RetroMeter WP2 – Methodology Development

April 24

Document control

ESC programme name	**SIF Application Alpha - RetroMeter**
ESC project number	ESC00890
Version*	1.0
Status	Final
Restrictions*	Confidential
Release date	27/03/2024
External release ID	(if project release process requires it)

* Refer to the Information Classification Policy

Review and approval

	Name	Position
Author	Connor Galbraith Paula Bendiek	Data Scientist Data Scientist
Reviewer(s)	Samuel Young Richard Dobson	Practice Manager, AI & Data Science Business Lead, Digital
Approver	Rebecca Sweeney	Business Leader, Homes

Revision history

Date	Version	Comments
15.03.2024	0.1	Initial revision
22.03.2024	0.2	Reviewed
27.03.2024	1.0	Approved

Contents

1. Executive Summary

This report presents the findings from the Alpha phase of the SIF RetroMeter project, which aims to develop a robust methodology for quantifying the energy savings from heating-related retrofits (like insulation and improved heating systems) in UK homes. Accurately measuring these savings, known as Metered Energy Savings (MES), is crucial for building trust in retrofits, enabling new financing options, and supporting grid planning. However, it is challenging to measure savings in a way that accounts for variations in weather, energy prices, and occupant behaviour. The project focused on the most common retrofit use case in the UK - fabric upgrades to homes with gas heating pre-intervention. It evaluated three core methodologies:

- OpenEEmeter, an open-source implementation of the CalTRACK method, which uses weather data and statistical models to predict a home's energy use in the absence of a retrofit, against which their actual metered consumption can be compared.
- Comparator groups, where the predictions are adjusted based on the consumption patterns of similar homes that did not receive a retrofit, to account for non-weather externalities (like energy prices) that OpenEEmeter cannot capture.
- Physics-based modelling, to estimate the portion of savings that may be taken back as increased comfort rather than bill savings. This part of the methodology in particular incorporates the work by the Smart Meter Enabled Thermal Efficiency Ratings (SMETER) project.

The first two methodologies were tested on smart meter data from approximately 3,000 homes, with the physics-based methodology tested on 15 homes from the SMETER project. Key findings were:

- OpenEEmeter alone was not sufficiently accurate, largely due to its inability to account for the sharp increase in energy prices experienced in the winter of 2022/23, and the resulting behaviour changes during the energy crisis. It systematically overestimated consumption in the post-retrofit period.
- Using comparator groups eliminated OpenEEMeter's systematic over-prediction and greatly improved accuracy, especially when comparators were matched based on the similarity of their energy consumption profiles. Grouping candidate homes into portfolios of as few as 5 homes brought accuracy within industry guidelines. Further collaboration with the Smart Energy Research Lab (SERL) and Hildebrand was explored for maintaining these comparison groups over intervention periods.
- The physics-based approach shows some promise for estimating comfort take-back when aggregated to larger portfolios, though further validation is needed. A better understanding of how the availability of smart meter and internal temperature data affects HTC and energy demand estimation performance was attained. The approach developed in this Alpha phase project will complement and help springboard further work alongside the conclusions of SMETER to UK-centric efficiency modelling.

The work has made a significant step forward by demonstrating that the RetroMeter methodologies provide a suitably robust foundation for assessing energy savings from small groups of homes in the UK. Focus now needs to shift to creating the conditions for large scale adoption, which include:

- Establishing a mechanism for ongoing access to smart meter data for comparison groups.
- Testing the effectiveness of comfort takeback estimates on a larger group of homes.
- Turning the methodology into an open standard and open-source software tool that are easy to apply.
- Creating regulatory incentives to encourage the use of MES in publicly funded retrofit programmes.

2. Approaches to Measuring Metered Energy Savings

This section introduces the concept of Metered Energy Savings (MES), and ties in the conclusions of the Discovery report that were taken forward in this Alpha phase part of the RetroMeter project. Section 2.4 will briefly overview the methodology discussed in the Discovery phase report and highlight any changes that have been made during Alpha.

2.1 What are Metered Energy Savings?

As the UK continues to progress towards reducing its carbon emissions and increasing domestic energy efficiency, it is becoming more and more important for homeowners to upgrade their properties with retrofits including improved insulation, more efficient heating systems, and solar panels. However, it can be challenging to accurately measure how much energy and money these upgrades actually save the homeowner, once you factor in variations in weather, energy prices, and behavioural changes. This is where the concept of MES comes in.

MES refers to the process of calculating the financial, carbon, and energy savings attributable to a heating-related retrofit in a home, resulting from energy efficiency upgrades or retrofits to the property. Rather than relying on estimates or averages, MES aims to calculate the actual energy saved by comparing the home's energy usage before and after the upgrades, controlling for other variables that may distort the true impact of the retrofit. There are several key reasons why accurately measuring MES is valuable both for the occupier and the wider energy system:

- MES allows households to verify that the upgrades they paid for are delivering the promised savings on their energy bills. This builds trust and confidence in retrofits as a mechanism to lower bills.
- Quantifying the savings makes it easier to invest in upgrades through financing options where the upfront costs are paid back over time from the monthly energy savings.
- Utilities and grid operators can use MES data to understand the impacts of efficiency programs and factor them into their demand forecasting and grid planning.
- Proven energy savings create a financial value that can potentially be monetised through energy efficiency incentive programs or carbon credit markets.

At its core, accurately measuring MES for a property requires creating a 'counterfactual' – a projection of how much energy the home would have used without had the retrofit not occurred, accounting for all the main factors that might influence the home's energy consumption. The methods for generating these counterfactual predictions need to be robust to the availability of data from the property, particularly in cases where the proportion of missing data is high. In addition, they must also try to account for changes in occupant behaviour such as 'comfort take-back', where occupiers set the internal temperatures higher after insulation upgrades in response to their newly increased heat retention.

2.2 Use Cases

The Discovery report introduced several use cases for MES, including important but currently niche situations where full gas to electric heating conversions take place alongside fabric retrofits with Time of Use (ToU) tariffs. It was concluded that the Alpha phase would focus on modelling gas-heated homes rather than electric, specifically targeting use case 1 (gas heated home, fabric retrofit only). This is primarily due to:

- Prevalence: The majority of UK homes are currently gas heated, making this the most common and impactful use case to address initially and establish the RetroMeter methodology across the largest pool of possible users.
- Simplicity: Gas prices are not dependent on time of use, eliminating the need to develop a more complex and less reliable half-hourly model. Daily or monthly models are sufficient for gas, which simplifies the methodology development and improves accuracy.
- Data availability: Requiring internal temperature data for a year pre-retrofit poses a significant barrier to widespread adoption. By focusing on gas and utilising smart meter data (available for around half of homes in the UK), the methodology can be more easily implemented and scaled.

Concentrating on this use case allows for the development of a broadly applicable methodology that is effective for the most common heating configuration in the UK. This approach simplifies the initial development process while still enabling the methodology to be extended to other use cases in the future as needed. The goal is to create a strong, verified methodology that can be built upon and adapted rather than attempting to devise a one-size-fits-all solution from the outset.

2.3 Existing Approaches and Limitations

A comprehensive overview of the MES landscape can be found in Metered Energy Savings 2022[1] and the Discovery phase report. A summary is provided here for information.

Historically, expected energy savings from building upgrades have been calculated using physics-based modelling approaches like RdSAP in the UK, implemented by building engineers at the point of retrofit install. This single static prediction however is not verified post-install using actual energy consumption readings, limiting its use as a rolling MES measure. MES approaches based on smart meter data aim to compare actual metered energy use before and after an intervention to quantify the savings, assuming other variables like occupancy and behaviour remain unchanged. These include:

- CalTRACK – Developed primarily for portfolios of commercial buildings, it uses a series of interpretable linear model based on heating/cooling degree days and time of week, requiring 12 months of pre-intervention consumption and external temperature data at daily, hourly, or monthly granularity. The open-source implementation of the CalTRACK standards, OpenEEmeter, is the core

[1] Young *et al.* (2022)

counterfactual model used in this project. The methodology was developed in the US by Recurve, and is currently maintained by the LF Energy.

- SENSEI – Similar to CalTRACK, aims to identify more distinct building operating states to better model occupancy variations, but requires hourly energy usage, external temperature, and other manually defined features including occupancy. It uses somewhat more complex gradient-boosted decision trees, offering less interpretability than CalTRACK. SENSEI was tested alongside OpenEEmeter in Metered Energy Savings 2022 and found to offer limited upside for significantly more technical difficulty.
- The Great Energy Predictor III competition on Kaggle also compared various modelling approaches. Top performers used large ensembles of models, especially gradient boosted trees. Careful data pre-processing to handle missing data and anomalies was also found to be critical – while not a significant barrier to deployment, the lack of standardisation in domestic energy data means that this is a mostly manual process.

2.4 The RetroMeter Approach

A basic approach to metered energy savings might be to simply look at a household’s pre-retrofit energy use and compare it to a household’s post-retrofit energy use. However, this does not account for the fact that the winter after the retrofit might be colder or warmer than the winter before it.

2.4.1 Adjusting for Weather

OpenEEmeter accounts for the impact of weather on energy consumption using mean hourly external temperature and metered energy consumption in the pre-retrofit ‘baseline’ period to fit regression models that also account for seasonal and other calendar effects. RetroMeter employs the ‘daily’ version of OpenEEmeter, generating a counterfactual each day for what the energy use would have been given the weather conditions. More detail on this approach is given in Section 4.1.

2.4.2 Adjusting for Society-Wide Factors

OpenEEmeter alone is incapable of accounting for society-wide factors that can influence energy use, for example: energy price changes causing people to cut back on their consumption, or changes to home heating practices during COVID lockdowns. These make simple before-and-after comparisons based on weather alone inaccurate for measuring MES.

The comparator methodology builds further on OpenEEmeter by comparing the energy use in the ‘candidate’ household post-retrofit, to energy use in the same period for similar households which have not had a retrofit. This can help separate out the energy changes due to retrofit from the energy changes happening in society more broadly. How these ‘comparator’ households are matched to the candidate is a key driver of how effective the modelling approach is, and this project investigated several approaches based on the availability of smart meter data prior to the retrofit.

2.4.3 Adjusting for Household Behaviour

Lastly, households may decide to keep their home at a warmer temperature after the retrofit than before the retrofit, because it cost them less to warm their home after the retrofit. One way to account for this comfort take-back is to take the observed post-retrofit internal temperatures and apply them to a physics-based model of the house pre-retrofit to calculate what the energy demand would have been.

This requires the use of a heat transfer coefficient (HTC), which is a measure of the rate at which the heat generated in a home is typically lost through leakage, with the pre-retrofit HTC estimated by correlating the pre-retrofit weather with the pre-retrofit gas usage.

This model focuses on gas usage, but requires additional data on things that influence heat generation and loss in a home. For example, it assumes that a certain proportion of electricity usage generates heat in the home indirectly, including through electric cooking and kitchen appliances, electronics, and lights. The model factors in heating from the sun, estimated using weather data including the external temperature and the solar irradiance. The model also accounts for baseload gas usage, defined as the gas used for purposes other than space heating such as domestic hot water and cooking.

2.4.4 Performance Metrics

The metrics used to assess the efficacy of the counterfactual models have been introduced in the Discovery report and are repeated here for reference. The Normalized Mean Bias Error (NMBE) and the Coefficient of Variation of the Root Mean Square Error (CVRMSE) are recommended by ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) for evaluating MES modelling.

The NMBE measures whether the model consistently over- or under-predicts energy use – in other words, if there is a systematic bias. Positive and negative errors can cancel out, so NMBE alone is not sufficient to evaluate accuracy. It is defined as:

$$NMBE = \frac{\sum(\hat{y}_i - y_i)}{n-1} / \bar{y}$$

The CVRMSE measures the size of prediction errors regardless of whether the model is over- or under-predicting. Prediction errors are squared, so larger errors are penalized more than small ones. CVRMSE normalizes the error by the mean observed energy usage:

$$CVRMSE = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-1}} \Big/ \bar{y}$$

where y_i is the observed energy use at time i, $\hat{y}_i$ is the predicted energy use, n is the number of time intervals, and $\bar{y}$ is the mean observed usage. The lower the CVRMSE, the higher the accuracy.

Figure 1 helps to visualise the difference between the two metrics as the balance between bias and variance, similar to the distinction made between accuracy and precision.

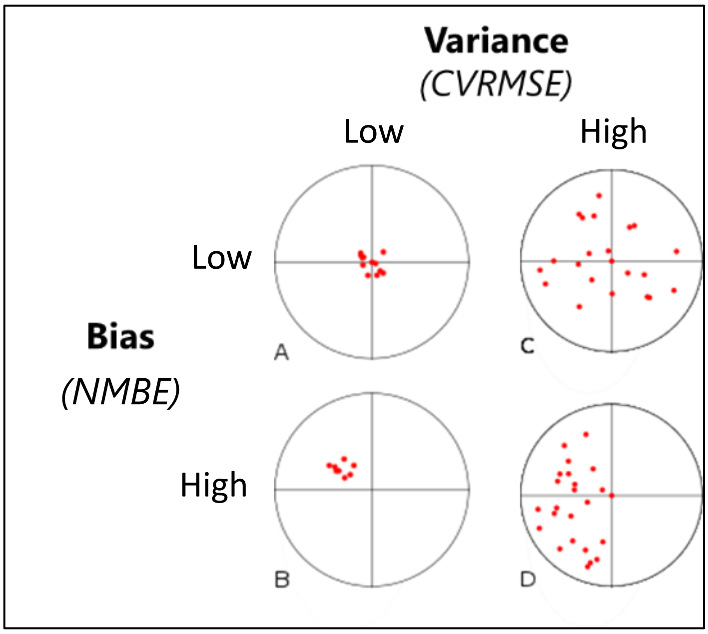

Figure 1 - Visualising what CVRMSE and NMBE are measuring - the balance between accuracy and precision.

The CVRMSE cannot be negative, but as NMBE can represent both systematic over- and under-prediction, the objective is to bring both metrics as close to zero as possible. The following sections will demonstrate that in general, a higher CVRMSE implies a higher NMBE, and vice versa.

2.5 Aims and Objectives

Work Package Two of the RetroMeter Alpha phase is focused on establishing whether an open MES methodology can be applied in a UK domestic gas-heating environment. By testing the RetroMeter approach described in this section on a large enough sample of properties, we intend to produce results with statistical significance that can then be used to inform the scaling up of the method to real-world retrofit MES projects.

We will not only be looking at the raw performance metrics, but also the feasibility of applying the RetroMeter approach to houses in an array of circumstances, including where no smart data is available for comparison groups prior to retrofit, where a heat pump installation is part of the intervention, or where no internal temperature measurements are available. We hope to lay the groundwork for a standardised, industry-wide approach to MES in the UK.

3. Smart Meter Dataset Analysis

Acquiring residential smart meter data in the UK is hard. Governance, privacy concerns, siloing, and in some cases strict sandboxing can limit the extent to which research projects such as this can reach the scale necessary to produce statistically significant results. Fortunately, the Catapult was able to engage with Hildebrand, a leading energy analytics provider, to procure a dataset of over 16,000 properties with half-hourly gas smart meter readings for this analysis. For the more detailed requirements of the physics-based methodology, the Catapult leveraged data published by the SMETER programme for highly granular data including internal temperature.

This section introduces these datasets and outlines their attributes, how they were cleaned, and a breakdown of their quality metrics.

3.1 Smart Meter Dataset Attributes

The set of gas smart meter data contains 16,759 anonymised properties, each with half-hourly gas consumption readings in kWh with varying start dates, all ending at the start of September 2023. Each property is matched to its nearest public weather station from which we acquire hourly external temperature readings in Celsius for the duration of the time that the property has valid meter readings. Across the full dataset, there are 58 unique weather stations with a fairly even distribution over England. It was evident that there was a moderate class imbalance across the stations, with several matching a disproportionate number of properties compared to others, although without an indication of how close each property is to their nearest station, it is not possible to tell from this alone whether it will impact the sampling in any meaningful way.

Each property was also provided with basic information about the building, summarised in Table 1 below. The totals for each characteristic may not sum exactly to the total number of properties in the dataset due to missing values. Given that 21.7% of UK households occupy a flat or maisonette[2], this dataset appears to under-represent flats, alongside modern properties built within the last twenty years. However, given that older houses are more likely to be the ones needing retrofits, this under-representation is less important.

[2] ONS Census 2021

Table 1 - Summary of smart meter dataset property metadata.

CHARACTERISTIC	VALUE	COUNT
Current EPC Rating	*A*	37
	B	944
	C	4582
	D	8019
	E	2656
	F	442
	G	79
Potential EPC Rating	*A*	449
	B	8141
	C	6359
	D	1487
	E	274
	F	33
	G	16
Property Type	*House*	14430
	Bungalow	1282
	Flat	887
	Maisonette	160
Built Form	*Detached*	6179
	Semi-Detached	5728
	Mid-Terrace	2936
	End-Terrace	1739
Main Heating Source	*Boiler heating, mains gas*	16,161
	Other (electric, oil, storage)	598
Property Age Band	*pre-1900*	1109
	1900-1929	2029
	1930-1949	2617
	1950-1966	2800
	1967-1975	2035
	1976-1982	1019
	1983-1990	1326
	1991-2002	1912
	2003-2006	676
	2007+	319

Another important factor is the number of contiguous days of readings available per property. The OpenEEmeter methodology requires at least two years of readings to conform with the CalTRACK standard – one year for the baseline period, and one year for the reporting period. As such, many of the properties in the dataset cannot be used for the RetroMeter methodology as they started recording their gas usage too late to be included. Since Hildebrand have an increasing number of homes in their database, more properties would be available with enough readings to be usable by OpenEEmeter if this project were to be extended in the future.

