC input data and the type of retrofit undertaken according to BEIS records. The results focus on Type A methods applied to the GHG-V data due to the lack of retrofits during data collection for the GHG-LAD sample.

Calculation of the expected change in HTC due to retrofit relies on both the quality of EPC data that provides a base model and GHG-V database that should detail the work that was undertaken. The poor data quality detailing retrofits was discussed above (p1322) whilst the construction of the EPC-data Model to predict ΔHTC (p3434) requires detailed knowledge of the retrofit. The treated areas (e.g. whole wall area for external insulation or just one façade, whole loft or one section) were not known, nor were sufficient construction details to accurately predict the pre- or post- retrofit thermal performance of these elements. This represents a major limitation to the work and it was not possible to tell how much of the of the difference between EPC-data Modelled and SMETER-estimated thermal performance this accounted for. However, we also note the findings above (p5757), that the SMETER estimated in-use HTCs of homes is significantly lower than those calculated using EPC input data. A lower than expected pre-retrofit HTC corresponds to a higher than expected total thermal resistance of the properties

Thermal resistance sums as the reciprocal of the resistance of individual layers ($R_{total} = \sum_k 1/R_k$), where R_{total} is the total thermal resistance, and R_k the thermal resistance of the k^{th} element. It follows that when the HTC of an element is lower than expected, corresponding to a higher than expected thermal resistance, retrofitted insulation has a lower than expected impact on total heat loss. The results above therefore suggest that the SMETER-estimated impact of GHG retrofit should be lower than that expected based on an EPC-data Model.

The following sections present the change in in-use HTC derived by the UCL-PTG and EDF-Deconstruct+ models, compared to the change in HTC estimated using the EPC-data Model.

Result plausibility and presentation

Based on the analysis presented in the previous sections, results of the mean difference in HTC, split by insulation intervention, are presented in the following section. The method-specific plausibility criteria above were applied. Results are presented if they meet the following additional criteria, for all methods:

A single measure was installed
The main heating output is captured by the smart-meter consumption measurements and could be collected via the DCC
The secondary heating is metered (if present)

- The properties did not have solar thermal water heating or heat pump space heating
- The properties did not have an electric vehicle charging at home
- The properties did not have photovoltaic electricity generation (except for results explicitly reporting on this)

Note that homes are included unless we have evidence that any of these criteria are not met, for example if the participant did not answer the survey question regarding electric vehicle ownership, we include them in the analysis.

HTC changes due to GHG-V retrofit

The change of in-use HTC (ΔHTC) following GHG-V retrofit is discussed in this section, estimated by UCL-PTG, EDF-Deconstruct+ and the EPC-data Model. Figure 21Figure 21 shows the distribution of the in-use ΔHTC following retrofit for the 146 homes for which all methods returned an estimate, according to the availability of data and the reliability requirements applied to each and also the distribution with different samples per method. It shows a broad range of predicted ΔHTC based on EPC input data, a broad distribution of in-use ΔHTCs calculated using UCL-PTG and a much narrower distribution of results for EDF-Deconstruct+. Both UCL-PTG and EDF-Deconstruct+ return some positive changes in HTC after retrofit.

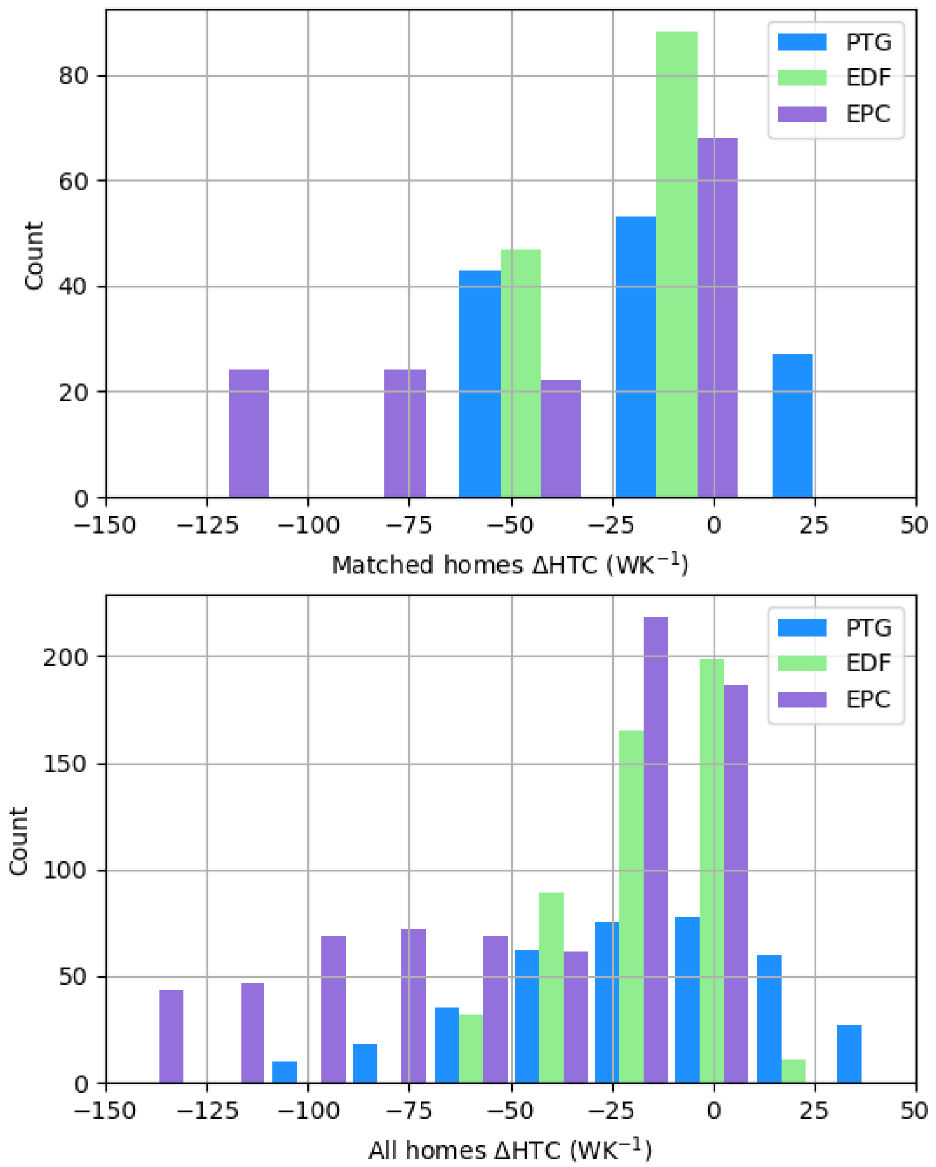


Figure 21. (Top) Histogram of ΔHTC as calculated by UCL-PTG, EDF-Deconstruct+ and EPC-data Model method for 146 homes with plausible HTC results for both before and after retrofit for all three methods. (Bottom) Histogram of ΔHTC as calculated by the three methods for all homes with plausible before and after retrofit HTCs from each method. Bars with N<10 have been suppressed.

The repeatability of in-use HTC estimation using SMETERs is critical to their use for evaluating the impact of retrofit on the thermal performance of homes, as in this project. Typically, the repeatability of a measurement may be taken to be the consistency of results when the same parameter is measured successively under identical conditions. As discussed above (p10), the in-use HTC of a home is not constant and instead varies according to environmental factors (such as windspeed) and occupant behaviour (such as internal door or window opening). As such, we should not expect in-use HTC results to exactly match when estimated from different data since the in-use HTC represents the period during which data has been measured.

The variation of in-use HTC that may be routinely observed, or directly associated with specific mechanisms, is not explored here. However, a snapshot of this issue is provided by the pre-

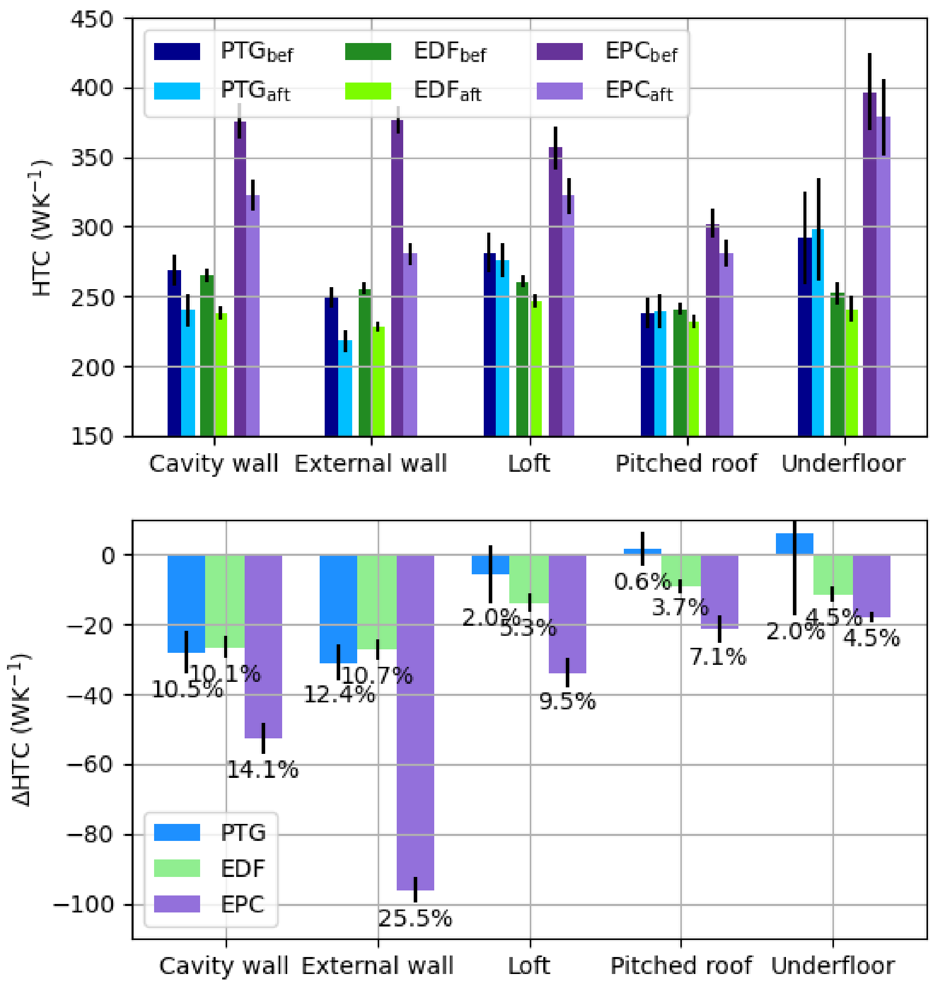
/post- retrofit in-use HTCs for the UCL-PTG method where a significant number of post-retrofit in-use HTCs were higher than those before installation of the GHG measure. If no changes were made to the property apart from retrofit, with no changes in behaviour (window opening, heating operation etc), retrofit is obviously expected to reduce the HTC. Indeed, even with very low-quality retrofit it is highly unlikely that any home would perform worse after retrofit than before it; thus these results indicate that the model is unable to fully represent the thermal behaviour of the home, such as changes to behaviour, controls or technologies within the home.