3.2 Data Quality and Cleaning Process

The CalTRACK standards[3] stipulate a minimum level of smart meter data quality for the counterfactuals to be reliable in practical applications. The most important requirements for RetroMeter are:

- All meter readings must be positive, with duplicate timestamps rendered as the average of both (or more) readings.
- The number of missing days in the baseline period should not exceed 10%, or 37 days for a typical 365-day year.
- Values of 0 are considered missing for electricity data, but not for gas.
- When aggregating granular readings up to daily, no more than 50% of values can be missing, i.e. in one day of 48 half-hourly readings, no more than 24 can be missing. These missing values can be infilled using the average consumption over the non-missing periods. For gas smart meter data, where it is normal to have many half-hour periods with zero consumption, this requirement can be stringent. We have assumed that periods of zero consumption that last longer than a month constitute missing readings.
- Hourly temperature data may not be missing for more than six consecutive periods, otherwise the day is treated as a missing day. Any fewer can be linearly interpolated.
- Extreme values should be flagged for review if greater than three interquartile ranges above the median. For this project, we have chosen to cap these using windsorisation.

After applying each of these requirements, the original dataset of 16,759 homes was reduced down to 3,048, according to the conditions outlined on the x-axis in Figure 2.

[3] https://docs.caltrack.org/en/latest/methods.html

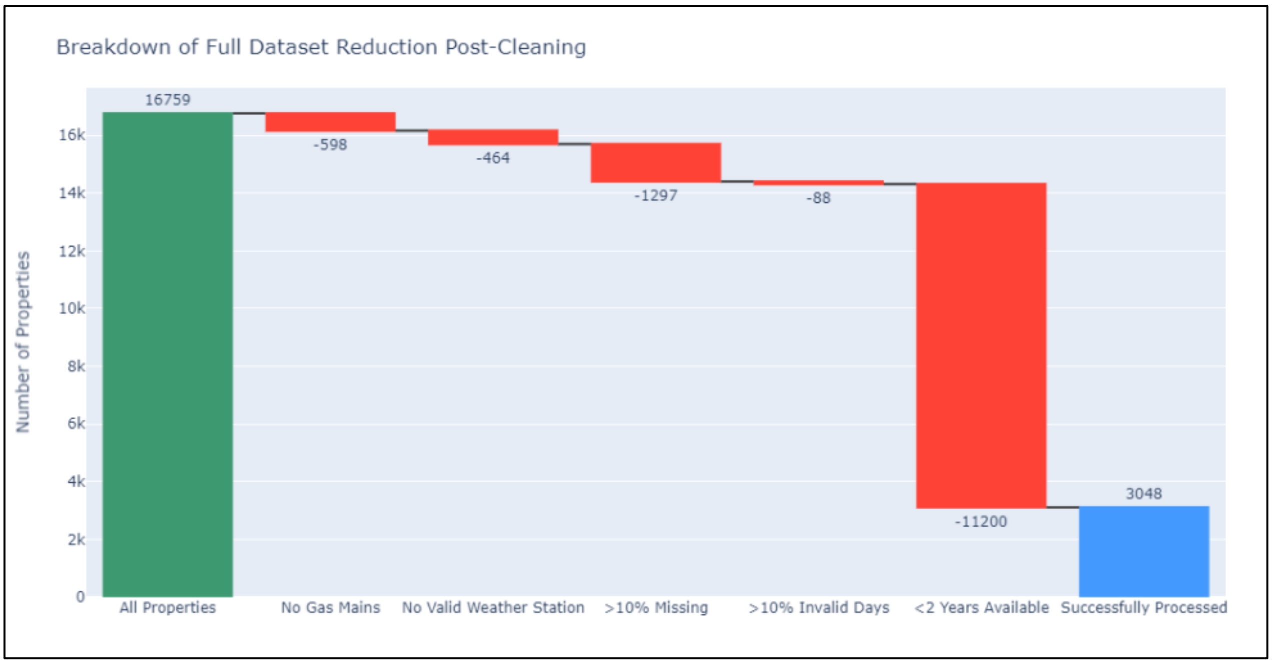

Figure 2 - Causes of property removal from the original dataset to the final number successfully processed.

Figure 3 outlines the proportion of periods affected by each of the four key CalTRACK sufficiency criteria, for the number of properties on the y-axis, applied cumulatively in the order presented in the legend. Despite the number of properties that needed to be excluded from the dataset due to having fewer than the complete two years of data necessary for a full baseline and reporting period, relatively few properties needed to be excluded on the grounds of the sufficiency criteria.

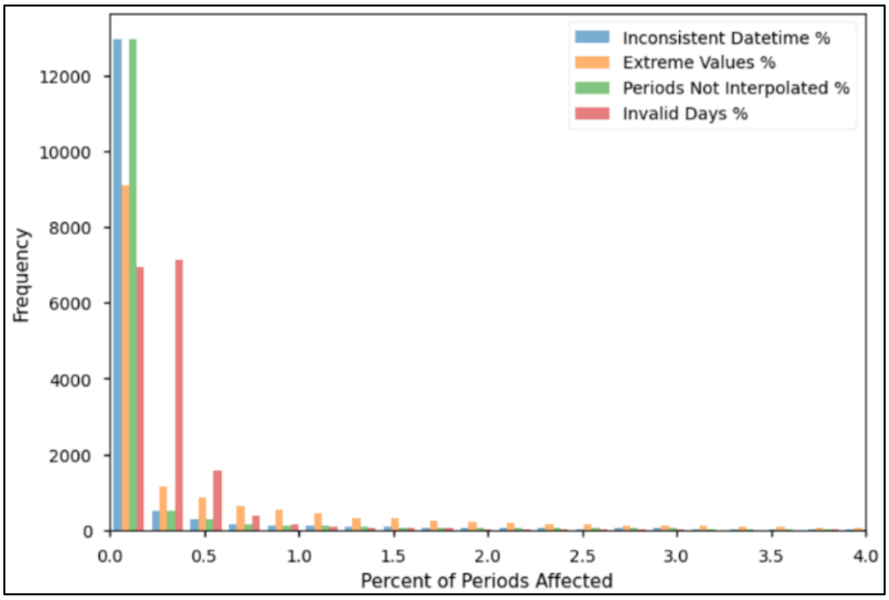

Figure 3 - Proportion of half-hourly or daily periods over all properties affected by CalTRACK data sufficiency criteria.

3.3 Property Data Used for the Physics-Based Model

To develop the physics-based model required data from homes that included internal temperature, co-heating Heat Transfer Coefficient measurements, and solar irradiance. Data from 15 homes that were part of the Smart Meter Enabled Thermal Efficiency Ratings (SMETER) project was used to develop the model (Allinson, et al., 2022).

Data from an additional 15 homes from the same project was due to be published in January 2024 and so the RetroMeter project intended to use that data to validate the physics-based model to ensure it was not over-fitted to the first 15 homes. Unfortunately delays to publication of that data meant that validation of the results has not been possible, and therefore the physics-based model may not generalise well to all homes.

4. Applying OpenEEmeter to Individual Properties

A naïve approach to creating a counterfactual energy usage post-retrofit would be to use the energy consumption profile from the year before. If it's colder, however, the household would likely have used more energy to compensate – so a more sophisticated method is needed to adjust for the range of reactions households will have to changes in temperature. OpenEEmeter is designed to tackle this challenge.

4.1 How OpenEEmeter Works

OpenEEmeter is an open-source Python package developed by Recurve, now maintained by the LF Energy, that provides an implementation of the CalTRACK MES methodology. For RetroMeter, we have used version 4.0.0a2, a pre-release available from the start of the project that benefits from many improvements to the daily model made over version 3. To preserve the comparability and transferability of this project's results, we have ensured that no material changes were made to the OpenEEmeter package and that the CalTRACK data sufficiency requirements were fully adhered to. More information on these requirements can be found in the technical annex.

The OpenEEmeter daily method relies on a set of linear models with three energy use regimes: heating, cooling, and baseload, which are fitted to the baseline year's temperature and energy consumption data. The CalTRACK methods attempt to identify external temperatures at which heating and cooling begin to be required in the home, known as heating and cooling balance point temperatures. For RetroMeter, where only gas heating is considered, there will never be any energy demand for cooling, so only the heating model and balance points are considered.

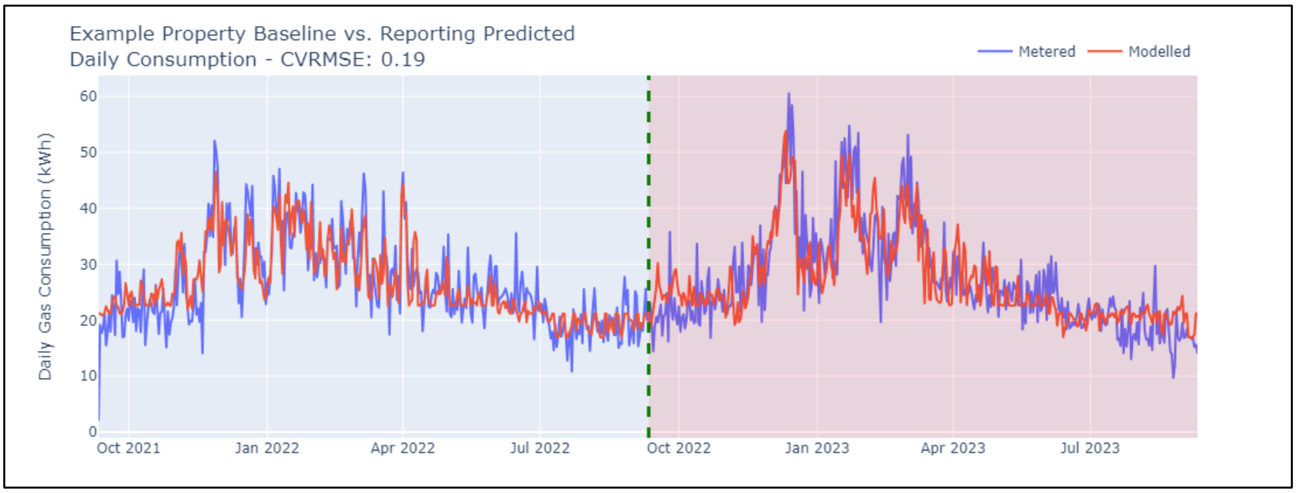

Figure 4 - OpenEEmeter fitted to a single property baseline year (in blue), with predicted consumption for the reporting period (in red)

The predicted energy demand each day is regressed on the total difference between the external temperature and the balance point temperature over the course of a day (this is referred to as heating degree days). OpenEEmeter will fit the best model from a range of candidates to each season of the year, divided into winter, summer, and 'shoulder' for spring and autumn, but does not account for comfort take-back, external factors, or changes in occupancy patterns.

For the reporting period, the fitted model will then predict the consumption given the hourly temperature readings for each day of interest. Figure 4 illustrates what this looks like for an example property that has not changed its behaviour much in the reporting period compared to the baseline period – this can be seen by how well the modelled and observed consumption values in the reporting period align, indicating that this property has a predictable response to external temperature without much influence from other factors.

A particular strength of OpenEEmeter is that it does not require a year of internal temperature readings prior to the retrofit, only external temperature readings that can be acquired easily from publicly available weather station data streams. Providing the property has had a functioning smart meter for at least a year prior to retrofit, the most recent 13 months of consumption data can be extracted to fit the baseline model, with evaluation performed on a rolling basis after the installation period is complete.

As there are no interventions performed on properties within our dataset during the testing period, the expected MES is zero. Therefore, minimising the prediction error and bias of the counterfactual is the metric against which OpenEEmeter will be assessed. Additionally, in practice a retrofit can take weeks and sometimes months the complete; this is referred to as a 'blackout period', and is typically ignored in real metered energy savings applications, with the baseline period ending at the point where works begin, and the reporting period commencing once they are finished. For the purposes of this study, the blackout period is assumed negligible – the reporting period begins as soon as the baseline period ends.

4.2 Principal Results

As described in Section 3, the properties with sufficient data for a full year of baseline and reporting period testing spanned the winters of 2021/22 and 2022/23. The winters are the most important part of the year as this is where the majority of a property's heating demand will occur. Figure 5 describes the mean temperature over the baseline and reporting winter periods, where it can be seen that during the reporting period, most months were colder than the baseline period. This suggests that, all else being equal, OpenEEmeter should predict that the average property consumes more energy during the reporting period than the baseline period.

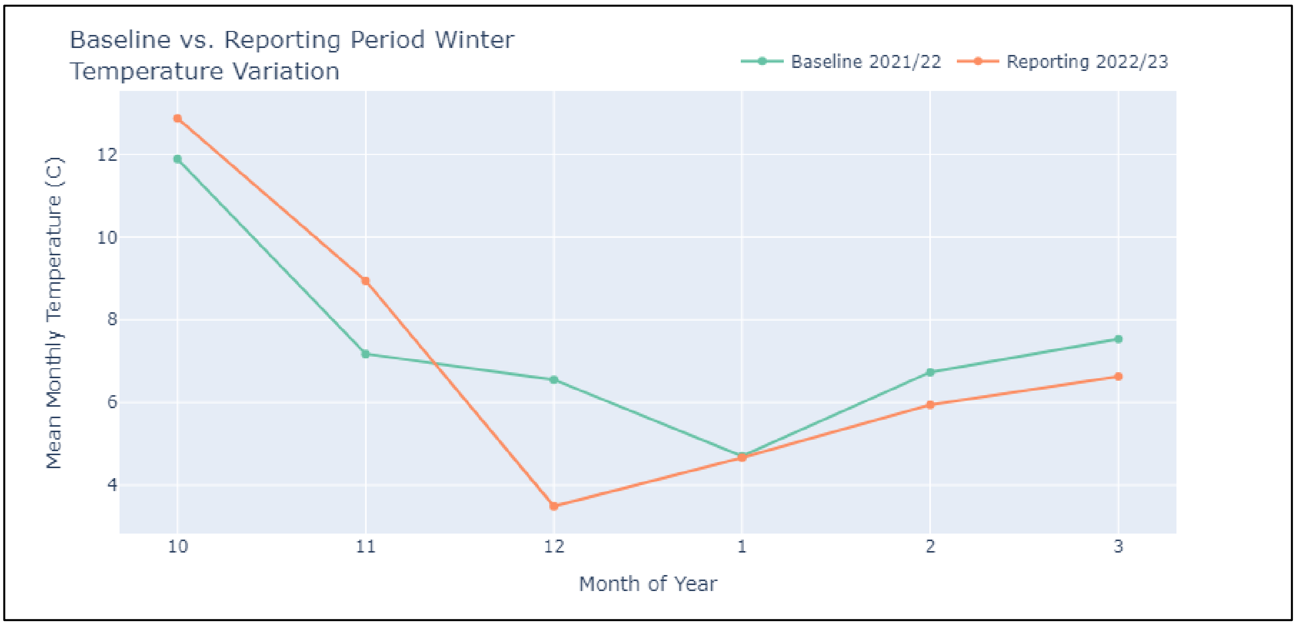

Figure 5 - Mean monthly winter temperature over the baseline and reporting period.

However, the winter of 2022/23 also saw a significant spike in the price of gas, as shown in Figure 6 using an index tracking the real-terms gas price normalised to 2010, which proved to be a stronger external factor than the lower temperatures on domestic consumption habits, resulting in lower mean consumption. We therefore expected OpenEEmeter to overestimate the reporting period consumption, as it is only capable of incorporating the external temperature and not the gas price increase.

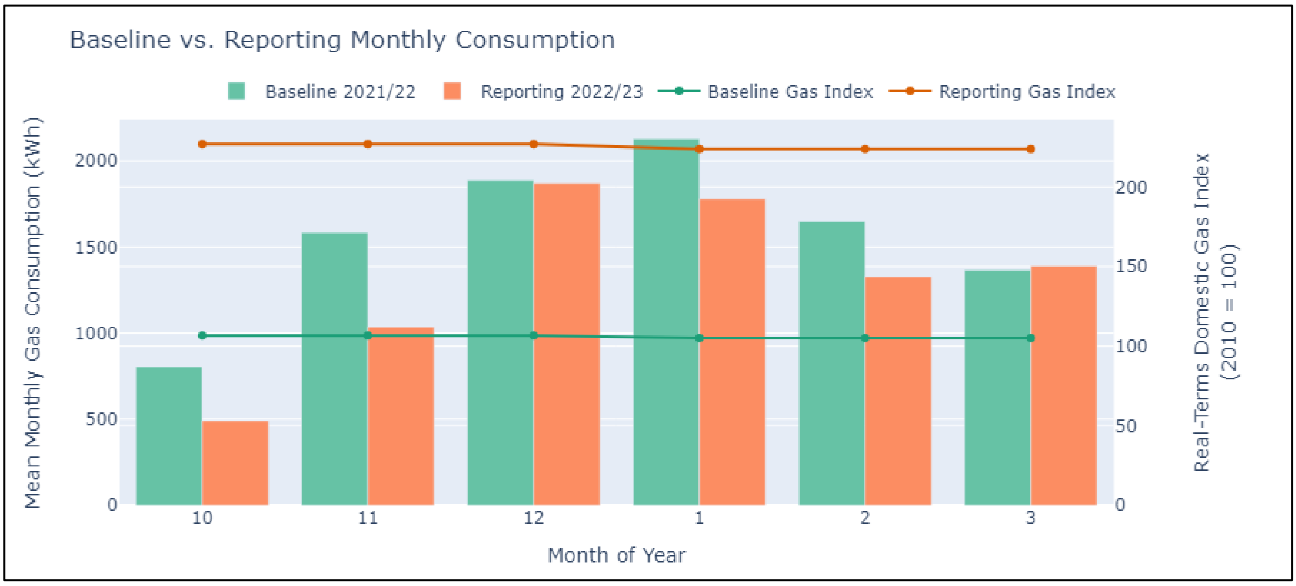

Figure 6 - Mean winter baseline/reporting period measured gas consumption and domestic gas retail price index.