The mean ΔHTC (± standard error on the mean) following retrofit for all matched homes is (-57.7 ± 4.7)WK^{-1} for the EPC-data Model, (-19.6 ± 4.0)WK^{-1} for UCL-PTG and (-21.5 ± 2.2)WK^{-1} for EDF-Deconstruct+. This shows that the EPC-data Model predicts much larger changes in HTC than are found by either the UCL-PTG or EDF-Deconstruct+ methods, and that both empirical methods find very similar results on average as noted above (p5757). All homes with plausible in-use HTC pre- and post- retrofit are included in this analysis, including those identified above where the SMETER estimates an unexpected increase of in-use HTC. Whether this results in a systematic bias or higher "random" error (which may not be truly random) associated with the ability of the SMETER to accurately represent the thermal performance of the home is not known and cannot be determined using the data available to this project.

Figure 22 shows the mean before and after HTC for the GHG-V homes calculated using different methods and split by insulation measure. Note that to maximise the sample size matching homes are not presented for each method (all homes with a successful HTC/in-use HTC estimate both pre- and post- retrofit are included in each instance). This figure clearly illustrates the significant difference between HTCs estimated using EPC input data and in-use HTCs and highlights the generally good agreement between UCL-PTG and EDF-Deconstruct+.

The in-use ΔHTC associated with GHG retrofit, shown in Figure 22, is also significantly lower than that predicted by the EPC-data Model, aligning to expectation as discussed above (p57,68). The large standard errors for underfloor insulation for both UCL-PTG and EPC-data Model are associated with a relatively low number of properties with this retrofit (N<30).

The same trends to Figure 22 are shown in Figure 23 which shows the pre- and post- retrofit HTCs and the ΔHTC on GHG retrofit for the reduced sample of matched homes across all analysis methods. As found in Figure 22, the EPC-data Model predicts a much larger change in HTC on retrofit that the in-use HTC estimated using either SMETER. The UCL-PTG and EDF-Deconstruct+ results show good agreement, with the mean falling within the standard uncertainty for all measures except pitched roof insulation (for which there are only 25 results).



N	PTG	EDF	EPC
Cavity wall	94	94	141
External wall	136	138	179
Loft	131	118	186
Pitched roof	118	124	129
Underfloor	29	26	48

Figure 22 (Top) The HTC calculated from different methods for each insulation measure, showing the mean and standard error and (bottom) the corresponding ΔHTC for insulation measures. These results present data from all households where methods were successfully applied before and after, and the main heating fuel was metered (i.e. different homes are presented for each method). The accompanying table shows the sample sizes for each measure and method.

It is not possible to identify the cause of the observed large difference between EPC-data Modelled and SMETER in-use HTC results. In particular, due to the lack of high-quality retrofit data and ground-truth HTC or expert survey, it isn't possible to distinguish between inaccuracies caused by assumptions of the retrofit undertaken or the physical properties of the homes compared to the EPC (such as non-representative U-values in modelling), any bias in SMETER results and differences in retrofit performance due to workmanship issues, or material selection. However, for external wall insulation the EPC-data Model predicts a ΔHTC almost three times that observed with in-use HTCs; this is not explainable through the walls simply performing better than expected pre-retrofit[1]. Whilst it isn't possible to explore the cause of this disparity in greater detail, we highlight the assumption that the full external wall area is insulated in retrofit, which was made in order to calculate the post-retrofit HTC using the EPC-data Model, may or may not represent the actual retrofit undertaken, this again highlights the importance of data quality in interpreting results.

ΔHTC upon GHG retrofit including uncertainty analysis

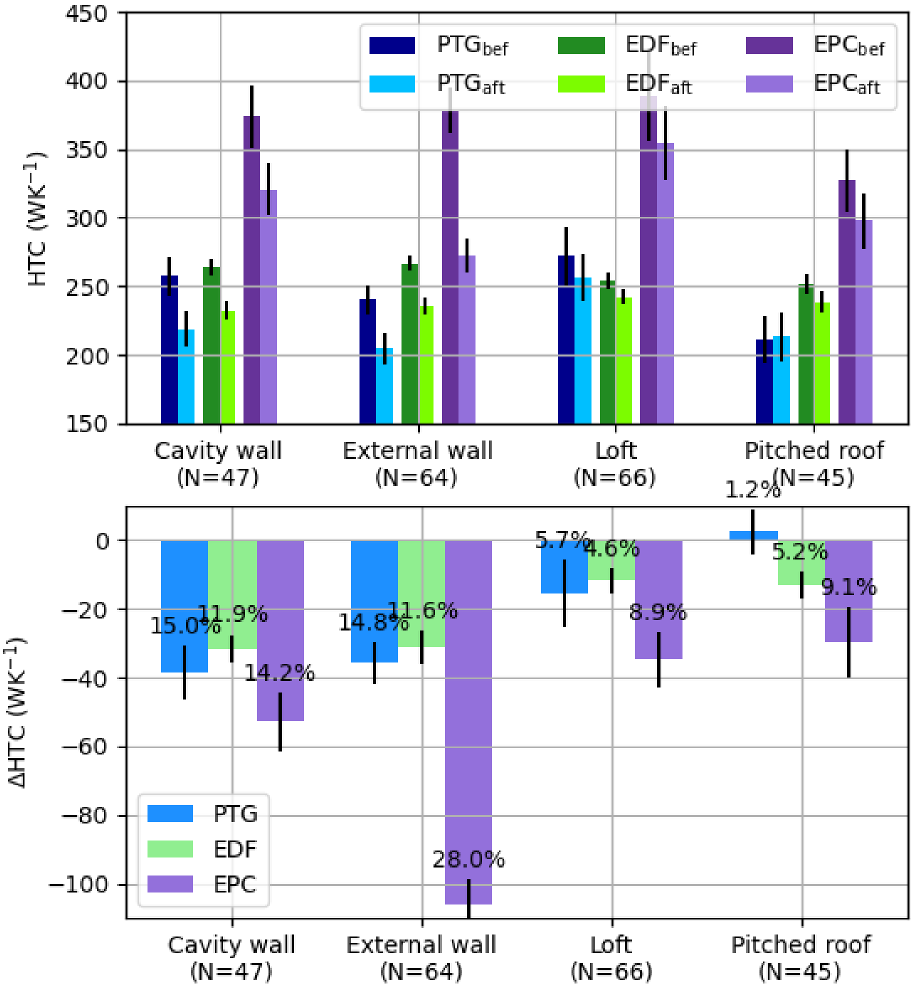


Figure 23 (Top) The HTC calculated from different methods for each insulation measure, showing the mean and standard error and (bottom) the corresponding ΔHTC for insulation measures. These results present data from households where all methods were successfully applied, and the main heating fuel was metered (i.e. the same homes are presented for each method) and therefore represents a reduced number of in-use HTC results than previous figures. Results for underfloor insulation are suppressed as they fall below the 10 sample minimum required for statistical disclosure.

---

[1] As an example consider a highly simplified model of a semidetached home with solid walls, sized at the average floor area of a typical UK home. Pre-retrofit conductive losses through the wall are approximately 132WK^{-1} with an initial U-value of 1.6 Wm^{-2}K^{-1} (107WK^{-1} if we assume a starting U-value of 1.3Wm^{-2}K^{-1}) and final conductive heat loss of 22WK^{-1}. Retrofit aligned to PAS 2035 where all walls are treated should result in a much larger decrease in HTC than observed.

The distribution of pre- and post- in use HTC of homes retrofitted through the GHG leads to the mean ΔHTC discussed above, alongside the distribution around this mean, represented here by the standard error. However, the SMETER methods used in this work derive the in-use HTC and the associated uncertainty in it. This section investigates the agreement between methods to estimate ΔHTC considering their uncertainty intervals.

Errors pre- and post- retrofit (p28) were added in quadrature, assuming independence of each estimate[2], and multiplied by 1.96 to obtain the 95% confidence interval for UCL-PTG ΔHTC:

$$CI_{\Delta HTC} = 1.96\sqrt{\alpha_{HTC,b}^2 + \alpha_{HTC,a}^2}$$

Where $CI_{\Delta HTC}$ is the 95% confidence interval for ΔHTC, and $\alpha_{HTC,b}^2$ and $\alpha_{HTC,a}^2$ are the uncertainties on the before and after in-use HTC respectively.

EDF provided 95% confidence intervals for each of their before and after HTC results, these were simply added in quadrature to obtain 95% confidence intervals for EDF-Deconstruct+ ΔHTC:

$$CI_{\Delta HTC} = \sqrt{CI_{HTC,b}^2 + CI_{HTC,a}^2}$$

Where $CI_{HTC,b}^2$ and $CI_{HTC,a}^2$ are the confidence intervals on the before and after in-use HTC results respectively.

Table 10 Proportion of in-use ΔHTC (+/- uncertainty bands) in encompassing with EPC-data Model ΔHTC, for matched samples of homes between methods.

Measure	Proportion of empirical ΔHTC (+/- uncertainty) encompassing EPC-data Model ΔHTC		Number of homes
	UCL-PTG	EDF-Deconstruct+	
Cavity Wall	0.70	0.48	27
Loft	0.59	0.50	44
Pitched Roof	0.63	0.46	24
External Wall	0.23	0.09	43
Flat Roof	1.00	0.50	4
Underfloor	0.60	0.80	5
All	0.51	0.38	150

[2] In uncertainty analysis errors may be added in quadrature if independent, or simply added if this isn't the case. Since different parts of the extended timeseries data was used for in-use HTC estimates pre- and post- retrofit, it was assumed that the errors are independent (e.g. errors in smart meter data). However, we acknowledge that some errors are not independent, such as any associated with the occupants. These confidence intervals therefore represent a lowest case and are conservative in determining when ΔHTC estimates agree.

Table 10 shows the proportion of homes for which the EPC-data Model ΔHTC was within the uncertainty of the SMETER methods. Only homes for which estimations from all three methods were available are presented in this analysis. In line with SAP modelling, no uncertainty is associated with the HTC calculated using the EPC-data Model; on this basis it agrees with the empirical method in 51% of cases for UCL-PTG and 38% of cases for EDF-Deconstruct+. However, there is notable variation by measure. The agreement is worst for external wall insulation, with 23% agreement for UCL-PTG and 9% for EDF-Deconstruct+. Small sample sizes throughout this comparison reduce the likelihood of agreement between methods (see p8282 for discussion of the required sample sizes). As previously highlighted, the EPC-data Model predicts a large change in HTC for external wall insulation, but the SMETERs find a much smaller change on average.

The mean ΔHTC for the retrofit measures that can be reported (minimum 10 homes) and associated mean error on each estimate are presented in Table 11. This highlights that on an individual property level, these Type A SMETERs are not suitable for determining the efficacy of retrofit for this sample of homes. The change in HTC on retrofit for individual homes is further explored in the following section.

Table 11 Mean ΔHTC according to three methods of estimation, with mean uncertainties for methods using measured data. These are for the matched samples of houses across all three methods.

Measure	Mean ΔHTC (WK^{-1})			Number of homes
	UCL-PTG (mean error)	EDF-Deconstruct+ (mean error)	EPC-data Model	
Cavity Wall	-38.0 (44.5)	-32.1 (33.7)	-52.4	27
Loft	-13.7 (70.2)	-11.8 (31.2)	-36.1	44
Pitched Roof	2.4 (37.1)	-11.8 (34.0)	-31.0	24
External Wall	-33.5 (40.7)	-31.1 (29.7)	-105.6	43
All	-19.6 (49.7)	-21.5 (31.9)	-57.7	150

ΔHTC on retrofit and home contextual characteristics

The relationship between ΔHTC and the characteristics of homes is explored in this section. The change to in-use HTC upon retrofit is expected to vary according to physical features of the property, such as the floor area (greater *surface* area is associated with greater losses for otherwise identical buildings and increases in surface area will normally be reflected in increases in floor area), property age (due to changes in construction techniques) and the starting energy efficiency of the home. Such factors form the basis of the EPC-data Model, where the EPC survey is used to identify the characteristics and dimensions of the building. However, the ability of SMETERs to accurately estimate the thermal performance of buildings depends on how the SMETER model represents the real energy demand, which varies according to a large number of factors including those physical components of a building and the practices of occupants (p30). As the energy demand reduces (lower area, greater insulation etc), the impact of errors that don't relate to total heating demand (e.g. data errors) increases. Similarly, as the HTC decreases, the proportion of non-heating energy use increases compared to heating demand, the impact of changes to ventilation (such as window opening) potentially increases and the proportion of demand from water heating is expected to increase. The relationship between ΔHTC and the characteristics of homes was therefore investigated to both illustrate how in-use HTCs vary and whether any trend in the relationship between in-use and modelled ΔHTC is apparent. The following characteristics were investigated:

- Building age – this is often related to the thermal performance of the building, as different wall types and insulation levels are common in different periods of construction. SERL survey question B9, "approximately when do you think your accommodation was built", was used for this analysis.
- Building type – this is related to the external wall area compared to party wall area, with detached homes having the most external surfaces and flats the least. SERL survey question B1, "what type of accommodation do you live in", was used for this analysis.
- EPC current energy rating (A-G) – these bands are related to the EPC-modelled cost of heating, lighting and hot water for the home, with homes that are the most expensive receiving G-ratings, and cheapest costs receiving A-ratings. The least efficient homes will tend to be the most expensive (although the fuel type also impacts the cost), and the least efficient homes will tend to have larger improvements to HTC following retrofit.
- Floor area – larger homes generally have larger external surface areas than smaller ones, and so it is expected that they would tend to have larger improvements in HTC following retrofit. The EPC variable reporting total floor area was used for this analysis.

A subset of these variables are presented below (building age and EPC current energy rating).

The reliability of SMETERs in returning HTC estimates is discussed above (p5541), with the number of results for each method summarised in

Table 7. Dividing the sample into those with a SMETER in-use HTC followed by categorisation according to insulation measure and then contextual characteristic led to small groups of homes in many cases, reducing the statistical significance of results, groups with small samples are not presented. The largest possible sample sizes are presented, whereby all plausible results are included for each HTC estimation method, without requiring matched samples where all three methods return a result.

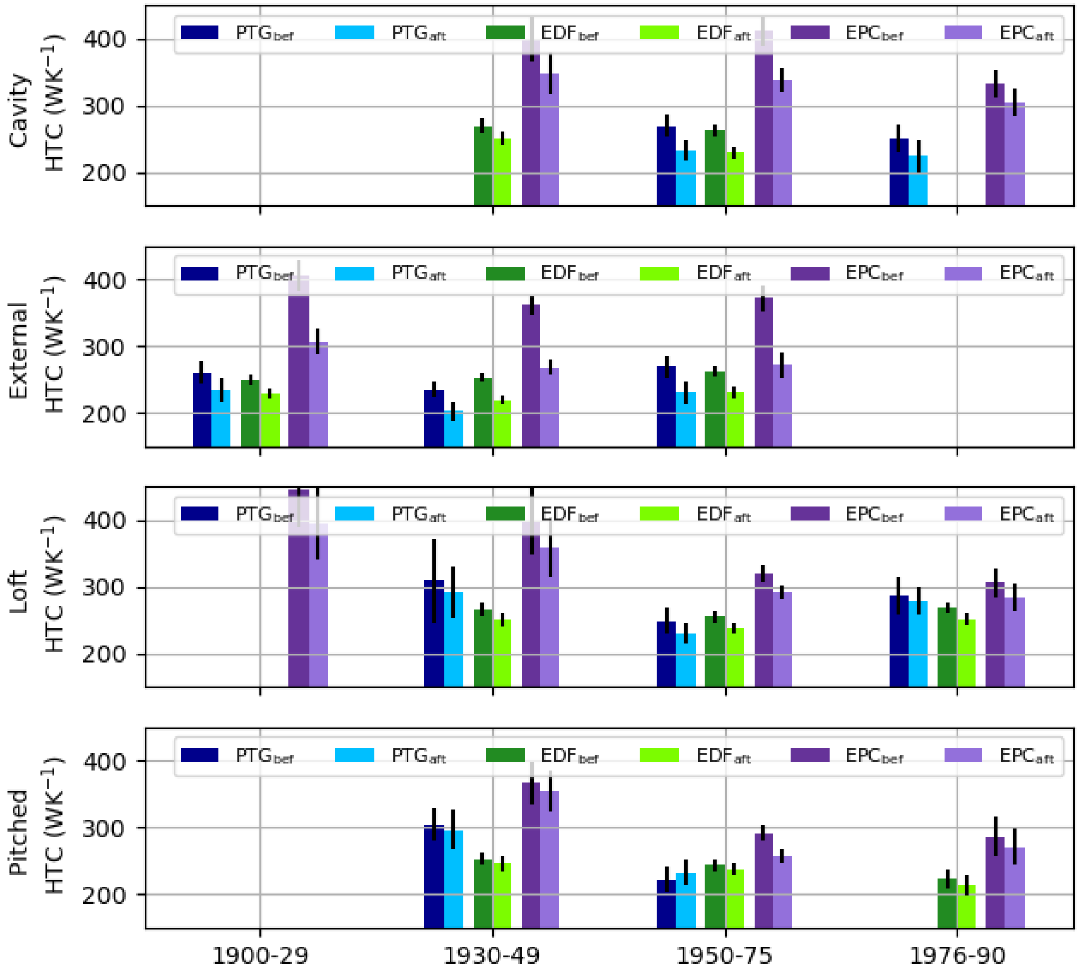


Figure 24 Mean HTC split by insulation measure and building age for UCL-PTG, EDF-Deconstruct+ and the EPC-data Model. Note that homes were not required to have results for all three methods available so different groups of homes are represented by each bar.