Figure 7 presents the results of fitting the OpenEEmeter daily model to all 3,048 properties with at least two years of contiguous gas meter readings. Alongside, we have also presented the results from Metered Energy Savings 2022 for comparison, where version 3 of the OpenEEmeter daily model was applied to 42 homes with gas smart meter readings across a range of years prior to 2022. The blue (hourly) and green (monthly) lines indicate the ASHRAE criteria (Guideline 14-2014, Measurement of Energy and Demand Savings) for a modelled counterfactual to be sufficiently accurate that the difference observed between the counterfactual and the post-intervention actual metered values can reasonably be attributed to the intervention, rather than error in the model. These are:

"The computer model shall have an NMBE of 5% and a CV(RMSE) of 15% relative to monthly calibration data. If hourly calibration data are used, these requirements shall be 10% and 30% respectively."

In the absence of a defined acceptable daily error, we assume it should fall between the 30% and 15% figures quoted above. It is clear that the majority of RetroMeter properties fall well outside of the necessary accuracy for use under the ASHRAE guidance, with median daily, monthly, and annual CVRMSE values of 53.4%, 34.2%, and 18.7% respectively. It is also notable that the RetroMeter properties performed poorly compared to Metered Energy Savings 2022 with daily and monthly median CVRMSE values of 37.4% and 20.3%. This is likely due to the absence of the significant rise in gas prices from the baseline to the reporting period that was observed for the RetroMeter properties. It is also possible that the considerably smaller pool of properties has led to some sampling bias in computing the median.

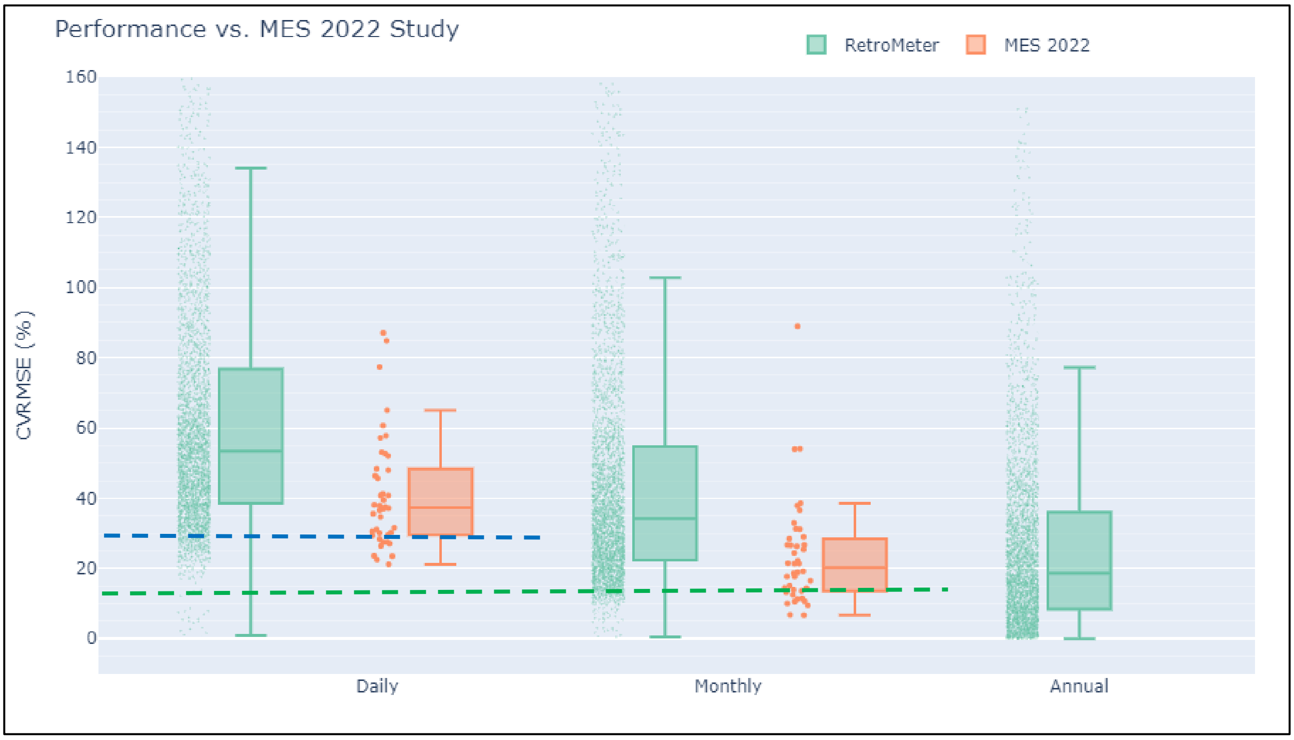

Figure 7 - Distribution of CVRMSE from OpenEEmeter applied to 3,048 properties, aggregated by day, month, and year.

The net bias introduced by OpenEEmeter can be seen in Figure 8, where the prediction error on the y-axis is plotted against the bias on the x-axis for each property in the RetroMeter dataset fitted with a daily counterfactual. The majority of the points lying to the right of the 0-line on the x-axis indicates that most properties have an over-estimated counterfactual, in line with the hypothesis that OpenEEmeter cannot account for the changes to behaviour induced by energy price alone.

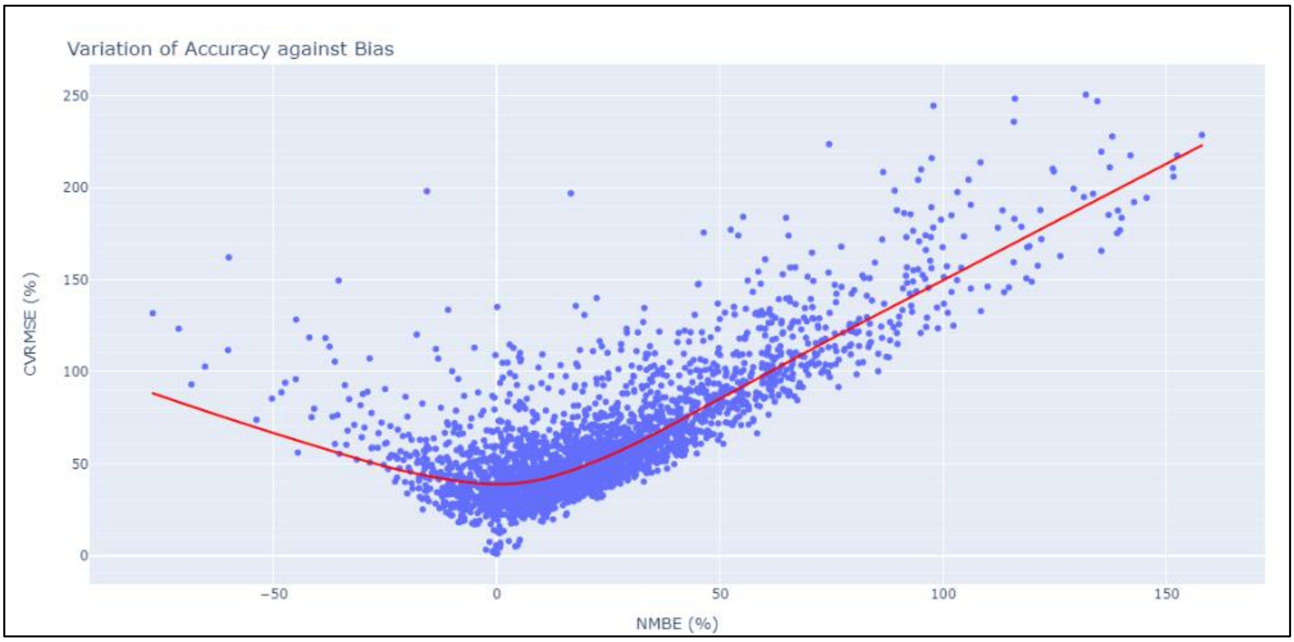

Figure 8 - OpenEEmeter daily counterfactual prediction error against bias with LOWESS trendline.

4.3 Analysis and Variations

A pair of key questions following on from these results were:

- Is it possible to identify ahead of time which properties are likely to perform better?
- To what extent are these results impacted by the choice of baseline and reporting year?

To explore the first question, we broke down the error distribution by property archetype and baseline year energy consumption, to ascertain whether any of these characteristics that are known prior to any interventions can influence the performance of the model.

Binning the properties by their total gas consumption in the baseline year as in Figure 9 revealed a downward trend in error as consumption increases, but an upward trend in bias. One explanation for this is that those properties with the lowest levels of gas consumption (<10,000 kWh/annum) exhibit high degrees of unpredictability and variance, possibly driven by occupancy changes between the baseline and reporting year. An increasing bias may indicate that those properties with particularly high consumption in the baseline year were those who cut back on their consumption the most as prices rose; this would cause OpenEEmeter to overestimate their consumption during the reporting period more severely than those who were consuming a more 'normal' level of energy.

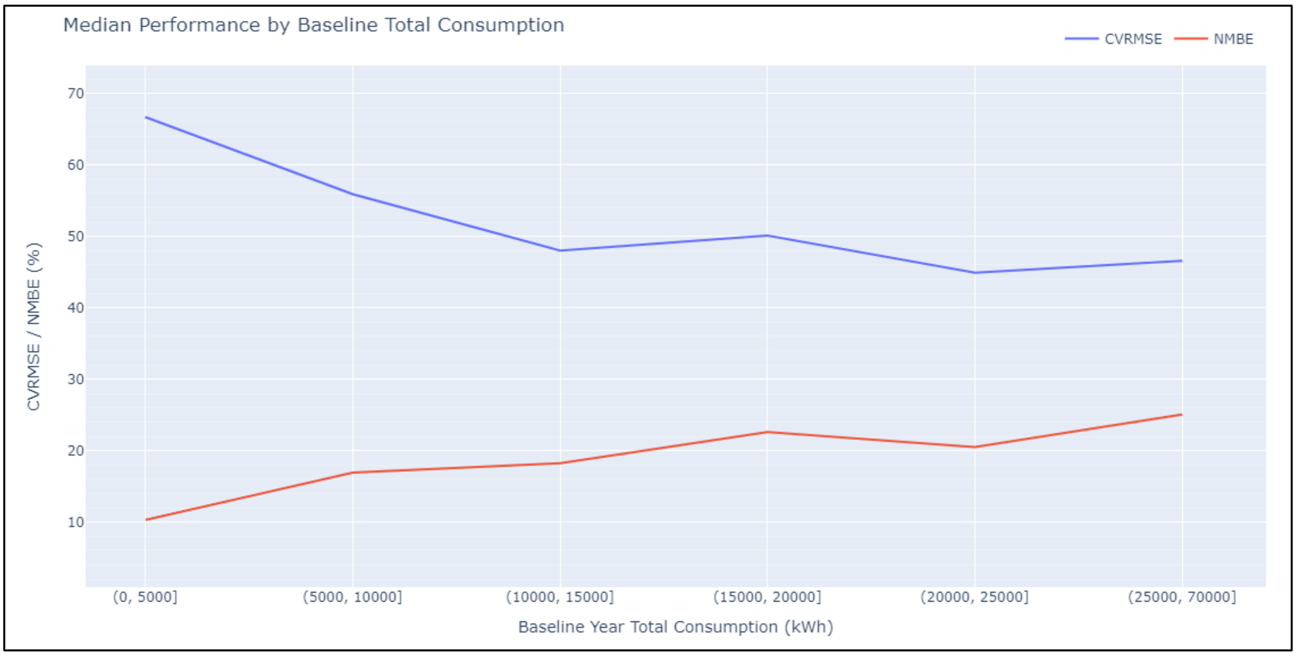

Figure 9- Median CVRMSE and NMBE by Total Baseline Consumption Bin

Similarly, binning by the age of the property showed an increase in both error and bias with age, although to a lesser extent than energy demand. Intuitively, older properties are less likely to be well-insulated, leading to a noisier relationship between the energy consumed and the internal temperature achieved. It was not possible to test this hypothesis rigorously as the distribution of EPC ratings over the properties was significantly imbalanced, with over 98% having either a C, D, or E rating; however, across these three ratings there is a clear trend of older properties tending to be E and below, and newer properties more likely to be C and above. There was found to be no similar relationship between accuracy/bias and EPC rating.

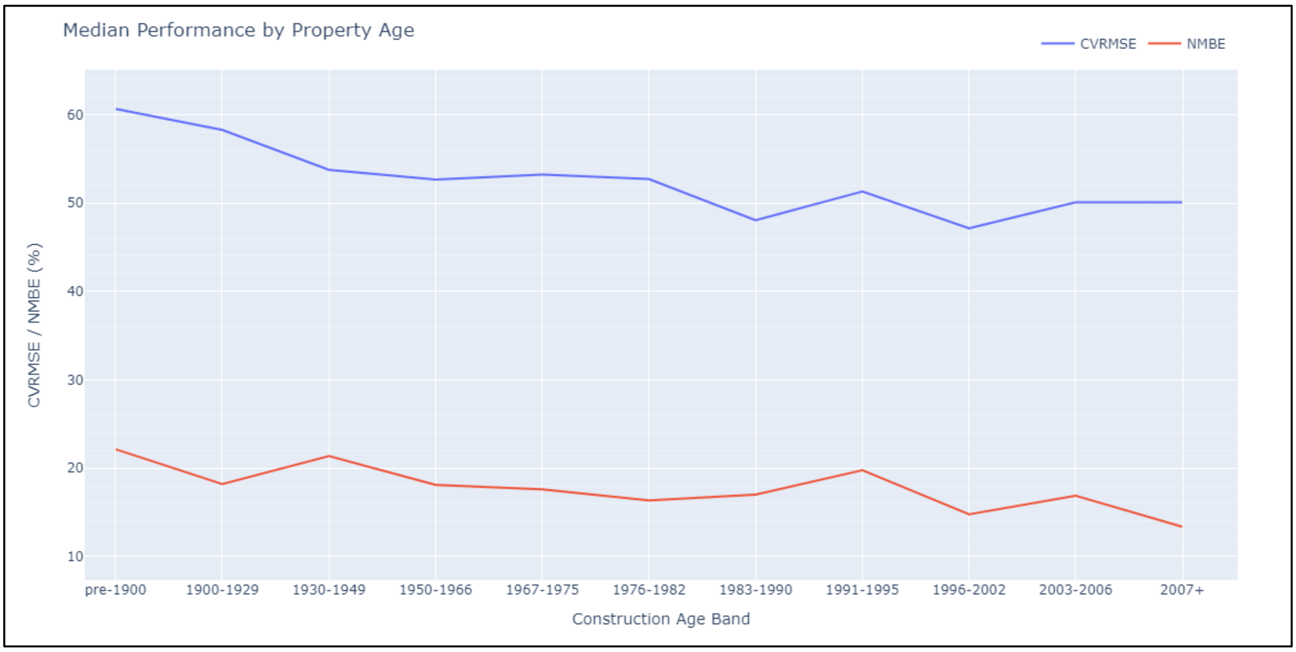

Figure 10 - Median CVRMSE and NMBE by property construction age band.

To explore the additional bias introduced by OpenEEmeter, we also compared the accuracy of the counterfactual against a naïve estimator that uses each day's metered consumption in the baseline year as the prediction for the same day in the reporting period. We would expect a naïve estimator to under-predict consumption in the reporting period of this study as the temperatures are known to be lower on average.

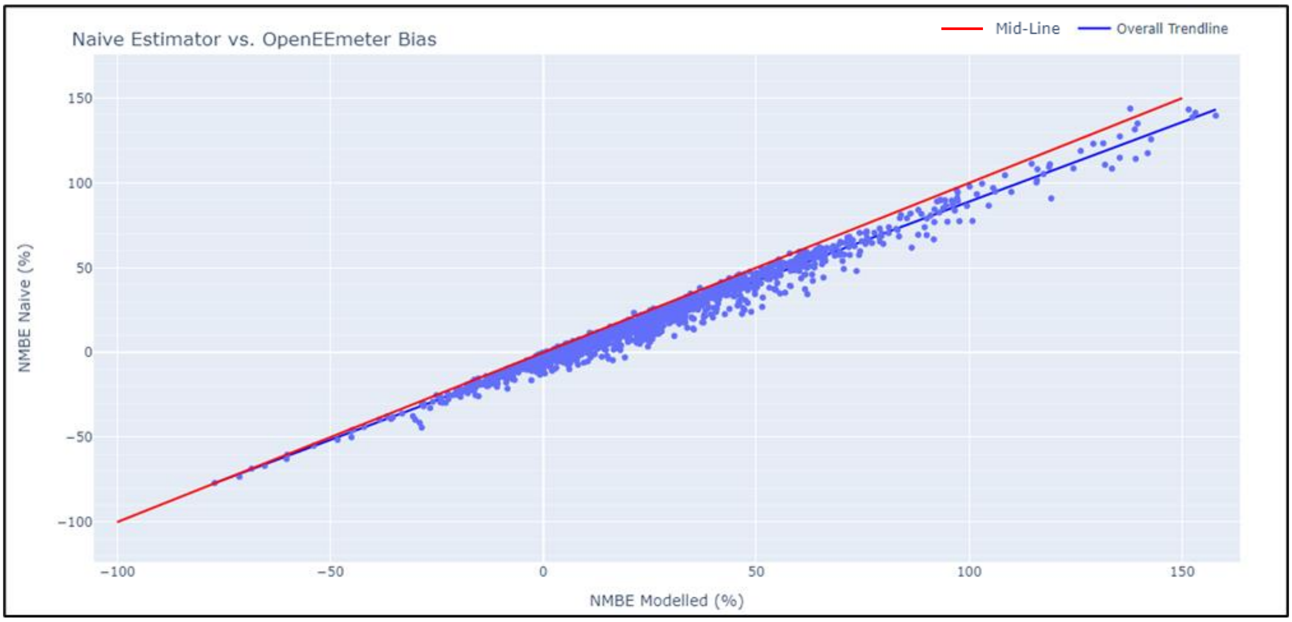

Figure 11 - Bias for each property from using OpenEEmeter and the naive estimator, where the baseline year consumption is used as the prediction for the reporting year.

Figure 11 describes the NMBE by plotting each property as a single point, with the NMBE from OpenEEmeter on the x-axis, and from the naïve estimator on the y-axis. There are two key insights from this figure – we know from the previous section that there is in general an over-estimation from the OpenEEmeter counterfactuals, most likely driven by the price impact coupled with colder temperatures. However, almost all the properties whose OpenEEmeter counterfactual over-predicted their consumption are more biased than if their baseline consumption had been used instead, as in the naïve model. The difference between the naïve and modelled NMBE indicates how much OpenEEmeter is attempting to compensate for the colder temperatures over the reporting period, with the overall NMBE above zero capturing the combined impact of both temperature and price.

In Figure 12, the CVRMSE is plotted in the same way for each property. While this confirms that OpenEEmeter is indeed superior accuracy-wise to the naïve estimator, there is another informative interpretation: the CVRMSE of the naïve estimator is equivalent to a measure of the *similarity* between the baseline year and the reporting year's consumption profile. With this view, we can see that there is a strong relationship between the accuracy of the OpenEEmeter counterfactual and the closeness of the consumption profiles from one year to the next. The distribution also widens as their error on both models increases - those properties with highly dissimilar baseline and reporting consumption profiles exhibit the widest range of OpenEEmeter prediction errors. This may be indicative of the effects of occupancy changes during the measurement period, reinforcing the notion that consistent occupant behaviour is necessary to OpenEEmeter to generate useful results.

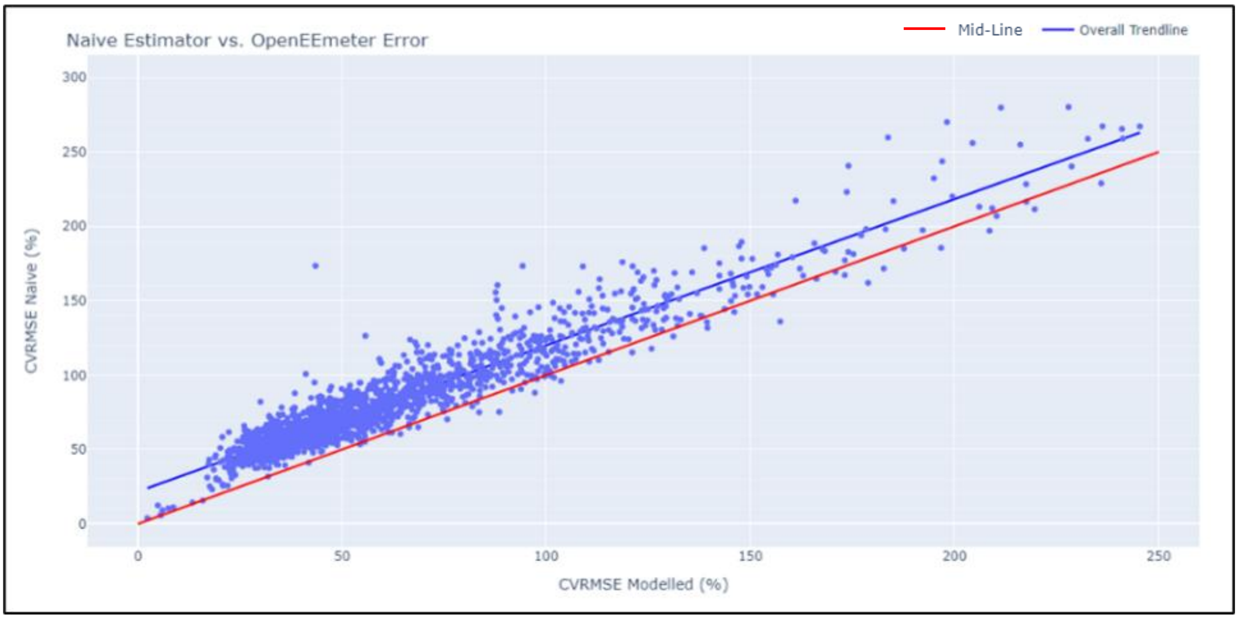

Figure 12 - Error for each property from using OpenEEmeter and the naive estimator, where the baseline year consumption is used as the prediction for the reporting year

4.4 OpenEEmeter Methodology Outcomes

These results have shown that OpenEEmeter alone is not sufficient to generate gas consumption counterfactuals for UK homes accurate enough to quantify metered energy savings, as the impact of externalities beyond temperature cannot be sufficiently accounted for.

The next section introduces the comparator methodology as a means of adjusting the OpenEEmeter counterfactual to remove the impact of externalities that affect all properties over a wide jurisdiction.

5. Using Comparator Groups to Reduce Model Bias

We will refer to 'externalities' in this project as those society- and economy-wide factors that impact residential energy use beyond just weather. The premise of the comparator methodology is that the impact of these externalities that OpenEEmeter cannot account for can be addressed using the counterfactuals from a set of similar properties not undergoing retrofit. This section details how this methodology was developed for the RetroMeter project and the resulting improvements to using OpenEEmeter by itself.

5.1 What are Comparator Groups?

A comparator group is a set of properties that have been matched based on some similarity metric or qualitative grouping to a given candidate property, one that is undergoing a retrofit intervention. Crucially, the comparator properties neither have existing retrofits nor plan to receive any interventions for the duration of the candidate property's baseline and reporting periods. This ensures that the comparator counterfactuals only capture the non-retrofit impacts to energy consumption experienced by the candidate property.

Figure 13 visualises this process with simplified consumption time series before and after the retrofit intervention. By adjusting the candidate property's counterfactual through subtracting the prediction error from the comparator group properties, the 'error' that remains between the adjusted counterfactual and the measured consumption of the property is the impact of the retrofit on energy consumption.

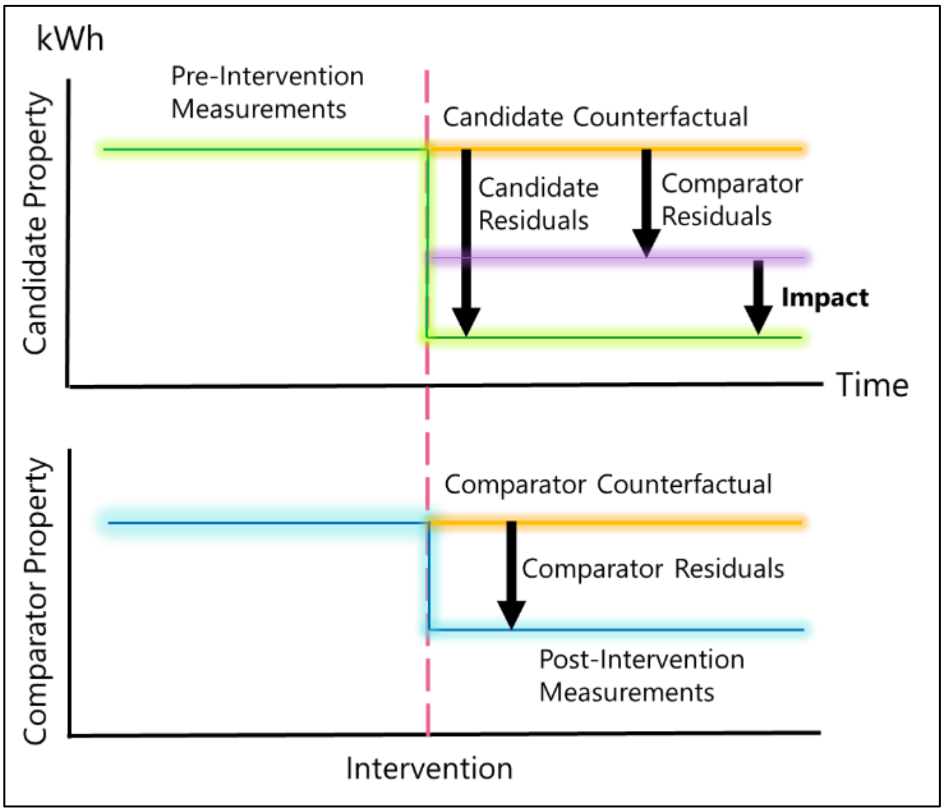

Figure 13 - How counterfactuals are adjusted from comparator property counterfactual error to leave the 'impact' of an energy-saving intervention.

Note that this approach still assumes that there have been no behavioural or occupancy changes in the candidate property other than those that might be reflected in similar homes without retrofits. While the physics-based methodology attempts to account for comfort take-back, a change of tenant[4] or significant alterations to work patterns, for example, remain confounding factors and will be impossible to distinguish from the energy efficiency impact. The most important factor is that the properties in the comparator group should respond in the same way to external non-temperature factors in the same manner as the candidate property, allowing them to be extracted from the counterfactual error.

Figure 14 offers an example of what this adjustment process looks like in practice. As before, the modelled consumption should ideally match the observed consumption in the reporting period because no retrofit has been performed. We can see that the adjusted counterfactual in light green is closer to the measured consumption in red than the original OpenEEmeter counterfactual in blue. See Appendix 8.1 for more detail on how this adjustment is carried out and justified.

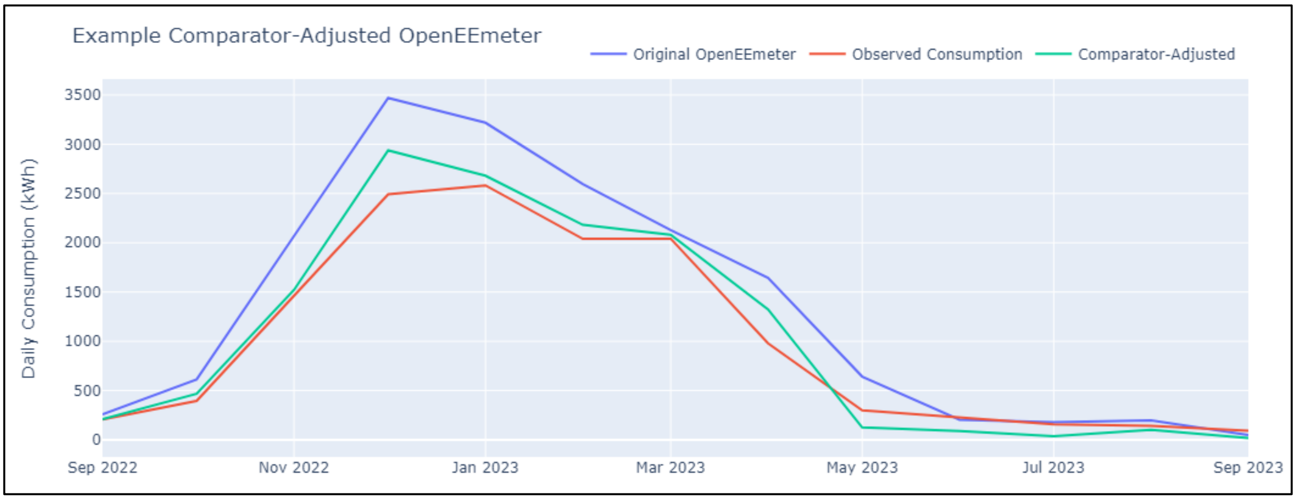

Figure 14 - Example application of comparator methodology.

The comparator methodology was evaluated in two distinct cases, based on whether the candidate property has smart meter available pre-retrofit to match comparator properties to. In each case, a different set of potential matching processes were tested and evaluated for accuracy. In addition, a range of temporal and portfolio aggregation levels were assessed to determine the best trade-off between granularity and predictive power. As was seen in Figure 7, aggregating up to higher timeframes improves accuracy, at the expense of needing to wait longer post-retrofit to attain an impact estimate. Similarly, aggregating multiple candidate properties into a portfolio improves accuracy by smoothing out the noise at the individual level, at the expense of the ability to attribute impact to any one property within that portfolio. Ultimately it is at the discretion of the end-user and application as to how far this trade-off is acceptable; the goal of RetroMeter is to indicate what the expected performance is likely to be along that spectrum.

[4] The DCC does track this information currently, although is in the process of consulting on how to share it more widely.

As there are no interventions in our dataset, the concept of a 'treatment' group, the term typically used to describe the properties undergoing retrofit in other studies, does not translate perfectly into our experiment design. Essentially, in order to extract the most value from the 3,000 properties available to us, each property takes its turn being a 'treatment group' of one, with the remaining properties in the pool acting as the control group from which the best-matching comparators are drawn. When testing the impact of portfolio size on performance, the properties within the portfolio become the treatment group, in that none of them are eligible to be comparators for other properties in the portfolio. We shall continue to use the term candidate to refer to the property whose MES is being measured.

5.2 Principal Results

The first test of the comparator methodology was performed using only each property's archetype data. A segment of the overall pool was created for each candidate property in turn with the same property type and built form, as well as the same or adjacent EPC rating and age band. If we created property segments based strictly on properties with the same type, built form, EPC rating, and property band, there would be many cases where the segment would be too small to select enough comparator properties from. Therefore, we also allowed matching on adjacent EPC and property age bands to compensate – for example, properties with an 'F' EPC rating were also segmented with 'E' and 'G' properties. As there are no quantitative metrics in this iteration, the comparator properties were drawn at random from the segmented group.

5.2.1 Matching Only on Archetype Removes Bias but Increases Variance

Figure 15 compares the results from the comparator methodology using only the property archetypes to match, against the OpenEEmeter and MES 2022 results from the previous section. Despite the accuracy of the model decreasing, the net positive bias that OpenEEmeter introduced due to the price increase was successfully removed, as shown in Figure 16 (only the daily aggregation is illustrated; as NMBE is an absolute measure of bias, it does not change when the predictions are aggregated).

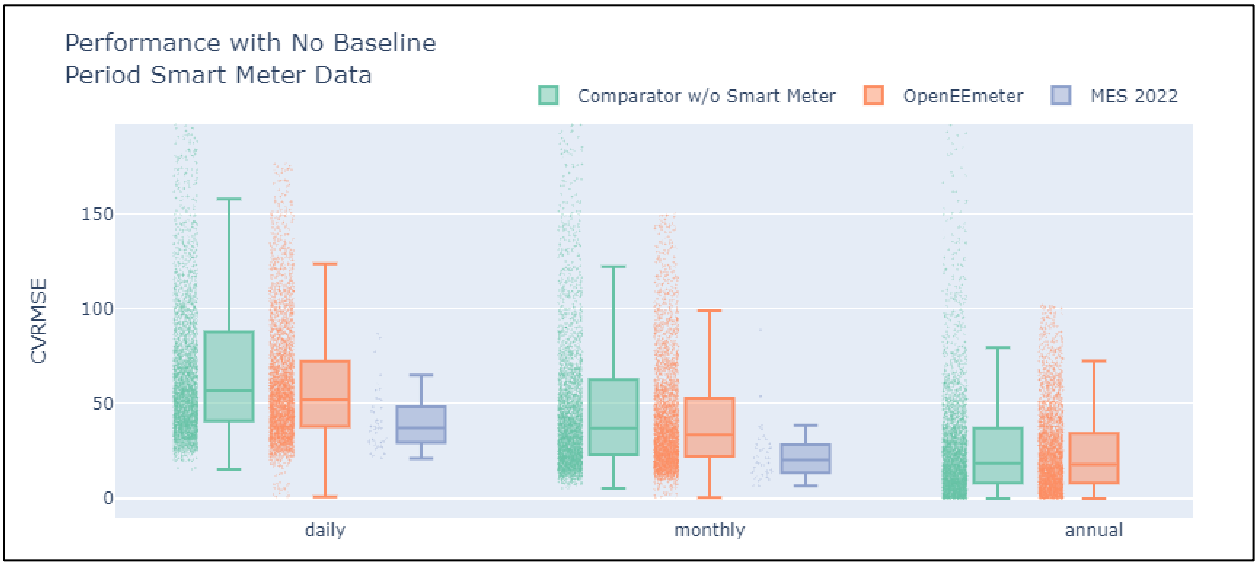

Figure 15 - Comparator methodology performance using property archetype information only

It is notable also that the net bias of the MES 2022 results is also around zero, further highlighting the impact of the gas price increase in this project. In addition, the impact of restricting the comparator properties to be selected from the same weather station as the candidate was examined, although the performance was worsened considerably, most likely due to the high class imbalance between stations. Other studies that have applied matching over the same climate zone have had access to considerably more properties than are available to us in this study, therefore we cannot conclude from this result that we should not match on climate zone. In practice, a larger pool of more evenly distributed property-to-weather stations would be needed to explore this requirement.