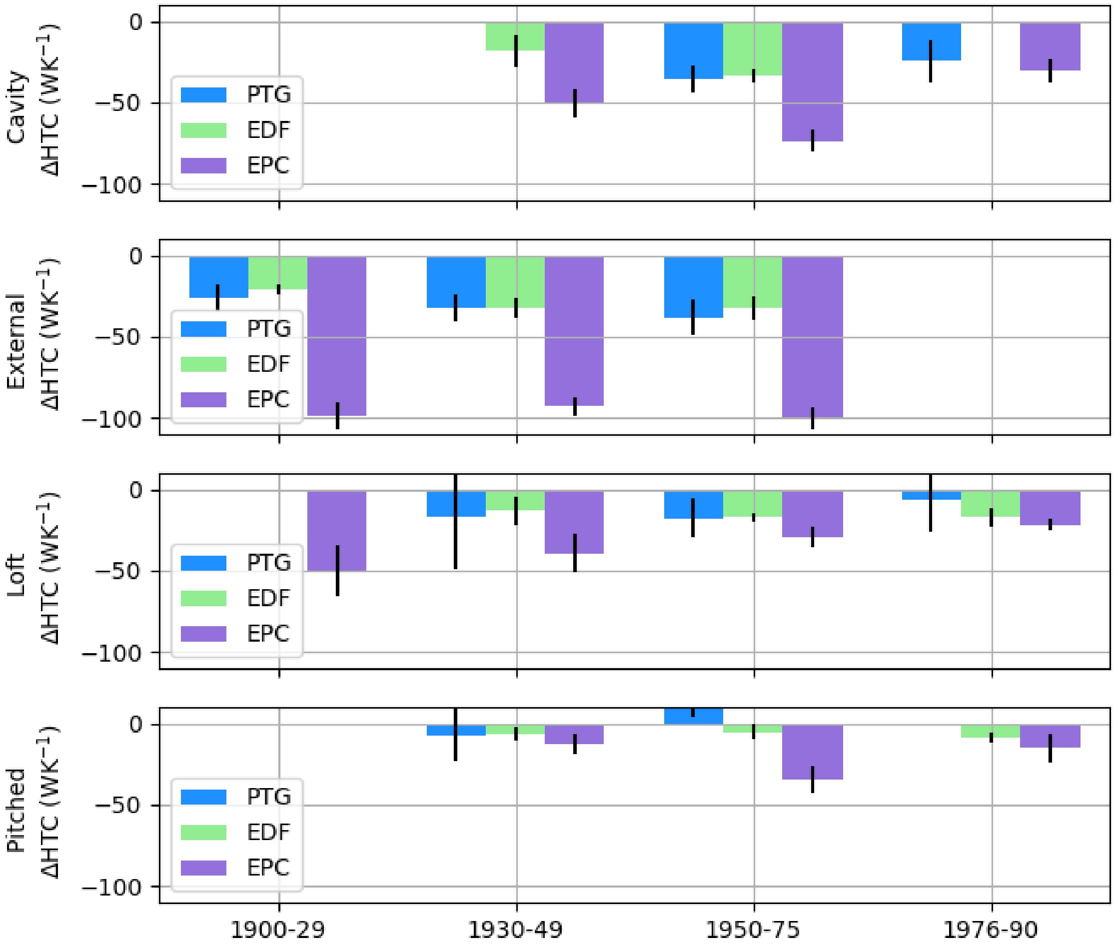


Figure 25 Mean ΔHTC split by insulation measure and building age for UCL-PTG, EDF-Deconstruct+ and the EPC-data Model. Note that homes were not required to have results for all three methods available so different groups of homes are represented by each bar.

Figure 24 and Figure 25 show the in-use and EPC-data Model HTCs of GHG-V homes in the study and the ΔHTC on retrofit. The trend in the EPC-data Model, of generally decreasing HTC and ΔHTC for newer buildings, illustrates the expected improvement in thermal performance for newer buildings due to changes in construction materials and methods; variation of other factors such as building size may also contribute to this outcome. However, the in-use HTCs and ΔHTCs for both UCL-PTG and EDF-Deconstruct+ exhibit generally slight reductions in HTC as the age category of properties becomes more recent and less clear trends in ΔHTC. Larger samples are required to determine if such trends are present in matched homes across all methods, with contextual variables refined to minimise the effects of known influences such as floor area, and whether any performance gap can be linked to specific aged properties.

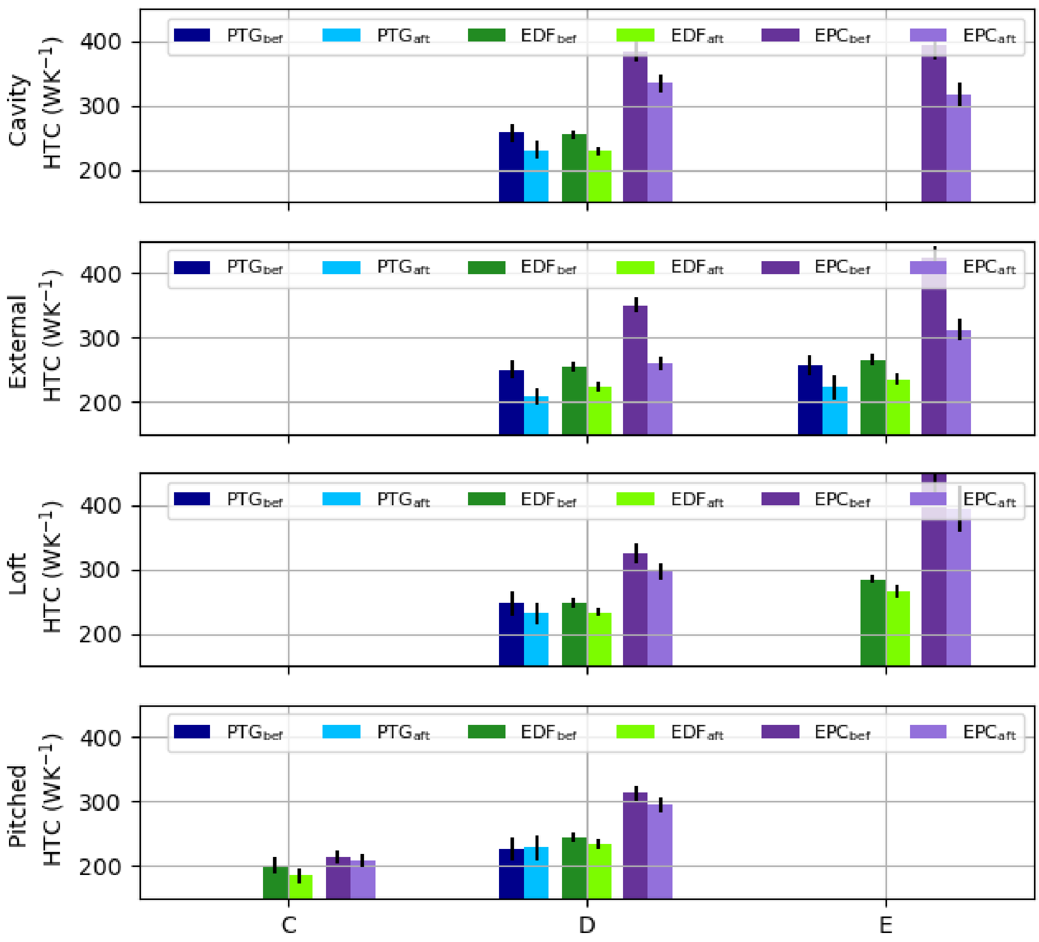


Figure 26 Mean HTC split by insulation measure and current energy efficiency rating for UCL-PTG, EDF-Deconstruct+ and EPC-data Model. Note that homes were not required to have results for all three methods available so different groups of homes are represented by each bar.

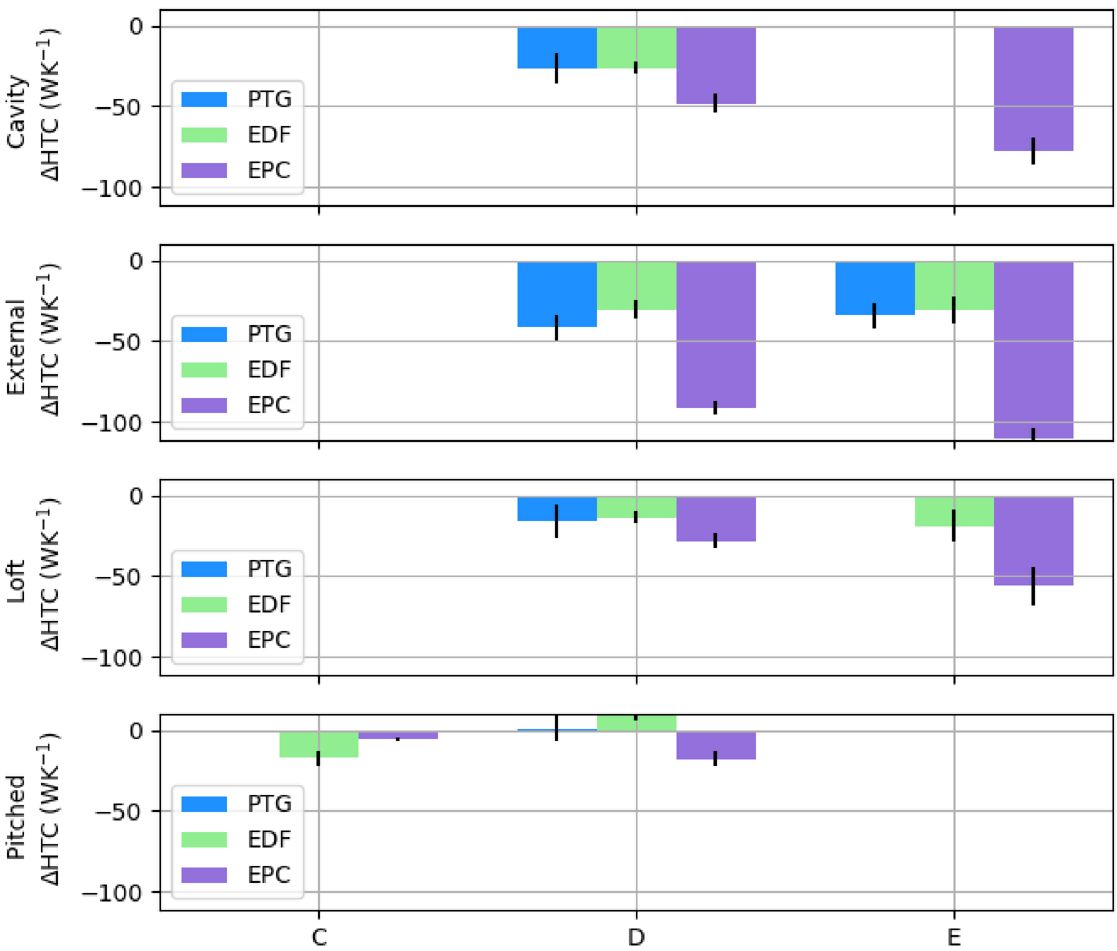


Figure 27. Mean ΔHTC split by insulation measure and current energy efficiency rating for UCL-PTG, EDF-Deconstruct+ and EPC-data Model. Note that homes were not required to have results for all three methods available so different groups of homes are represented by each bar.

Figure 26 illustrates the low range of EPC bands in the sample, a combination of the number of homes with EPCs (p23) and the EPCs of those homes (p19). The sample is weighted towards lower EPC bands, as expected since they are installing GHG measures, which are targeted towards worse performing homes. As expected, both the EPC-data Model and SMETER in-use HTC values are higher for lower EPC band homes. Similarly, Figure 27 shows that the EPC-data Model estimates larger ΔHTC values for lower EPC bands for all interventions. This is expected since homes in lower efficiency bands are generally assumed to have poorer thermal performance and the installation of insulation measures should improve their performance by a greater amount than those in higher EPC bands. The in-use ΔHTC is less clear; few results are available when subdivided by both EPC band and retrofit measure. However, for both external wall insulation and loft insulation Figure 27 shows no significant difference between the ΔHTC in bands D and E. This may be due to the real thermal performance of the homes pre- and post- retrofit, or by the ability of the applied SMETERs to represent this. A larger sample may enable a clearer result to be identified.