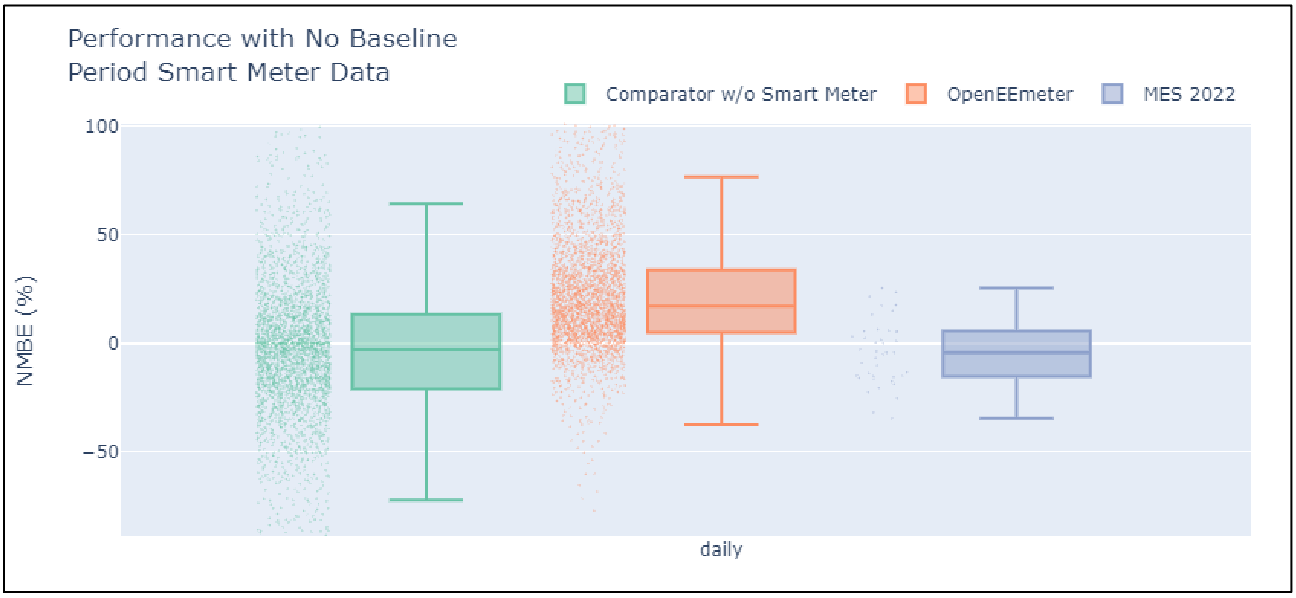

Figure 16 - Comparator methodology NMBE using property archetype information only.

The importance of each of the four property archetype characteristics available was evaluated by removing them one at a time from the full set of four and re-running the comparator methodology. For example, the impact of EPC was evaluated by segmenting the pool only by the candidate's property type, built form, and age band, selecting comparators with any ECP rating from that segment. Figure 17 shows that removing the age-band matching condition improves the median error compared to using all four archetypes by nearly two percentage points, while leaving out the built form worsens the error.

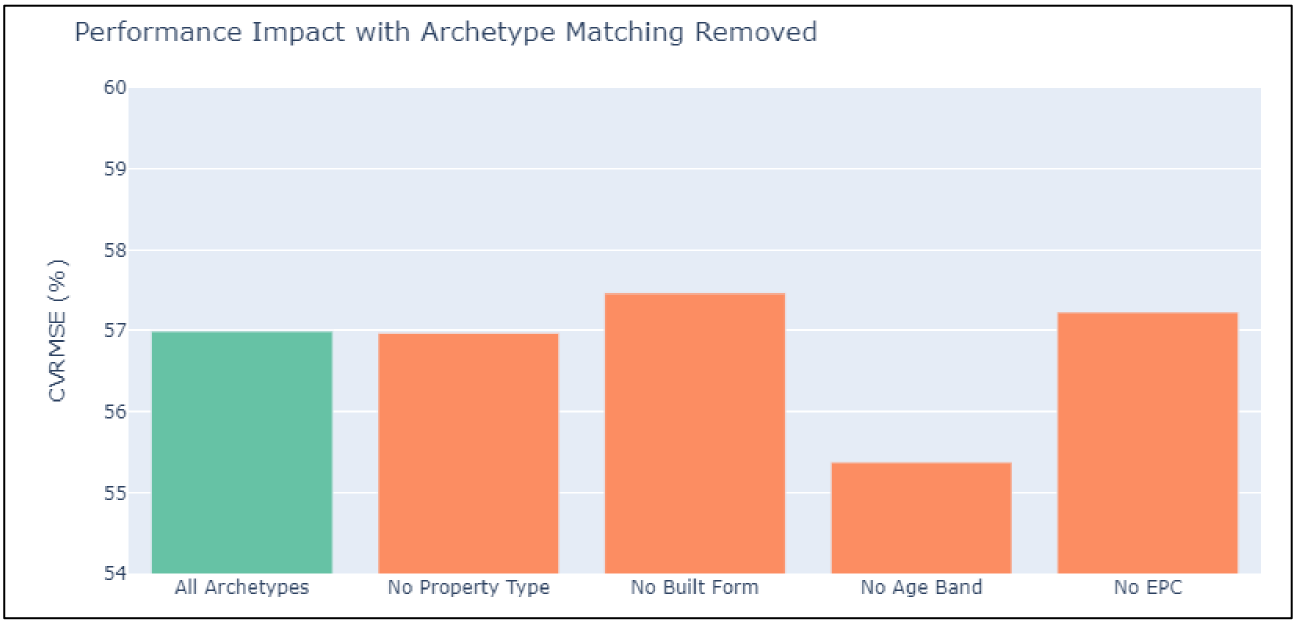

Figure 17 – Impact on median CVRMSE of removing each property archetype from the set used to match properties on performance.

5.2.2 Matching on Pre-Intervention Smart Meter Data Performs Better

If smart meter data for comparator properties is available prior to the candidate property's intervention, then we can further refine the matching approach beyond simply selecting randomly from the archetype property segment. The simplest approach is to select properties from the segment according to how similar their total and peak consumption during the baseline year was[5].

However, this leaves out a lot of information contained in the daily consumption profiles of the properties. The shape of the demand curve for a property can be used to infer occupancy and other behavioural factors that are important for estimating how the household may respond to externalities in the reporting period. Rather than try to infer these factors directly, they can be implicitly utilised by using the consumption profiles themselves to match candidate and comparator properties. Many metrics exist for quantifying the similarity of two time series, but for simplicity we have opted to continue using the CVRMSE as a de-facto indicator of how close two baseline consumption profiles are – the lower the error, the more similar they are and the more likely they are to be matched.

Matching on profile similarity alone was found to be the best-performing metric for individual properties. The combinations of factors and parameters tested are summarised in Figure 18, showing the median CVRMSE achieved across each aggregation and method, including results from several other heuristics described in Section 5.3. Although the difference in performance across the latter three feature combinations is small, it is notable that the best-performing matching approach is also the simplest – giving the candidate

[5] Specifically, by taking the Euclidean distance between any quantitative variables and selecting the closest matches.

property the widest possible pool of comparators to match against, without pre-filtering by archetype or baseline filter, appears to work best.

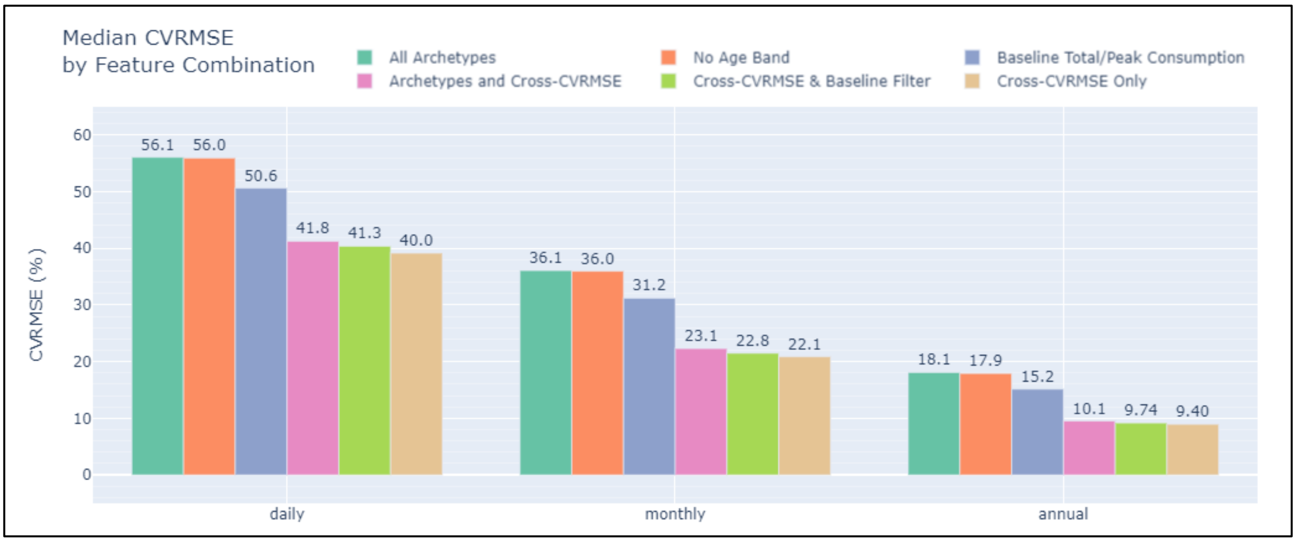

Figure 18 - Median CVRMSE from each feature combination tested.

5.2.3 Aggregating Portfolios of Properties Performs Best

Whilst these results demonstrate an improvement on matching against property archetype data alone, they are on average not quite accurate enough for meeting the ASHRAE guidelines at the monthly level. An alternative approach that has also been explored in previous studies is to combine the consumption profiles of multiple properties into a portfolio, against which the total impact of any interventions is measured. This has the effect of smoothing out the variance in individual property's consumption, as can be seen in Figure 19 as the number of properties per portfolio is increased.

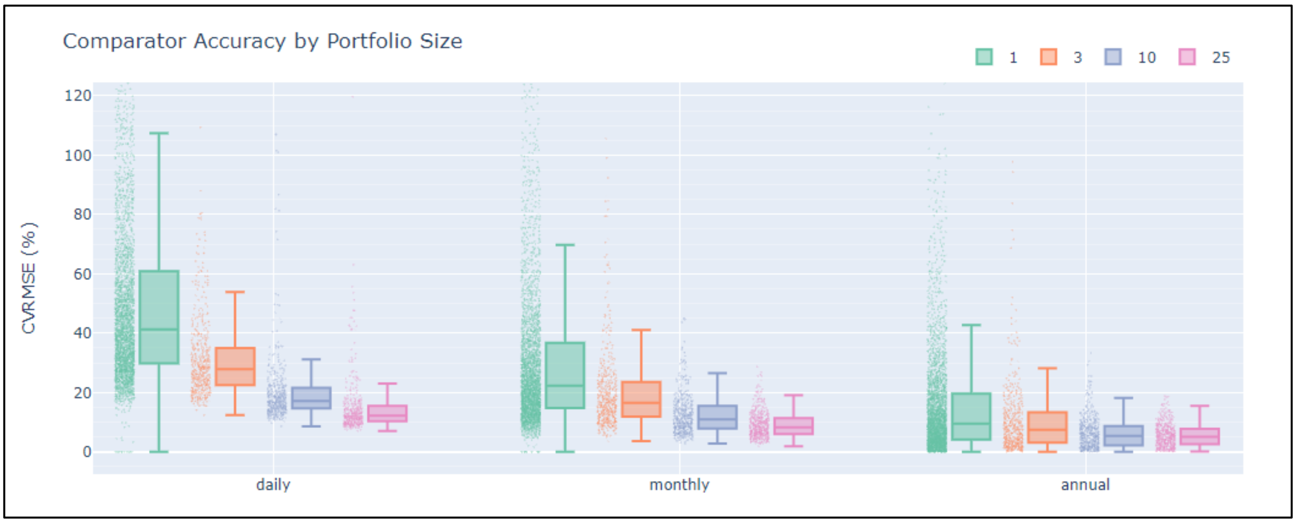

Figure 19 - Comparator accuracy by portfolio size.

The portfolio aggregation methodology involves matching each candidate property within the portfolio to its own set of comparator properties as before, with the restriction that no candidate within the portfolio can be a comparator property for another candidate, nor can any candidate property share comparators between their own groups. While this method successfully reduces the error to as little as 5% at the annual, 25-property portfolio point, it comes with some practical caveats that end-users must be aware of:

- The candidate properties within the portfolio must have had their interventions completed at around the same time, so that their baseline and reporting periods line up. This is necessary for ensuring that each property is fully represented at each timestep of the aggregated reporting period.
- They must also be sufficiently physically close to each other so that the same external temperature readings can be applied to each.
- MES cannot be disaggregated and attributed to individual properties with this approach.

These limitations imply that the portfolio aggregation approach is best suited to cases where a group of properties, managed by the same owner and on a single estate or terrace for example, can be retrofitted at the same time, and tied to a monitoring mechanism this is satisfied with attributing the MES to the project as a whole rather than individual properties.

5.3 Analysis and Variations

There are many factors and modifications that can be made to potentially improve the performance of the comparator methodology. To find the optimum combination within the scope of this project, we selected a number of options from both expert opinion and through assessing approaches tested in the literature to test against our dataset.

5.3.1 Finding the Appropriate Number of Comparators

Intuitively, the more comparator properties that are matched to the candidate property, the more smoothed out the counterfactual adjustment will be, eliminating more of the noise in the comparator group. To test this, we repeated the comparator matching using the property archetypes excluding the age band, matching on the total and peak baseline consumption, with 5, 10, and 25 comparators per property respectively. Figure 20 reveals an interesting result – the performance of the comparator methodology is improved at the daily level when adding more comparators, but is *worsened* for the monthly and annual aggregations.

One explanation for this may be that the monthly and annual aggregations already remove enough of the noise in the candidate property's counterfactual prediction that the comparator properties only end up increasing it post-adjustment. Whereas at the daily level, the smoothing of comparator properties' counterfactuals as they are added together before adjusting the candidate property reduces the noise in the candidate counterfactual without introducing additional variance.

In practice, it is more likely that the MES will be assessed on a monthly or annual rolling basis. For this reason, we chose to keep the comparator group size fixed at five properties for all iterations to get the best possible performance out of the monthly and annual aggregations.

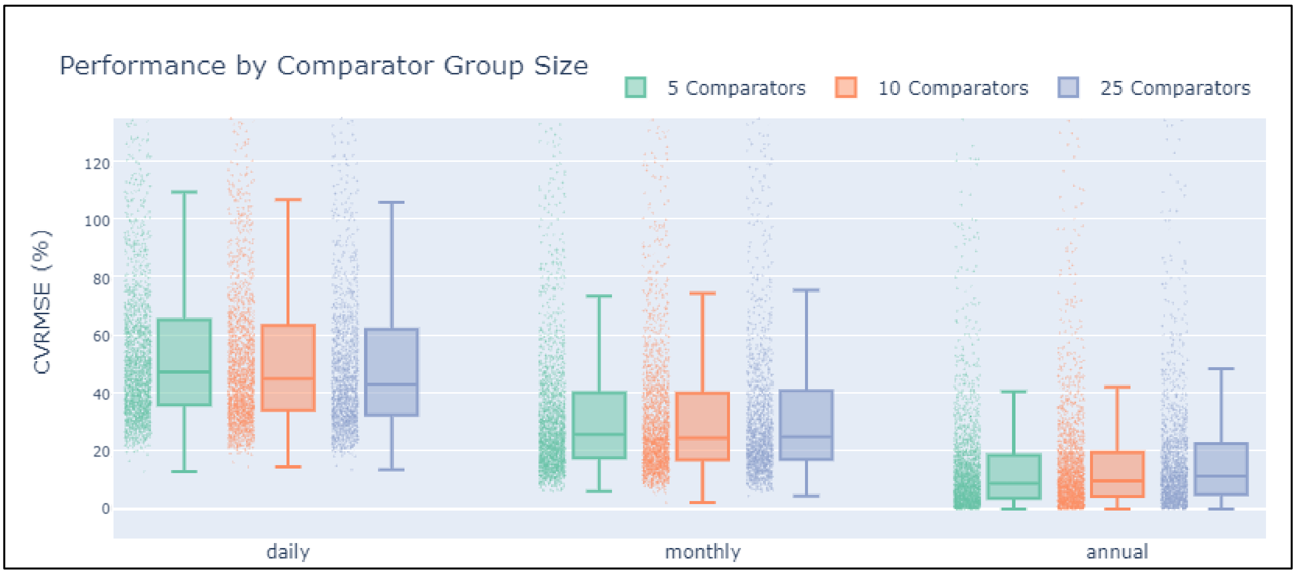

Figure 20 - Variation of performance with comparator group size

5.3.2 Other Factors that May Improve Matching Performance

Homes with hard to predict, noisy energy demand are unlikely to be as useful as comparator properties, as they are likely to introduce additional noise to the measurement of MES. Filtering on reporting period error would introduce a look-ahead bias, but it is reasonable to assume that properties that were hard to predict in the baseline period will also be hard to predict in the reporting period. Therefore, an effective way to remove high noise properties is to only use properties with a low baseline period CVRMSE for comparators.

Figure 21 illustrates the distribution of reporting period error, with the orange reading indicating those properties with a baseline period CVRMSE greater than 45%. This threshold was chosen after some testing to be the best trade-off between improving the quality of the comparator property counterfactuals, versus removing too many properties from the pool such that the best available comparators by consumption profile similarity to the candidate are weaker. As the size of the pool of possible comparators is increased, this threshold can be made stricter to further improve the comparator counterfactual quality without significantly worsening the number of remaining properties.

Note that this filtering is not applied to the candidate (i.e. retrofit) properties, only those being used for comparators. This approach assumes that any hard to predict candidate properties can be reasonably matched with easier to predict comparison properties, which is probably a reasonable assumption but may not hold for all homes.

Additionally, adjusting the candidate counterfactual by the percentage residuals of the comparator was found to perform better than using the raw residuals, likely as it normalises the differences in absolute consumption between candidate and comparator properties.

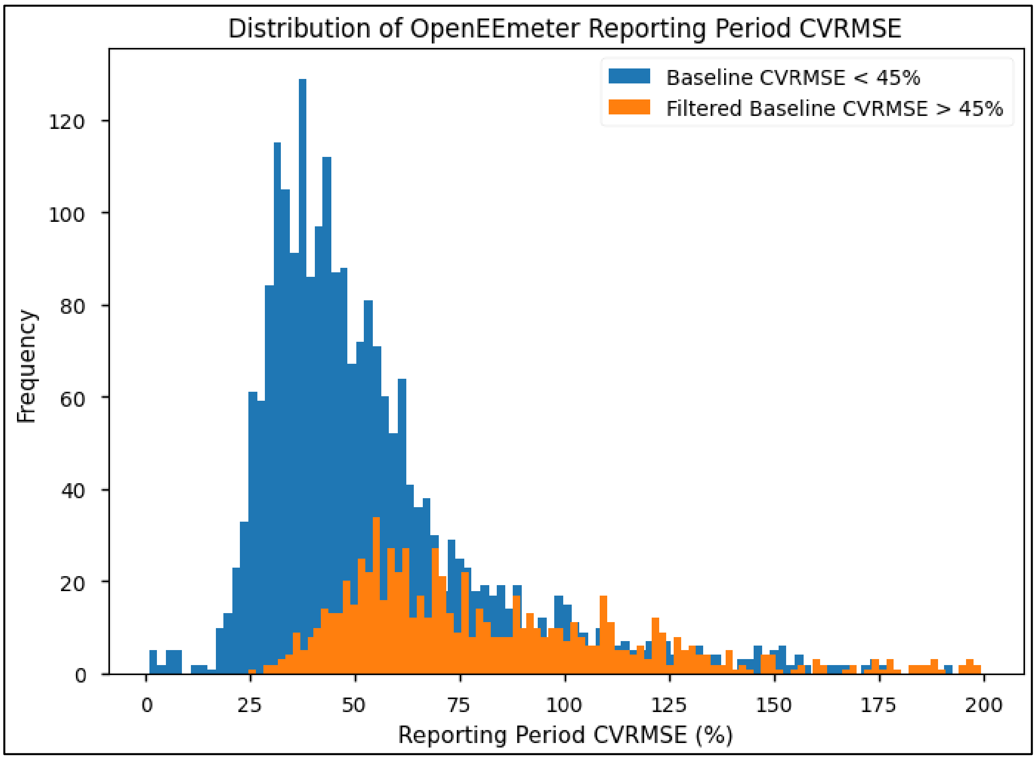

Figure 21 - Distribution of OpenEEmeter reporting period CVRMSE values, with baseline period CVRMSE threshold of 45% highlighted.

5.4 Comparator Methodology Conclusions

The comparator methodology has been shown to effectively control for the externalities that affected the performance of using OpenEEmeter by itself, improving the median error at the monthly level from 34% to 22%, whilst also eliminating the net positive bias. Aggregating the candidate properties into portfolios further improved the error, up to 8.3% at the monthly level with 25-property portfolios, beyond which the returns appear to diminish. We have demonstrated that you do not need a large number of homes to aggregate in order to produce acceptable accuracy, with only around 5 homes needed to bring the median monthly accuracy within the ASHRAE guideline value.

The comparator methodology has some limitations that would make it more complex to apply to real property interventions than OpenEEmeter alone:

- Maintaining access to a repository of comparator property smart meter data, against which any arbitrary candidate property could be matched and evaluated, will require close collaboration with an industry or academic partner – this is discussed in more detail in Section 7.2.
- Ensuring that none of the comparator properties matched to a candidate have any interventions or significant behavioural changes, such as occupancy, over the reporting period will not be possible, so a procedure for replacing comparator properties that are no longer valid during evaluation may be necessary.
- Behavioural changes for the candidate property post-intervention still cannot be accounted for. The physics-based methodology in the next section describes how this might be achieved.

6. Adjusting for Comfort Takeback through Physics-Informed Models

After a retrofit, occupants may change how they heat their homes, often increasing the internal temperature. This phenomenon, known as comfort takeback or comfort boost, means a portion of the energy savings are taken as enhanced comfort levels. The physics-based model, introduced in this section, aims to quantify the comfort takeback through a counterfactual physics-based model. This is shown in Figure 22.

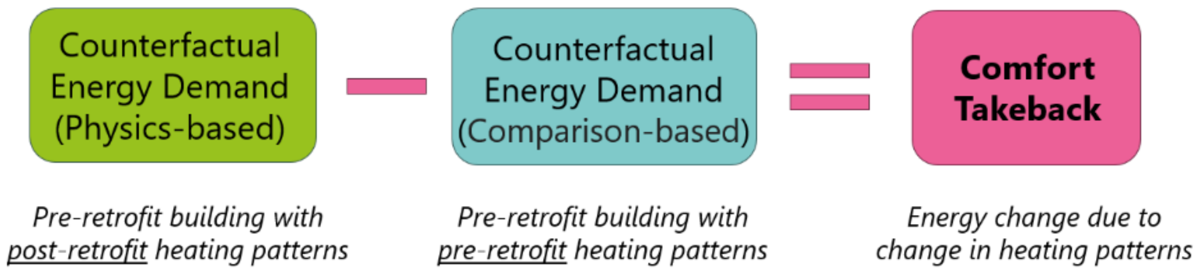

Figure 22- Estimating the Comfort Takeback by subtracting the Comparison-based Counterfactual Energy Demand from the Physics-based Counterfactual Energy Demand

6.1 How do we estimate the comfort takeback?

First, the pre-retrofit heat transfer coefficient (HTC) for each house is calculated. This can be done using existing techniques for measuring HTC, or using the methodology outlined below which relies on smart meter and external temperature data. The counterfactual energy demand can be estimated using the pre-retrofit HTC in combination with the post-retrofit internal temperature, as shown in Figure 23. It represents the energy demand that would have occurred if the occupants had been heating to those comfortable temperatures without retrofitting.

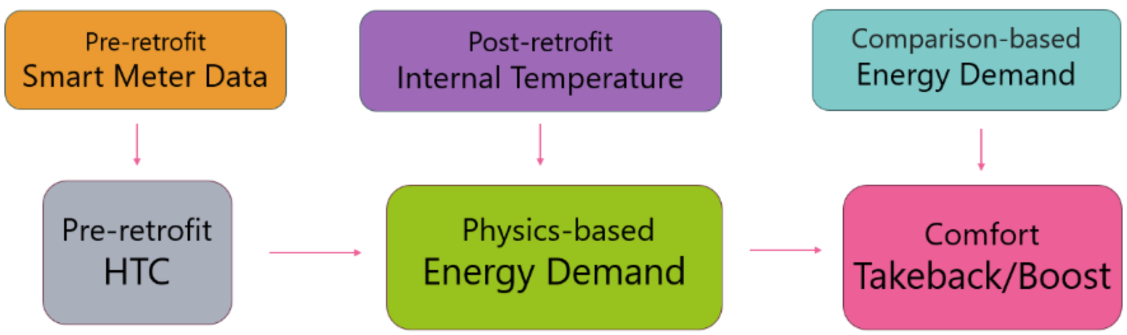

Figure 23 - The counterfactual HTC energy demand (or physics-based demand) is calculated using the pre-retrofit HTC and the post-retrofit internal temperature.

6.2 HTC Estimation from Smart Meter Data

6.2.1 What is the HTC?

The heat transfer coefficient (HTC) is a measure of a building's thermal performance, given in W/K. It quantifies the power required to keep the house at a steady internal temperature for every degree difference between the inside and the outside temperature. A lower HTC means that the house requires less energy to stay warm. In the UK, the average HTC is 330 W/K according to the Cambridge Housing Model (Chambers & Oreszczyn, 2019).

6.2.2 How do you calculate the HTC?

The HTC can be estimated by calculating the best line of fit between the temperature difference ($T_{in} - T_{ext}$) and the power required to maintain a steady internal temperature, as shown in Figure 24. The slope of the line represents the HTC.

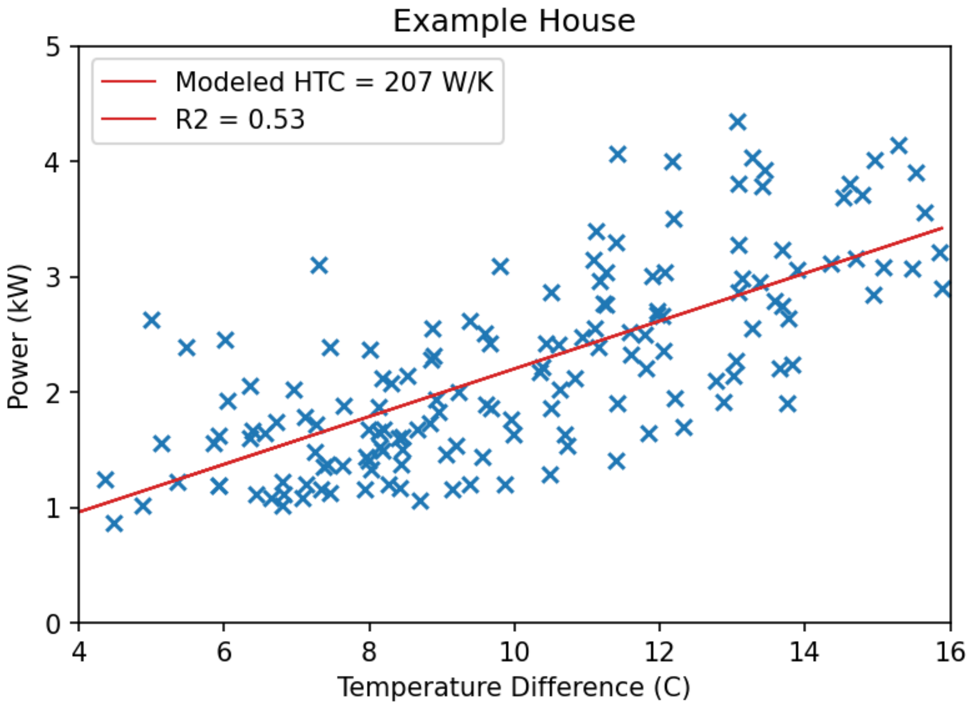

Figure 24 - The HTC is the slope of the regression line between the power and the temperature difference.

6.2.3 Basic Method & Applied Filters

To get the total energy consumption, both electricity and gas usage are accounted for, as including electricity yielded better results (lower CVRMSE and lower absolute NMBE). This can be partially explained by the usage of additional electric heaters in the UK. Even without secondary heating, electricity consumption mostly dissipates as heat. Energy is then aggregated to a daily level and the average power calculated. A balance point (the external temperature at which heating is assumed to turn on) of 15°C is used for all houses. This could be further individualised, but we found that improvements are marginal. Due to the chosen balance point, we filtered for days with mean average external temperature below 15°C. We further filtered for days with gas consumption above daily baseload gas consumption.

6.2.4 Daily Baseload

The daily baseload is calculated by filtering for high temperature (above 15°C) and high solar irradiance (above 50 W/m^2) days. These are days that typically do not require heating. Thus, the energy used during those days can be attributed to baseload activities, e.g. cooking and hot water usage. The median of the resulting distribution is deemed to represent the daily baseload for each house, as illustrated in Figure 25.

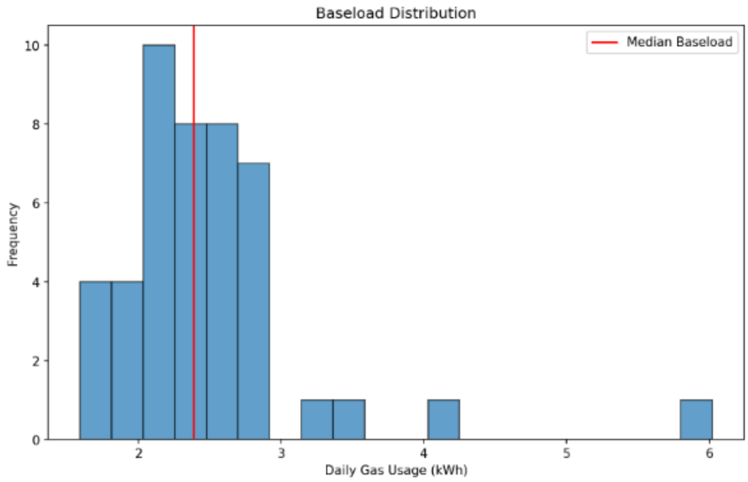

Figure 25 - Distribution of Gas Usage after filtering for Non-heating Days.

6.2.5 Internal Temperature Not Available

Pre-retrofit the internal temperature is not available. Instead, we can use the external temperature on the assumption that the internal temperature remains constant. This shifts the regression line without changing its slope – and the HTC is based on the slope. However, using only external temperature yields less reliable results compared to using the internal temperature. This suggests that the internal temperature does in fact vary.

In general, omitting the internal temperature from the model leads to the underestimation of the HTC. This is likely due to some days where the house is heated to lower internal temperature than usual. Since internal temperature is not available pre-retrofit, we need to approximate these low internal temperatures days.

6.2.6 Residual Filter,

Removing days where the internal temperature is low should improve accuracy of HTC estimation. However, as internal temperature is not available, another way to identify these days is required. Identifying days with lower than expected power usage for a particular external temperature seems a good proxy. First, we regress on all days, resulting in the red regression line in Figure 26. Subsequently, the residuals for all data point are calculated as the distance from the regression line. Data points below the first quartile (i.e. 25%) are eliminated, marked by red crosses in Figure 26. Note that if the internal temperature was in fact constant, this would be removing valid data, but it would not be expected to change the slope of the line (and therefore the HTC).

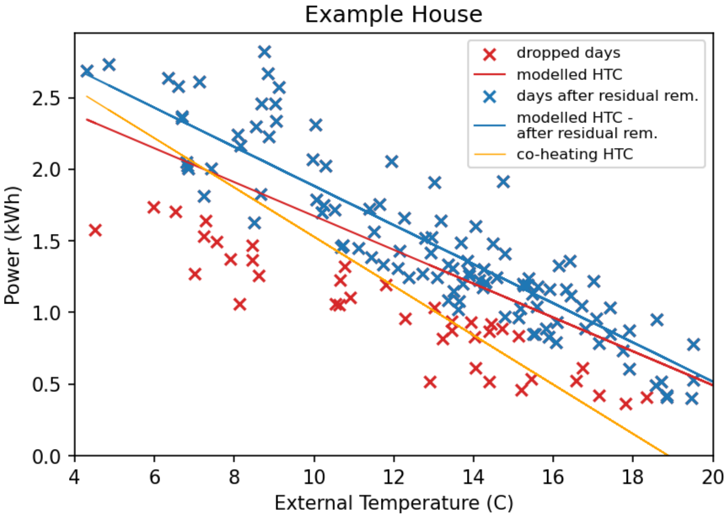

Figure 26 - HTC regression before (red) and after (blue) residual removal, and co-heating HTC (yellow)

6.3 HTC Model Results

The HTC can be estimated from only smart meter and weather data, although not as accurately as when the internal temperature is available. Without the residual filter adjustment for low internal temperature days, the model tends to underestimate the HTC, as can be seen in the left-hand graph of Figure 27. However, using the residual filter improves the model fit and reduces the bias as shown in the right-hand graph in Figure 27. Note that the model might have overfit to the 15 SMETER houses due a lack of anticipated validation data.