SERL Observatory control

Inter-year comparison of energy performance may be affected by factors that change between the years, such as large differences in energy prices (note this study took place before the energy price increases following Russia's 2022 invasion of Ukraine), weather that isn't adequately characterised by the SMETER or incremental energy improvements across the stock. A control analysis of the SERL Observatory was therefore undertaken to investigate whether any stock-level effects may be taking place that could influence results of the GHG-SMETER study. The SERL Observatory is a nationally representative sample of over 13,000 GB households providing smart meter and contextual data.

A random 10% sample of 1324 homes from the SERL Observatory (Webborn, et al., 2021) was analysed with the UCL-PTG method with data from 2020 and 2021, as a control to the GHG-V sample. Figure 28 shows the distribution of HTC results for the SERL control sample and the same for the recruited GHG-V households. Whilst both plots suggest a shift to lower in-use HTC in 2021 compared to 2020, the change in mean in-use HTC (± standard deviation) for the SERL Observatory is small: (234.4 ± 99.4)WK^{-1} in 2020 vs (228.8 ± 103.7)WK^{-1} in 2021. The reduction of in-use HTC for the SERL sample could result from home efficiency improvements to homes in the cohort, behavioural changes or uncertainties in SMETER results; we note that impacts of the Covid-19 pandemic could affect energy use and home operation over the duration of this study.

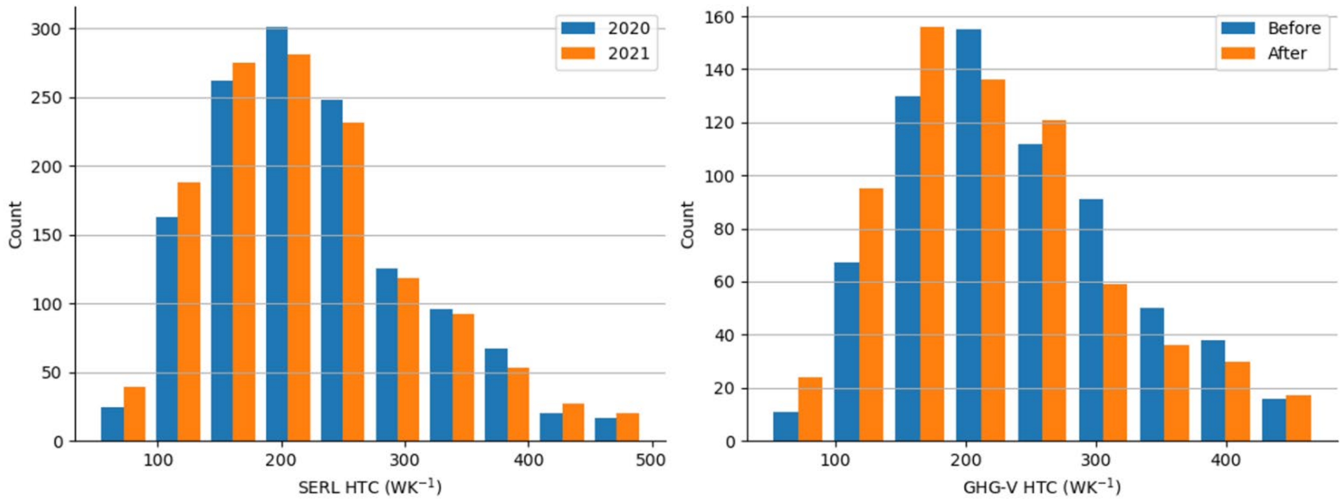


Figure 28 The before and after in-use HTC distributions for the SERL control sample (left) and the GHG-V sample that received fabric retrofits (right).

T-test results demonstrate a significant difference in before and after in-use HTC for the GHG-V sample receiving fabric retrofit measures (t(1390)=3.36, p=0.0008), whilst the difference between the SERL control sample across 2020 and 2021 is not significant (t(2706)=1.42, p=0.155). This suggests that the GHG-V scheme has, as expected, led to significant improvements to the fabric of the retrofitted homes and that any stock-level effects that could influence the estimated in-use HTC are not significant.

The results for both samples are summarised in Figure 29, again demonstrating the increased difference among the GHG group of households.

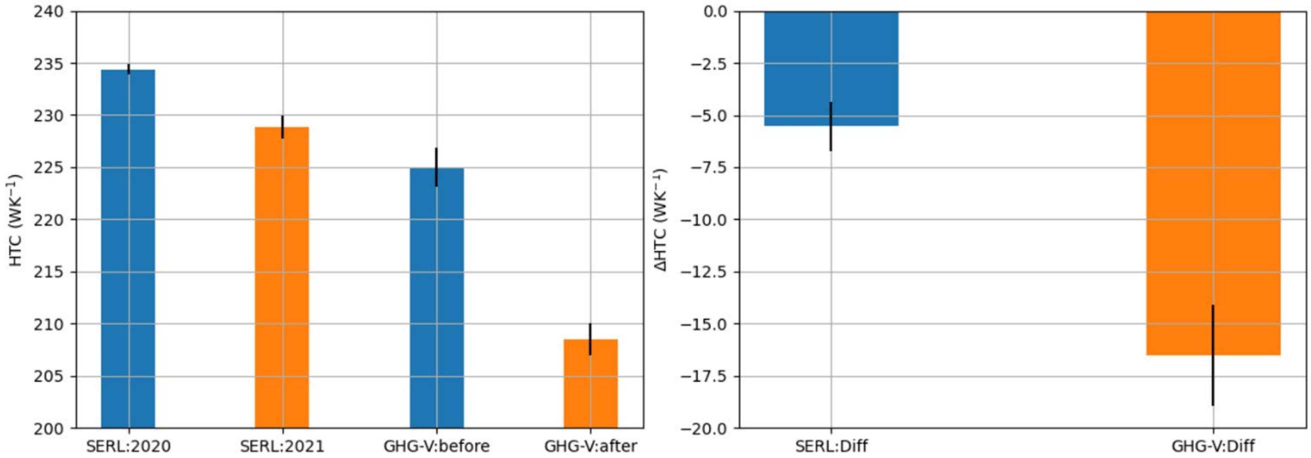


Figure 29 The mean HTC before and after and the ΔHTC with standard errors on the means across the sample for the SERL control group and the GHG-V sample.

Sample sizes required to identify the impact of retrofit

This section discusses the aggregation of results to investigate the impact of retrofit, or retrofit quality, over a sample of homes rather than for an individual home.

Whilst SMETERs have been shown to produce reasonably precise and repeatable estimates of the in-use HTC of individual homes in the TEST project (Allinson, et al., 2022), estimating a change in in-use HTC due to an intervention is a different challenge. It requires the uncertainty around the before-and-after HTCs to be small enough to reliably conclude that a meaningful change has occurred. For some measures, the expected change in HTC is relatively small and typically exceeded by the uncertainty around an HTC estimated by a SMETER (and potentially any empirical method), which means it would not be possible to detect the impact of the measure.

For example, in the SMETER TEST project (Innovation Competition) the mean confidence intervals were 18 and 21% for the two best performing SMETERs. These were Type B SMETERs with accuracy better than RdSAP over certain metrics. However, by comparison, Figure 23 shows that this uncertainty is greater than the mean expected ΔHTC for all but the most impactful measures examined in this study (i.e. solid wall insulation). The mean confidence intervals for the two SMETERs in the TEST project which were most like the Type A SMETERs were even higher, at 26% and 33%, which would rule out identifying the impact of the typical installation of those measures in this study too. Recent research has also estimated the uncertainty of co-heating tests, testing requiring unoccupied homes to derive a standardised measure of the HTC (p11), to be 5-27% over 14 case studies (Gori, D, Bouchie, & Stamp, 2023). , A ”natural variation” in HTC was identified due to changes to factors such as the ground temperature, infiltration and moisture effects which may exceed the ΔHTC of individual homes upon retrofit, they note that there is “little benefit” in more accurate measurements than this natural variation of HTC (Gori, D, Bouchie, & Stamp, 2023).

Research is ongoing to improve empirical methods to determine the thermal performance of homes and test methods may be published or developed that can determine smaller differences in HTC. However, an alternative approach is to study the average impact of retrofit over a sample of homes, rather than individual homes. This improves the statistical power and decreases uncertainty in the mean HTC estimate.

The size of sample required to have confidence in interpreting results depends on the applied SMETER method, the characteristics of the data to be analysed and the accuracy required to meet the purpose of the analysis (for example, investigation into a systematic performance gap in loft insulation requires high accuracy of results due to the expected relatively small change in thermal performance compared to wall insulation). The impact of sample size on the uncertainty of estimates of ΔHTC for the population of homes in the GHG-V study was investigated by repeatedly resampling homes from this population and analysing the results. This analysis assumes that the population has a mean ΔHTC and that sampling from it enables estimation of that mean. Such resampling was undertaken for both UCL-PTG and EDF-Deconstruct+ for insulation measures with a large number of households in the GHG-V group: cavity wall, external, loft, pitched roof and all measures combined. Sample sizes were increased from 10 to the total number of properties in the relevant category, in increments of 10; 100 resamples were undertaken for each sample size. This method therefore illustrates the application of subsamples of SMETER results to represent the total population of results.

Figure 30 shows the impact of resampling on the mean ΔHTC, the standard deviation on that mean and the combined error in ΔHTC for UCL-PTG and EDF-Deconstruct+. All three panels illustrate the increased stability of drawn samples as the sample size increases. Whilst variation in the standard deviation reduces as the sample size increases, tending towards the standard deviation of the whole cohort, it does not decrease with larger samples – this is a measure of the shape of the distribution of results, rather than an indication of certainty of the mean. By contrast, the error on the mean decreases as the sample size increases. Repeated measurement decreases uncertainty in the mean; the “acceptable” uncertainty depends on the required insights, but here we note an uncertainty of approximately 8WK^{-1} at a sample of 50 households, reducing to ~6WK^{-1} at 100 and ~4WK^{-1} at 200 for PTG; and approximately 1WK^{-1} at a sample of 50 households, reducing to ~0.64WK^{-1} at 100 and ~0.5WK^{-1} at 200 for EDF-Deconstruct+.

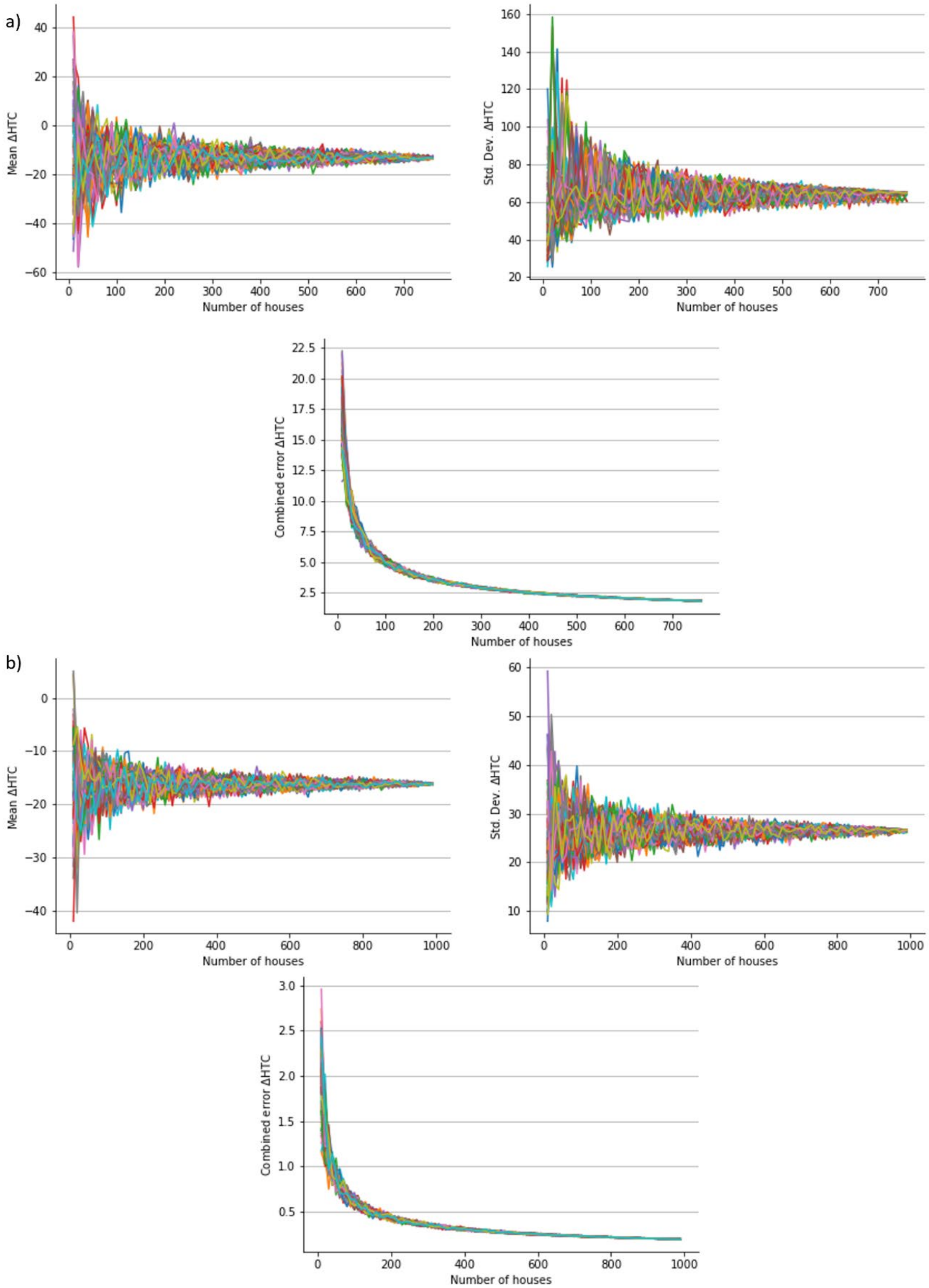


Figure 30 The repeated resampling results for all fabric retrofit PTG results (a) and EDF-Deconstruct+ results (b). All units are WK^{-1}.

The amalgamation of all retrofit measures increases the uncertainty on the mean ΔHTC on retrofit because the expected ΔHTC varies across measures and the physical properties of homes. Categorising into groups where the ΔHTC is expected to be more similar should decrease the combined uncertainty and this is potentially suggested in Figure 31 which shows the results for loft insulation alone ; whilst for a sample of 50 homes the uncertainty on the mean is similar to the whole sample, at 80 homes it has reduced to ~4WK^{-1} for PTG and ~0.6WK^{-1} for EDF-Deconstruct+. Larger sub-division of the GHG-V cohort is not possible due to the limitations of the sample sizes; however, studies that address the performance of either very large groups of homes, or focus on a specific type (for example a single development or construction size and type) are expected to similarly reduce the uncertainty on the mean of changes in ΔHTC, increasing the ability of SMETER methods to determine the performance of retrofit.

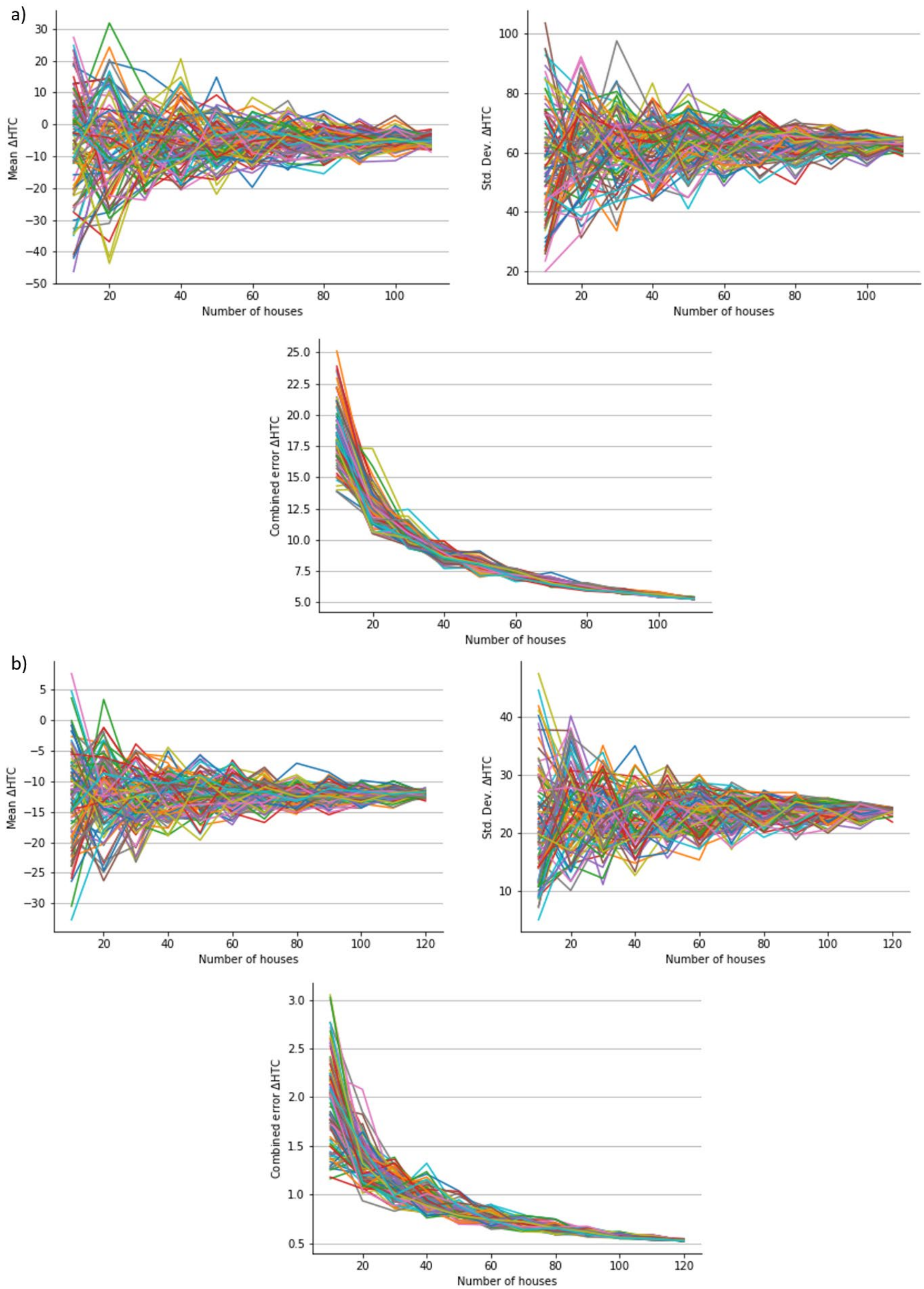


Figure 31 The repeated resampling results for loft insulation retrofit PTG results (a) and EDF-Deconstruct+ results (b). All units are WK^{-1}.

Discussion

The GHG-SMETER project applied Smart Meter Enabled Thermal Efficiency Rating (SMETER) methods to investigate the performance of properties retrofitted through the Green Homes Grant (GHG) schemes and to evaluate the ability of SMETER methods to characterise these homes and retrofits. The project analysed data from two cohorts of homes: the GHG Vouchers Scheme (GHG-V) and the Local Authority Delivery Scheme (GHG-LAD). This discussion brings together the key findings of the research, its implications and the future work that would support further development of SMETERs for the evaluation of retrofit performance.

Research design

The GHG-SMETER project was conceived after announcement of the GHG, which provided an opportunity to deploy SMETERs to investigate the impact of retrofit through the GHG and to explore how SMETERs perform in determining the thermal properties of homes that qualify for funding under such schemes. Aligning SMETER research to wider programmes of decarbonising and improving the quality of the housing stock (whether government, local government, estate management or finance led) leverages existing investment to provide cost-effective insights into both the efficacy of these schemes and the methods used to evaluate them. Such research can accelerate the development and wider implementation of tools and methods, such as SMETERs, at low cost. However, it also presents challenges that can limit the outcomes of the research.