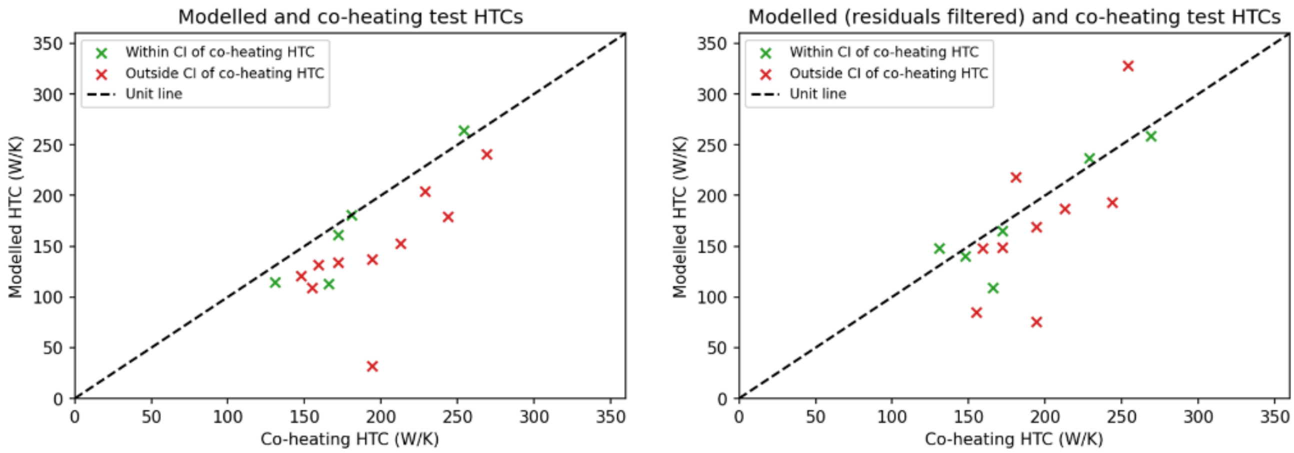

Figure 27 - Modelled HCT vs Co-heating HTC.
Left: without residual filter. Right: with residual filter.

6.4 Energy Demand Model

Now that we have the pre-retrofit HTC, we can use it to calculate the counterfactual heating energy demand. This can be done like in Equation 1. A breakdown of the approximate impact of each factor is shown in Figure 28.

$$\mathrm{P_{tot}} = \mathrm{HTC} * (\mathrm{T_{in}} - \mathrm{T_{ex}}) - \text{ solar gains - electric gains} + \text{base load}$$

Equation 1

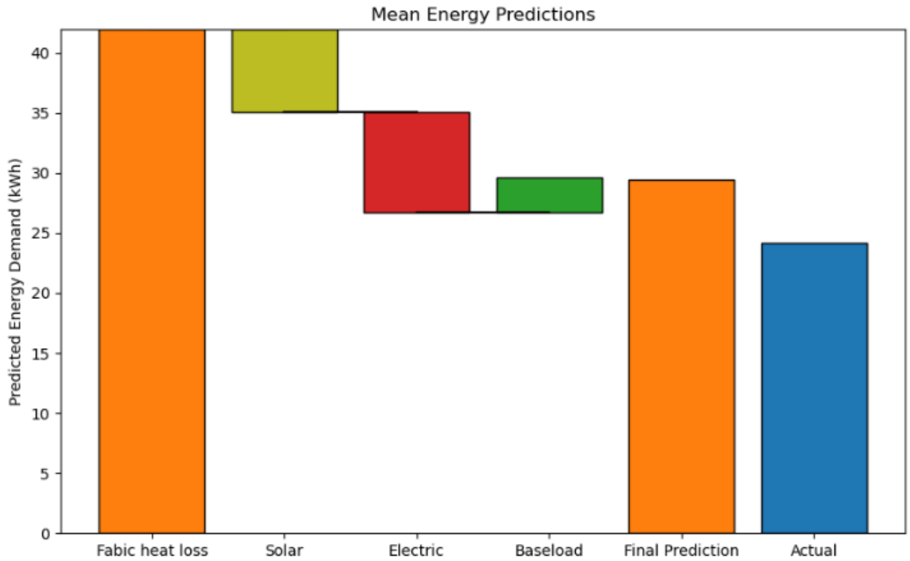

Figure 28 - Waterfall Graph of Predicted Energy Demand compared to Actual Energy Demand

6.5 Energy Demand Model Components

This section features a closer look at the individual factors determining the energy demand prediction, as shown in Equation 2.

$$\mathrm{P_{tot}} = \frac{[\mathrm{HTC}*(\mathrm{T_{in}} - \mathrm{T_{ex}}) - \mathrm{Irrad}*\mathrm{Aperture} - \mathrm{electricity}*F_{elec_{abs}})]}{boiler\ efficiency} + \text{base load} * (1 - F_{b\_abs})$$

Equation 2

6.5.1 Solar Gains

Solar gains refer to the heat absorbed from the sun, determined by multiplying irradiance by the solar aperture. To calculate the solar aperture, we used the Siviour method which is explained in more detail in Appendix 8.2.

6.5.2 Electric gains

The electric gains are derived from the total electricity consumption. While not 100% of electricity is converted into heat, assuming this is the case is simpler and tends to compensate for metabolic gains from people being in the house (since more electricity is used when people are in the house).

6.5.3 Baseload

The baseload is calculated as described in section 6.2.4 (Daily Baseload) and multiplied by (1 - F_{b_abs}), where F_{b_abs} is a baseload absorption factor. It indicates how much of the baseload is absorbed as heat (e.g. gas cooking heats a room). It is multiplied by (1 - F_{b_abs}) because 100% of the baseload is first added to the energy demand and then F_{b_abs} times the baseload is subtracted as gains.

6.5.4 Boiler Efficiency

Not all gas used by a boiler results in useful heat for the house, so a boiler efficiency needs to be assumed. This will vary from house to house, but a representative figure of 89% was assumed for this project, as it was the efficiency for the SMETER boilers (Allinson, et al., 2022).

6.6 Energy Demand Model Results

Figure 29 presents the outcomes of the energy demand prediction, compared with results from OpenEEmeter and the Comparison-based methodology. When the co-heating HTC is used as input into the energy demand model, it results in a median CVRMSE around 20%. The skew towards a positive NMBE, indicates a slight overprediction bias. Using the modelled HTC produces a similar CVRMSE as the comparison-based model. Without the residual adjustment, the predictions are biased towards underestimating the energy demand (indicated by the negative NMBE). However, adjusting the HTCs eliminates this bias.

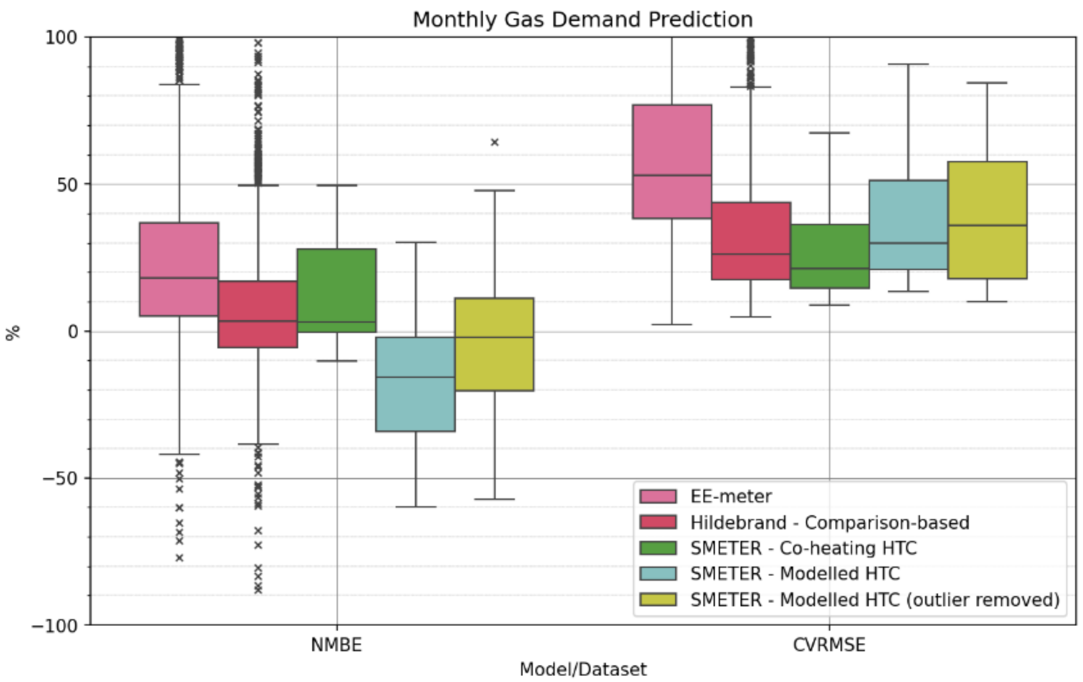

Figure 29 - Energy Demand Prediction CVRMSE & NMBE

Figure 30 shows the demand model results on a per house basis when using the co-heating HTC. Most houses fall within the acceptable range of CVRMSE (<30%) and NMBE (<15%). However, the accuracy of the energy demand model decreases when using the modelled

HTC as input. Whilst the model is probably not reliable enough to predict energy demand on a per house basis, the results might be useful when aggregated to a portfolio of houses.

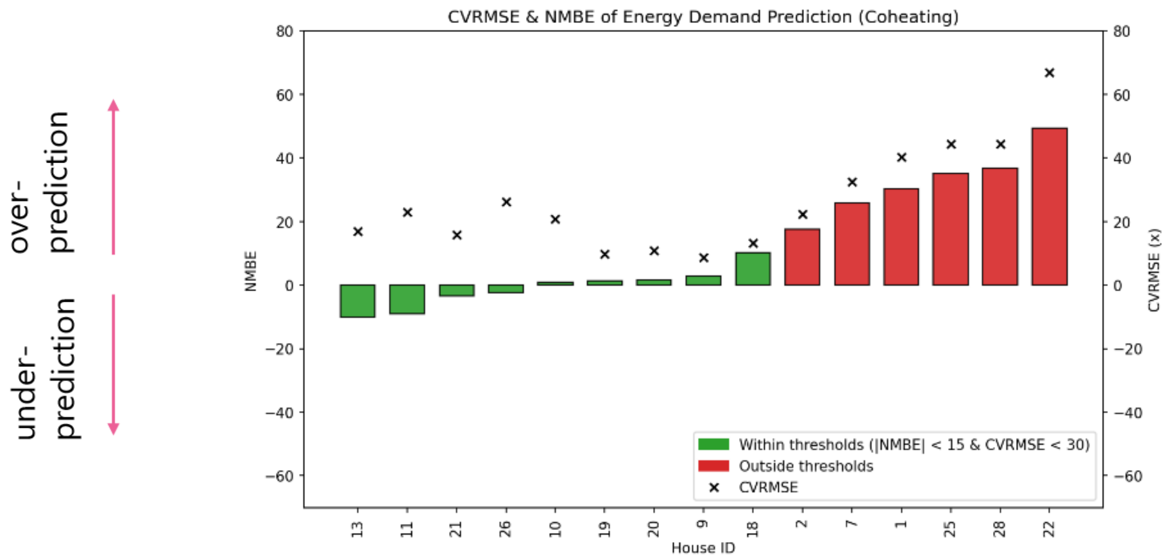

Figure 30 - Energy Demand Model Results on a per House Basis using the Co-heating HTC as Input

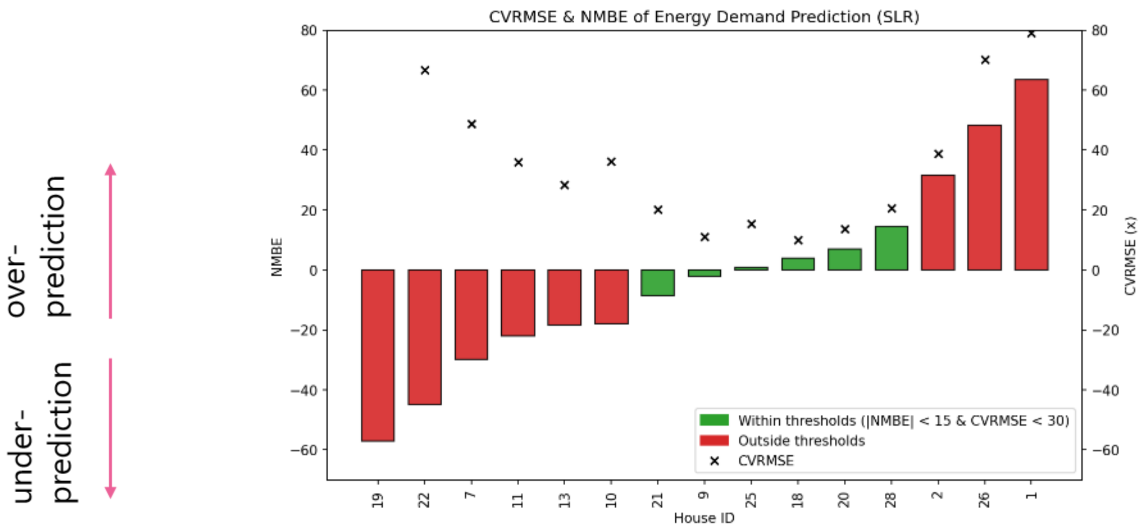

Figure 31: Energy Demand Model Results on a per House Basis using the modelled and adjusted HTC as Input

6.7 Suitability for Comfort Takeback Assessment

The HTC can be accurately estimated using internal temperature data. However, before retrofitting, this data won't be accessible in most cases. The HTC can still be determined using smart meter and weather data by applying the residual filter. Predicting energy demand with the modelled HTC yields results comparable to those of the comparison-based methodology. Since performance differs across individual homes, home-specific conclusions are more uncertain. Rough analysis suggests that comfort takeback levels greater than ~40% of the baseline energy consumption might typically be detectable for a single home (which is unlikely to be useful), or ~25% for groups of 10 homes. This suggests larger portfolios would be required to detect comfort takeback reliably.

7. Additional Considerations

In addition to the three core methodologies described in this project that make up the RetroMeter method, there are several more points that are either out of scope for a full analysis, or points that will require more work in the future should the method continue to be refined and applied.

7.1 Treating Properties with Heat Pump Interventions

For the case where a property intervention includes the conversion of its existing mains gas heating system to an air- or ground-source heat pump, the RetroMeter approach requires some modification. Heat pumps further complicate the evaluation of MES when there is also a fabric retrofit installed simultaneously, leading to two distinct cases:

- We only care about the total energy saved, so the MES of the heat pump and fabric retrofit can be treated as a single intervention.
- The MES impact of the heat pump and the fabric retrofit need to be disaggregated and treated separately. This requirement is more complex as the reporting period counterfactual now needs to be transformed into 'what would the energy consumption had been if the property *had always had a heat pump*?'.

In the first case, the application of RetroMeter is straightforward, with the reporting and baseline years in Figure 32 illustrated for a property simulated using the Catapult's Home Energy Dynamics (HED model, with an ASHP intervention at the start of January 2023. It can be seen that the metered consumption on a kWh basis has reduced considerably post-intervention. One simple interpretation is to measure the MES as the straight difference between the gas counterfactual and the metered ASHP consumption, but this ignores the obvious differences between a kWh of gas and a kWh of electricity.

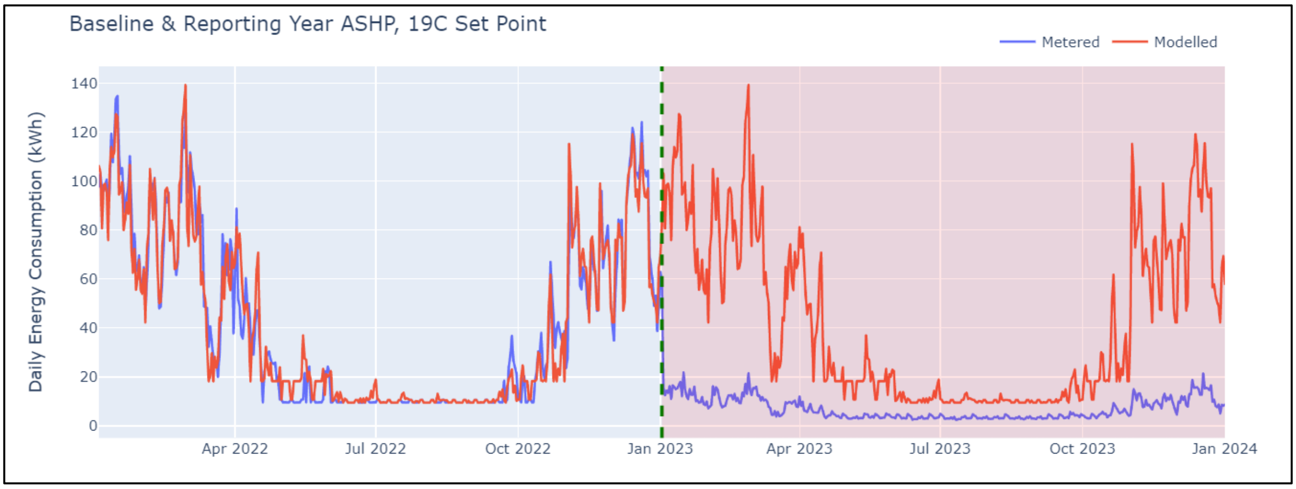

Figure 32 - Example OpenEEmeter application to a simulated property with an air source heat pump intervention.