The first challenge is that priority is naturally given to delivering the planned scheme rather than to supplementary research outcomes. It is important that research aiming to leverage wider investment does not impede the successful outcomes of that investment. A consequence of this issue can be that researchers struggle to access data, people or sites in a timely manner to facilitate their research. In the case of GHG-SMETER an example of this issue was the inability of the team to access GHG-LAD homes in advance of retrofit, to align internal temperature measurement to planned works. This was caused by the need to deliver the GHG-LAD scheme on a tight timescale which was challenging for the delivery agents to achieve; they did not have a planned programme of works long enough to be compatible with the timescales required to recruit and then measure homes in advance of retrofit. Whilst impossible to entirely solve the challenge of time pressures on planned works, some mitigation may be possible with the use of pre-agreed contracting and consent agreements to enable access to data for the Public Good (GDPR). To be effective, these agreements require development with researchers to ensure that they meet a reasonable range of research requirements. Similarly, contractual arrangements can be used to incentivise engagement with research as required.

Timely access to the often rich datasets held by different stakeholders can support energy and buildings research. In this project, access was granted to EPC input data after considerable negotiation and development of data sharing agreements. However, this is not the default and it is not necessarily an efficient way to grant data access. The introduction of standard Data Sharing Agreements along agreed principles for access for research undertaken in the public good, particularly when funded by UK government, would enable the high value of UK datasets to be better realised (such as EPC input data, MPC installation data, RHI data, Trustmark data etc).

Finally, data quality limits the value of data. In this case, data recorded through the GHG Scheme records insufficient to determine specifically what retrofit was undertaken (element area, product performance etc). The collation of such data is a challenge, in this example it required retrofit assessors and installers to complete documentation to a high standard. Work to improve standardised processes, design of data collection methods and incentives to complete high quality documentation will support both policy evaluation and future research utilising this data.

The reliability of SMETERs

The reliability of SMETERs to estimate the thermal performance, taken to be the heat transfer coefficient (HTC) of a home has different dimensions. These include the ability of the SMETER to return a result that may be taken to represent the thermal performance of the home, the repeatability of HTC estimates and the accuracy of the HTC. These issues are discussed here drawing on insights from the research.

As this project was undertaken on real occupied homes that participated in the GHG scheme, it was not possible, or economically feasible, to subject all participant dwellings to controlled tests of thermal performance, such as the aggregate heat loss or co-heating test (British Standards Institute, 2024; Wingfield, Johnston, Miles-Shenton, & Bell, 2010). This reflects the wider challenge of undertaking “ground-truth” testing of occupied homes – it is expensive and prohibitively disruptive to incorporate into most schemes given that not only must such tests be undertaken by highly-trained professionals, they must also be undertaken on unoccupied homes. It is therefore necessary to seek alternative routes, potentially including synthetic datasets, case studies and cross-validation studies, to determine the efficacy of SMETERs to estimate a home’s HTC. This project did not seek to determine what a mechanism to determine the validity of HTC results from SMETERs could be (which is addressed elsewhere by DESNZ (Department for Energy Security & Net Zero, 2024; National Physical Laboratory, 2024) and by the international scientific community (IEA EBC, 2024). However, it did investigate measures of internal robustness to determine whether the results of Type A and Type B SMETERs are plausible or not: whether they represent a possible HTC with sufficient goodness of model fit.

All SMETERs require internal robustness checks to determine whether the results are sufficiently robust (plausible) to be reported or not. The work that underpins this research highlights the value of access to large datasets, ideally with cross-validation or tested performance, in developing appropriate plausibility measures for SMETERs. The results show that plausibility rates are specific to the conditions, data availability and property characteristics, with occupant behaviours and heating control methods also expected to be major contributors, but untestable in this project. This highlights that SMETER developers will benefit from access to high quality datasets, enabling better SMETERs to be developed. It also highlights that SMETERs should be approved for different applications with care, considering for example that certain built forms (e.g. low energy use homes) may be poorly characterised by them at present.

Plausibility of HTC results and sample characteristics

The relationship between the plausibility rates (the proportion of SMETER results that pass internal tests for statistical robustness and adherence to the applied model) and various sample characteristics was investigated for the UCL-PTG and MLR/Siviour methods. A wide range of property characteristics were investigated in addition to external temperature, energy demand and dataset length. The small GHG-LAD sample limited the statistical power of analysis on this data; this discussion focuses on results from the GHG-V sample with the UCL-PTG method. Whilst these results are specific to the applied methods for this data, they may provide wider insights into the factors that affect SMETER reliability.

As expected, the higher the heating energy use of a property, the higher the chance of a plausible HTC result being returned: the metered energy use is higher providing a stronger “signal” in the “noise” of changing occupant practices and variable unmetered heat gains to the property. Plausibility rates therefore are positively correlated to floor area, negatively to EPC band, and increase with property age. The number of party walls (represented by the house type, that also includes co-variance of other factors such as property floor area) may also be linked to plausibility ratings, matching expectations as reducing the number of party elements reduces the potential for unmetered heat flows between properties. These factors all inform the property types for which SMETERs may be expected to most reliably characterise thermal performance.

These findings inform the potential application of SMETERs to support policy objectives. SMETERs are most likely to perform well on low-efficiency homes, such as those that would particularly benefit from retrofit and may be applied to determine the outcomes of policies or practice. Applications include to apprise householders about the thermal performance of their home and inform their decision-making, and to support government retrofit schemes such as ECO. By contrast, this evidence suggests that SMETERs are less likely to be robust when determining the performance of high-efficiency homes; however, due to the recruitment of homes into this research from the GHG scheme, a retrofit scheme, more research is required into SMETER application to high-efficiency homes to draw clear conclusions.

The UCL-PTG SMETER also returned a higher plausibility rate as the number of occupants increased, up to four and decreased thereafter, and the plausibility rates are also higher for more financially comfortable households. The application of SMETERs to GHG-LAD data further reinforced that underheating (potentially associated with fuel poverty) led to challenges in SMETER analysis due to the weak link between energy demand and external temperature. Application of SMETER methods to schemes aiming to address fuel poverty, such as ECO, may consequently be affected by this issue, and filtering or plausibility checks for SMETER methods in underheated homes may beneficial.

Finally, whilst no link between dataset length and the performance of Type B SMETERs was determined, longer datasets (up to a year) improved the performance of the Type A UCL-PTG. This highlights that for the purpose of policy evaluation, it is important to recruit participants into any evaluation study as soon as possible to maximise pre- intervention data (12 months historic data from the date of consent is available via the DCC). Importantly, short-timescale policy evaluation isn't possible with this SMETER method (UCL-PTG), which greatly benefits from 9-12 months of data availability.

Home technologies supplying unmetered energy

The greater the unmetered energy input into a home compared to the metered energy input, the greater the challenge for a SMETER to accurately determine the home's thermal performance. Homes with unmetered primary or secondary heating were statistically significantly less likely to return a plausible in-use HTC than those with fully metered heating for UCL-PTG. This is not surprising and reinforces the importance of checking that home heating is metered before application of a SMETER, although the potential impact of secondary heating was small in this sample.

In contrast to space heating provision, no statistical significance was observed for the presence of solar PV or solar thermal panels on the plausibility rates for UCL-PTG. A small number of homes with PV was included in the sample, however the results could be taken to indicate that SMETERs that take account of solar gains may also account for the heat input due to the self-consumption of generated solar energy, potentially widening the applicability of SMETERs.

Comparison of HTCs between SMETERs

Acknowledging the lack of “ground truth” HTC, in-use HTC SMETER estimates were compared to each other and the EPC-data Model for the GHG-V sample. Both UCL-PTG and EDF-Deconstruct+ reported in-use HTCs with skew and mean lower than predicted by the model of homes with EPCs (SMETERs determine a better than expected thermal performance, which may be expected to lead to lower than expected energy demand). This may support other research that has highlighted a performance gap between EPC energy use prediction and measured energy use, which reported a widening gap as EPC band decreased whereby EPC energy demand estimates were higher than measured energy use (Few, et al., 2023). Overestimation of energy use has important policy (and practical) implications because the savings associated with retrofit or other measures will be reduced if total demand is lower than assumed. If, however, the discrepancy between the EPC and SMETER methods is caused by a systematic bias of the in-use HTC this not only has impact for those considering retrofit and for policies to support it, but also in the potential for SMETERs to inform policymaking at stock level and support the achievement of the UK's decarbonisation targets.

The impact of GHG retrofit on in-use HTC

The GHG-SMETER project investigated the change in in-use HTC (ΔHTC) following the installation of fabric retrofit measures through the GHG-V scheme using different Type A

SMETERs. ΔHTC estimated by UCL-PTG, EDF-Deconstruct+ (both Type A SMETERs), and an EPC-data Model were compared; with a control study of a sample of over 1000 GB homes indicating no significant shift in the characteristics of the national stock over the same period. Significant data quality issues for GHG scheme installation details lead to large uncertainty and potential over-estimation of ΔHTC using the EPC-data Model. It is not possible to determine the contribution of different effects to differences between the EPC-data Model.

The repeatability and natural variability of in-use HTCs is critical to this application of SMETERs. Not only does real HTC vary over time, which may need to be accounted for in detailed analysis, but estimates of HTC vary over time within any home due to normal behaviours such as variation in window or internal door opening, or increasing the heated area of a property. Further, such changes may be caused by retrofit, for example by occupants taking-back savings to provide better comfort across their whole property. Such outcomes are not necessarily undesirable, but highlight the challenge of evaluating the performance of retrofit. Despite these issues, some insights were drawn from the GHG-SMETER study.

Firstly, the EPC-data Model predicted much larger reductions in HTC due to retrofit than the SMETER methods found in empirical data. This may be caused by the disparity in modelling, but a major contributor in this analysis is also the significantly lower in-use HTCs for participant households compared to those expected from the EPC-data Model, as discussed above. This difference potentially aligns to a performance gap between the expected and actual thermal performance of the homes prior to retrofit (and may support previous research showing a performance gap between EPC predicted energy use and observed energy use (Few, et al., 2023)). This finding, regardless of the cause, highlights the importance of understanding the starting state before any retrofit before determining any potential savings from it and emphasises the potential for SMETERs to support retrofit decision-making and policy.

EDF-Deconstruct+ and UCL-PTG showed good agreement with each other in estimating the in-use HTC and ΔHTC due to retrofit, with the former reporting lower uncertainty. Whilst this does not validate the results of these methods, this finding suggests that SMETERs may be able to provide self-consistent insights that are broadly transferable across different Type A methods, supporting their application for decision-making at a stock and policy level.

Both EDF-Deconstruct+ and UCL-PTG detected statistically significant changes in HTC for cavity wall and external wall insulation, but not for loft and pitched roof insulation, which had smaller than expected impact on HTC. This demonstrates the potential for Type A SMETERs to identify the change in thermal performance of a stock of homes (noting that the magnitude of the change cannot be verified in this work). However, the accuracy of SMETER methods provides a limitation to this use-case and for certain applications, such as determining whether a retrofit meets thermal performance expectations where the detection of small changes in HTC may be required.

In light of the potential challenges in applying SMETERs to individual homes, the potential to use Type A SMETERs to investigate the performance of a stock of homes was investigated by considering the variance of the mean of ΔHTC by repeated random sub-sampling from the total sample of each retrofit intervention studied. This method showed significant promise in determining the average impact of retrofit to high accuracy, with the standard deviation on the mean determined by the number of homes in each sample. The acceptable uncertainty in the mean depends on the application but this exploratory study suggests it may be reduced to ~6WK^{-1} at 100 and ~4WK^{-1} at 200 for PTG; and approximately 1WK^{-1} at a sample of 50 households, reducing to ~0.64WK^{-1} at 100 and ~0.5WK^{-1} at 200 for EDF-Deconstruct+. Filtering to select homes of similar characteristics was shown to decrease this uncertainty further. Whilst more research is determined into the accuracy of these Type A SMETERs and their determination of ΔHTC, this result is promising for their application to support decision-making at stock level or policymaking/policy evaluation.

References

Allinson, D., Roberts, B., Lomas, K., Loveday, D., Gorse, C., Hardy, A., . . . Revell, L. (2022). *Technical Evaluation of SMETER Technologies (TEST) project: full report.*

Bennett, G., Elwell, C., & Oreszczyn, T. (2019). Space heating operation of combination boilers in the UK: The case for addressing real-world boiler performance. *Building Services Engineering Research and Technology, 40*(1), 75-92.

Bennett, G., Elwell, C., Lowe, R., & Oreszczyn, T. (2016). The importance of Heating System Transient Respone in Domestic Energy Labelling. *Buildings, 6*(3), 20.

BRE. (2014). *SAP 2012 The Government's Standard Assessment Procedure for Energy Rating of Dwellings.* BRE.

British Standards Institute. (2017). BS EN ISO 13789:2017: Thermal Performance of Buildings. Transmission and Ventilation Heat Transfer Coefficients. Calculation Method.

British Standards Institute. (2024). BS EN 178871:2024: Thermal Performance of buildings. In situ testing of completed buildings. Data collection for aggregated heat loss test.

Chambers, J., & Oreszczyn, T. (2019). Deconstruct: A scalable method of as-built heat power loss coefficient inference for UK dwellings using smart meter data. *Energy & Buildings, 183*, 443-453.

Department for Business, Energy & Industrial Strategy. (2020). *Green Homes Grant: Local Authority Delivery Phase 1B Guidance for Local Authorities.*

Department for Business, Energy & Industrial Strategy. (2021). *Energy Follow Up Survey: thermal comfort, damp and ventilation.*

Department for Business, Energy & Industrial Strategy. (2021). *Heating patterns and occupancy report.*

Department for Energy Security & Net Zero. (2023). *PAS 2035: 2023 Retrofitting dwellings for improved energy efficiency - Specification and guidance.*

Department for Energy Security & Net Zero. (2024). *Domestic hot water use Observations on hot-water use from connected devices.*

Department for Energy Security & Net Zero. (2024). *Mini Competition against an existing Framework Agreement: Provision of Building Energy Performance Specialist to the Smart Metering Implementation Programme.*

Department for Levelling Up, Housing & Communities. (2021). *English Housing Survey Headline Report 2020-21.*

Department for Transport. (2021). *Decarbonising transport: a better, greener Britain.*

Deurinck, M., Saelens, D., & Roels, S. (2012). Assessment of the physical part of the temperature takeback for residential retrofits. *Energy & Buildings, 52*, 112-121.

Farmer, D., Johnston, D., & Miles-Shenton, D.). Obtaining the heat loss coefficient of a dwelling using its heating system (integrated coheating). *Energy & Buildings, 117*, 1-10.

Fels, M. (1986). PRISM : An Introduction. *Energy and Buildings, 9*, 5-18.

Few, J., Manouseli, D., McKenna, E., Pullinger, M., Zapata-Webborn, E., Elam, S., . . . Oreszczyn, T. (2023). The over-prediction of energy use by EPCs in Great Britain: A comparison of EPC-modelled and metered primary energy use intensity. *Energy & Buildings, 288*.

Gori, V., & Elwell, C. (2018). Estimation of thermophysical properties from in-situ measurements in all seasons Quantifying and reducing errors using dynamic grey-box methods. *Energy & Buildings, 167*, 290-300.

Gori, V., D, J., Bouchie, R., & Stamp, S. (2023). Characterisation and analysis of uncertainties in building heat transfer estimates from co-heating tests. *Energy & Buildings*, 113265.

Griffiths, E., Greci, C., Kotrotsios, Y., Parker, S., Scott, J., Welpton, R., . . . Woods, C. (2019). *Handbook on Statistical Disclosure Control for Outputs.*

Hayton, J. (2009). *Technical paper supporting SAP 2009 A meta-analysis of boiler test efficiencies to comapre independent and manufacturers' results.* BRE.

HM Government. (2021). *The Building Regulations Approved Document L Volume 1.*

Hollick, F., Gori, V., & Elwell, C. (2020). Thermal performance of occupied homes : A dynamic grey-box method accounting for solar gains. *Energy & Buildings, 208*, 109669.

Huebner, G. M., McMichael, M., Shipworth, D., Shipworth, M., Durand-Daubin, M., & & Summerfield, A. (2013). Heating patterns in English homes: Comparing results from a national survey against common model assumptions. *Building and Environment, 70*, 298-305.

IEA EBC. (2021). *Building Energy performance Assessment Based on In-situ Measurements.*

IEA EBC. (2024, July 21). *Annex 94 - Validtion and Verification of In-situ Building Energy Performance Measurement Techniques*. Retrieved from https://annex94.iea-ebc.org/

Jack, R. (2015). *Building Diagnostics: Practical Measurement of the Fabric Thermal Performance of Houses.* Phd Thesis.

Johnston, D., Sidall, M., Ottinger, O., Peper, S., & Feist, W. (2020). Are the energy savings of the passive house standard reliable? A review of the as-built thermal and space heating performance of passive house dwellings from 1990 to 2018. *Energy Efficiency, 13*, 1605-1631.

Joint Committee for Guides in Metrology. (2012). *International Vocabulary of Metrology - Basic and General Concepts and Associated Terms, Technical Report.*

Li, F., Smith, A., Biddulph, P., Hamilton, I., Lowe, R., Mavrogianni, A., . . . Oreszczyn, T. (2014). Solid-wall U-values: heat flux measurements compared with standard assumptions. *Building Research & Information, 43*(2), 238-252.

Lowe, R., Summerfield, A., Oikonomou, E., Love, J., Biddulph, P., Gleeson, C., . . . Wingfield, J. (2017). *Final report on analysis of heat pump data from the Renewable Heat Premium Payment (RHPP) scheme.*

Lowe, R., Wingfield, J., Bell, M., & Bell, J. (2007). Evidence for heat losses via party wall cavities in masonry construction. *Building Services Engineering Research and Technology, 28*(2), 1-19.

National Physical Laboratory. (2024). *SMETER Validation Methodology (Unpublished).*

Orr, G., & Summerfield, I. (2009). *Final Report: In-situ monitoring of efficiencies of condensing boilers and use of secondary heating.*

Porritt, S. M. (2012). *Adapting UK dwellings for heat waves.* PhD thesis.

Ritosa, K., Saelens, D., & Roels, S. (2023). A large scale artificial dataset to evaluate the impact of on-board monitoring set-ups on the Heat Tranfers Coefficient assessment. *Energy & Buildings, 289*, 113061.

SERL. (2024, July 21). *Smart Energy Research Lab (SERL)*. Retrieved from https://serl.ac.uk/

Sharma, M. (2024, July 21). *How to Assess a Regression's Predictive Power for Energy Use*. Retrieved from kW Engineering: https://kw-engineering.com/how-to-assess-a-regressions-predictive-power-energy-use/

Stamp, S. (2015). *Assessing uncertainty in co-heating tests : Calibrating a whole building steady state heat loss measurement method.* PhD Thesis.

Summerfield, A. J., Lowe, R. J., & Oreszczyn, T. (2010). Two models for benchmarking UK domestic delivered energy. *Building Research & Information, 38*(1), 12-24.

Watson, S., Lomas, K., & Buswell, R. (2021). How will heat pumps alter national half-hourly heat demands? Empirical modelling based on GB field trials. *Energy & Buildings*, 110777.

Webborn, E., Few, J., McKenna, E., Elam, S., Pullinger, M., Anderson, B., . . . Oreszczyn, T. (2021). The serl observatory dataset: Longitudinal smart meter electricity and gas data, survey, epc and climate data for over 13,000 households in great britain. *Energies, 14*(21).

Wingfield, J., Bell, M., Miles-Shenton, D., South, T., & Lowe, B. (2009). *Lessons from Stamford Brook. Understanding the Gap between Designed and Real Performance.*

Wingfield, J., Johnston, D., Miles-Shenton, D., & Bell, M. (2010). *Whole House Heat Loss Test Method (Co-heating).* Leeds Beckett University.

Appendix

For the full report submitted by EDF:

Medina Vazquez, Gustavo; Thevapalan, Anushiya; Patel, Nium (2025). Estimating the effect of retrofitting with Deconstruct+. University College London.
Report. https://doi.org/10.5522/04/29314244.v1

https://rdr.ucl.ac.uk/articles/report/Estimating_the_effect_of_retrofitting_with_Deconstruct_/29314244/1?file=567353

This publication is available from: https://www.gov.uk/government/collections/heat-pump-research