For this reason, it is probably preferable to consider the cost and emissions equivalent of the MES, rather than just the raw energy difference. On a cost basis, domestic gas has both a lower unit rate and a lower daily standing charge – both must be taken into account for a proper comparison to be made. Taking the Ofgem price cap rates as of March 2024 of

7.42p/kWh and 29.6p/day for gas, and 28.6p/kWh and 53.4p/day for electricity, the baseline and reporting periods are transformed to a cost basis in Figure 33. The baseline period remains unchanged as both the metered and modelled consumptions are gas, but the reporting period shows that a saving is still achieved over the winter periods, albeit with a more expensive summer.

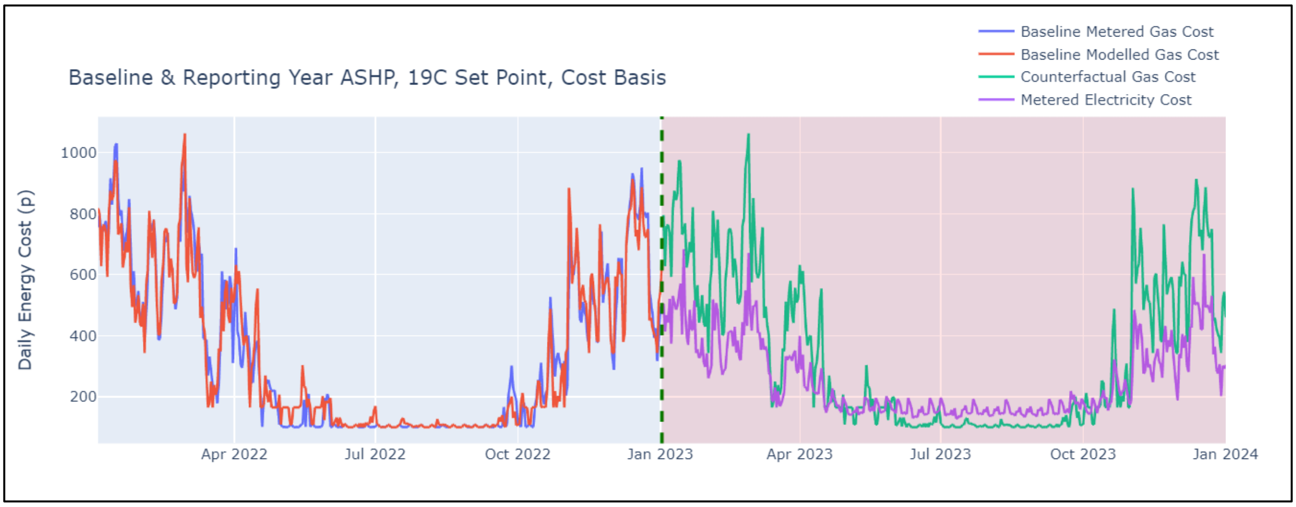

Figure 33 - AHSP intervention with baseline and reporting periods presented on a cost basis.

For the case where the heat pump and fabric retrofit interventions need to be disaggregated, a possible approach is outlined in the flow chart below. It requires some assumptions that may be difficult to justify in practice, including daily COP values for the heat pump and the average efficiency of the gas boiler.

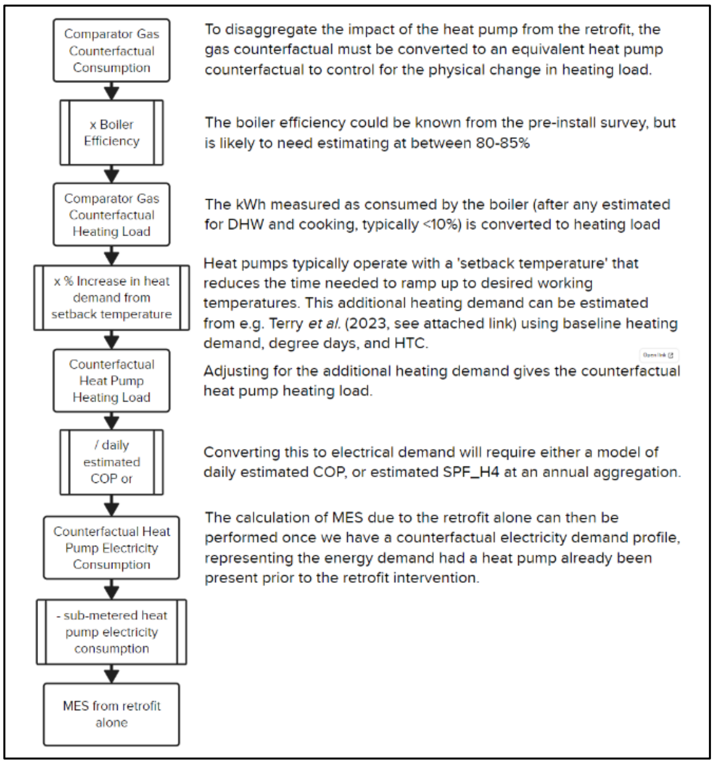

Figure 34 - Suggested process for disaggregating MES when a heat pump and fabric retrofit are both installed simultaneously.

In future work we hope to expand and test this methodology further, as the UK government hopes to meet its target of 600,000 heat pump installations a year by 2028. this use case is going to become increasingly common, and the RetroMeter approach must be suitable to unlock the funding necessary to accelerate the energy transition.

7.2 Maintaining Comparator Groups Through Evaluation Periods

Continuing on from Section 5.4, further work is needed to properly define a mechanism for maintaining up to date comparator groups that can be queried and monitored on a rolling basis to meet the needs of retrofit projects as needed. It is currently very time-consuming to access smart meter data through the DCC as an accredited 'Other User' (see here[6] for an overview of the process as of 2023) even for the most basic data requests.

The requirements of the comparator methodology are complex – not only are thousands of properties worth of consumption data needed, each with at least 12 months to satisfy the baseline year sufficiency requirements, but the consumption of those properties must also be kept up to date through the reporting period of the candidate property. In addition, there must be some way of knowing their rough location in order to match them to a weather station for external temperature readings. Running the matching algorithm requires access to a repository of all valid potential comparator properties at once – it cannot practically be run on a per-property consent basis as most Managed Service Providers (MSPs) offer currently.

One potential solution that prevents needing to extract any smart meter data from their managed, privacy-maintaining repositories is to partner with an existing DCC Other User with live and historical access to at least 3,000 properties. Instead of pulling the full dataset to a local compute instance, the matching algorithm and candidate property details are passed to the dataset provider and run on their system, with only the adjusted candidate counterfactual returned. This likely prevents the need for any additional household consent to be requested whenever a new comparator group is formed for candidate property, as no information is returned that might identify the comparison properties.

Another solution builds on the recent announcements from DNOs including UKPN and SSEN that they will be publicising aggregated smart meter readings from properties on their network. Matching the candidate property's baseline consumption profile to the aggregated readings would require normalising the magnitude of each profile, as it would not be possible to attribute the contribution of each comparator property's individual consumption to the aggregated group. If this challenge is solved, there would also need to be communication with the DNO to derive the external temperature readings for the properties, although as DNOs are small enough that their climate does not vary too much, an average across the region may suffice.

Recent publication of aggregated electricity smart meter data by DNOs might present

[6] https://www.linkedin.com/pulse/how-access-gb-smart-meter-data-matt-brake/

another source of data, but unfortunately this is only electricity data and there are no corresponding plans to publish aggregated gas smart meter data.

7.3 Further Work

One of the key goals of the RetroMeter project as a whole is to develop an open-source methodology that users in the UK retrofit ecosystem can take and test under real-world conditions to accelerate the validation and uptake of MES. The codebase developed for this Alpha phase project stage has successfully demonstrated that this is viable, but requires further work to build into a production-level software package ready for others to take and use without adaptation.

As electrical heating becomes more widespread in the UK, the methodology will also need extending to cover that.

8. Appendix

8.1 Comparator Residuals Adjustment Methodology

The candidate property's counterfactual (CF_{cand}) can be adjusted by either the absolute (res_{comp}) or percentage ($res\%_{comp}$) comparator counterfactual error. In the case where there is no intervention (H_0) we expect there to be no impact on the candidate's consumption beyond what will also affect the comparator group ($res_{comp} = res_{cand}$). Equation 3 outlines the process of removing the raw counterfactual error from the candidate property.

$$impact = res_{cand} - res_{comp}$$

$$Obs_{cand} = CF_{cand} - res_{cand}$$

$$H_0: impact = 0,$$

$$\Rightarrow res_{cand} = res_{comp}$$

$$\Rightarrow E(CF_{cand}) = CF_{cand} - res_{comp}$$

$$\Rightarrow CVRMSE\big(E(CF_{cand})\big) = 0$$

Equation 3 - Adjusting candidate residuals by the raw comparator residuals.

Adjusting by the percentage residuals was found to perform better than using the raw residuals, likely as it normalises the differences in absolute consumption between candidate and comparator properties.

$$res\% = \frac{CF - Obs}{CF}$$

$$Obs_{cand} = CF_{cand}(1 - res\%_{cand})$$

$$H_0: impact = 0$$

$$\Rightarrow res\%_{comp} = res\%_{cand}$$

$$\Rightarrow E(CF_{cand}) = CF_{cand}\big(1 - res\%_{comp}\big)$$

Equation 4 - Adjusting candidate residuals by the percentage comparator residuals.

8.2 Solar Aperture – Siviour Method

This section outlines the Siviour Method to calculate the solar aperture. The Siviour Method utilises a rearrangement of the steady-state heat equation to calculate the solar aperture (see Equation 5). The aperture is the intercept of the regression line between $\frac{P_{heat}}{\mathrm{T_{in}} - \mathrm{T_{ex}}}$ and $\frac{Irradiance}{\mathrm{T_{in}} - \mathrm{T_{ex}}}$ as shown in Figure 35 (Allinson, et al., 2022). In this case, the solar aperture is equivalent to 4 m^2. Note that while both vertical and horizontal irradiance can be used, it is important to consistently use the same type of irradiance in downstream applications (i.e. energy demand prediction).

$$\frac{P_{heat}}{\mathrm{T_{in}} - \mathrm{T_{ex}}} = -Aperture\frac{Irradiance}{\mathrm{T_{in}} - \mathrm{T_{ex}}} + HTC$$

Equation 5

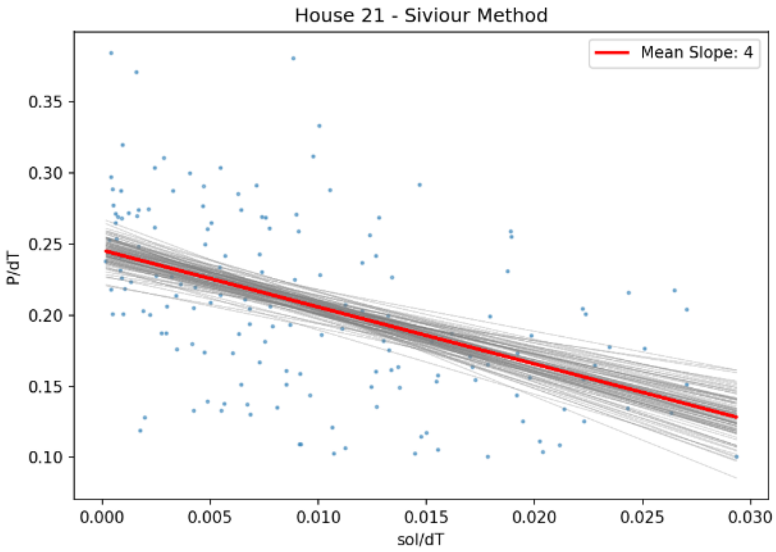

Figure 35: Example of Solar Aperture estimation as Intercept between $\frac{P_{heat}}{T_{in} - T_{ex}}$ *and* $\frac{Irradiance}{T_{in} - T_{ex}}$

9. References

Allinson, D. *et al*. (2022), *Technical Evaluation of SMETER Technologies (TEST) Project,* Department for Business, Energy, and Industrial Strategy.

CalTRACK (n.d), *The CalTRACK 2.0 Methodology*. Retrieved from CalTRACK.org: https://docs.caltrack.org/en/latest/methods.html on 2024-03-22.

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

Demand Side Analytics (2023), *Pacific Gas and Electric Company – Population NMEC Control Group Accuracy Assessment.* Available at: https://www.calmac.org/publications/PGE0476.01.pdf

OpenEEmeter (n.d.), *EEmeter ver. 4.0.0a2: tools for calculating metered energy savings.* Retrieved from https://github.com/openeemeter/eemeter on 2024-03-22.

Recurve (2023), *Comparison Groups for the COVID Era and Beyond,*

Webborn, E. and Few, J. (2022), *Smart Meter Consumption Data: Data Quality Report,* Smart Energy Research Lab.

Young, S. et al. (2022), *Metered Energy Savings: Unlocking home retrofit financing by reliably measuring energy savings* - Energy Systems Catapult. Available at: https://es.catapult.org.uk/report/metered-energy-savings

10. Licence/Disclaimer

Energy Systems Catapult (ESC) Limited Licence for SIF Alpha – RetroMeter
ESC is making this report available under the following conditions. This is intended to make the Information contained in this report available on a similar basis as under the Open Government Licence, but it is not Crown Copyright: it is owned by ESC. Under such licence, ESC is able to make the Information available under the terms of this licence. You are encouraged to Use and re-Use the Information that is available under this ESC licence freely and flexibly, with only a few conditions.

Using information under this ESC licence
Use by You of the Information indicates your acceptance of the terms and conditions below. ESC grants You a licence to Use the Information subject to the conditions below.

You are free to:

- copy, publish, distribute and transmit the Information
- adapt the Information
- exploit the Information commercially and non-commercially, for example, by combining it with other information, or by including it in your own product or application.

You must, where You do any of the above:

- acknowledge the source of the Information by including the following acknowledgement:
- "Information taken from SIF Alpha – RetroMeter, by Energy Systems Catapult"
- provide a copy of or a link to this licence
- state that the Information contains copyright information licensed under this ESC Licence.
- acquire and maintain all necessary licences from any third party needed to Use the Information.

These are important conditions of this licence and if You fail to comply with them the rights granted to You under this licence, or any similar licence granted by ESC, will end automatically.

Exemptions
This licence only covers the Information and does not cover:

- personal data in the Information
- trademarks of ESC; and
- any other intellectual property rights, including patents, trademarks, and design rights.

Non-endorsement
This licence does not grant You any right to Use the Information in a way that suggests any official status or that ESC endorses You or your Use of the Information.

Non-warranty and liability
The Information is made available for Use without charge. In downloading the Information, You accept the basis on which ESC makes it available. The Information is licensed 'as is' and ESC excludes all representations, warranties, obligations and liabilities in relation to the Information to the maximum extent permitted by law.

ESC is not liable for any errors or omissions in the Information and shall not be liable for any loss, injury or damage of any kind caused by its Use. This exclusion of liability includes, but is not limited to, any direct, indirect, special, incidental, consequential, punitive, or exemplary damages in each case such as loss of revenue, data, anticipated profits, and lost business. ESC does not guarantee the continued supply of the Information.

Governing law
This licence and any dispute or claim arising out of or in connection with it (including any noncontractual claims or disputes) shall be governed by and construed in accordance with the laws of England and Wales and the parties irrevocably submit to the non-exclusive jurisdiction of the English courts.

Definitions
In this licence, the terms below have the following meanings: 'Information' means information protected by copyright or by database right (for example, literary and artistic works, content, data and source code) offered for Use under the terms of this licence. 'ESC' means Energy Systems Catapult Limited, a company incorporated and registered in England and Wales with company number 8705784 whose registered office is at Cannon House, 7th Floor, The Priory Queensway, Birmingham, B4 6BS. 'Use' means doing any act which is restricted by copyright or database right, whether in the original medium or in any other medium, and includes without limitation distributing, copying, adapting, modifying as may be technically necessary to use it in a different mode or format. 'You' means the natural or legal person, or body of persons corporate or incorporate, acquiring rights under this licence.

Energy Systems Catapult is an independent research and technology organisation. Our mission is to accelerate Net Zero energy innovation.

Launched in 2015 by Innovate UK, the Catapult has built a team of more than 250 people, with a range of technical, engineering, consumer, commercial, incubation, digital, and policy expertise. They draw on sector-leading test facilities, modelling tools, and data collected from our back catalogue of more than 500 research projects.

We use that 'whole energy' system capability to support innovative companies -- small and large – to test, trial and scale their new products and services. Our impact comes when those innovators attract new customers, new investment, and new grants so they can thrive in the future energy system.

Based in Birmingham, Energy Systems Catapult is part of a network of nine world-leading technology and innovation centres, established by Innovate UK. The Catapult Network fosters collaboration between industry, government, research organisations, academia, and many others to transform great ideas into valuable products and services.

Energy Systems Catapult
7th Floor, Cannon House
18 Priory Queensway
Birmingham
B4 6BS

es.catapult.org.uk