WP1 – Data Access and Governance

May 2023

Document control

ESC programme name	**SIF ENW Metered Energy Savings**
ESC project number	ESC00585
Version*	1.0
Status	Approved: Contains reviewed and approved content.
Restrictions*	Open
Release date	08/06/2023
External release ID	WP1

Review and approval

	Name	Position
Author	Will Rowe Jessica Steinemann	Data Scientist Senior Data Scientist
Reviewer(s)	Sam Young	Practice Manager, Data Science & AI
Approver	Eveline Sleiman	Senior Commercial Manager - Homes

Revision history

Date	Version	Comments
08.06.2023	V1.0	Initial release

Contents

1. Executive Summary

This report focuses on the data required for developing and deploying a metered energy savings solution to measure the energy (and cost/carbon) that has actually been saved by a retrofit, what relevant datasets are available, and how they might be accessed. It also considers the practicalities and processes involved in collecting data from households that participate directly in the Beta phase.

The majority of UK properties are currently gas heated, and so development of the methodology in Alpha and Beta will focus on modelling gas rather than electricity. This limits the data required for methodology development to gas data (plus external temperature). Since gas prices are not time-of-use dependent, this also technically eliminates the need for half hourly data. However, in practice daily data is only available when a smart meter is installed.

Note that testing the accuracy of the counterfactual modelling in Alpha requires data from homes that have **not** had retrofits performed. This is because the first year of the data is used to develop the counterfactual for the second year, which is then compared to the actual energy use in the second year, with the expected difference being zero for a perfect model – because no retrofit has occurred.

The key data requirement is therefore daily gas data for 2+ years from homes that have not had retrofits performed. This must also be accompanied by external temperature data to allow for variations in weather conditions to be accounted for. This data ideally needs to be representative of UK housing stock so that we can be confident the methodology works well for all housing type and demographic mixes.

Developing and validating methodologies based on data from individual homes typically requires a sample size in the hundreds or low thousands – particularly if the aim is to validate its effectiveness across the entire UK housing stock.

However, given the comparison-based methodology proposed in WP2 (where homes are matched to similar properties that haven't undergone a retrofit to help adjust for external factors like energy price changes), much larger samples are required. This is because the accuracy of this methodology is expected to be quite sensitive to the number of homes available to select matches from. Similar work in the US typically leverages pools of smart meter data from hundreds of thousands of homes, but data at that scale is not currently available in the UK. However, we intend to test how effective those methods are when the pool of homes is 10-20k.

In addition, the development of a physics-based component of the methodology ideally requires some homes with gas smart meter data and known Heating Transfer Coefficient (HTC) (or at least gas smart meter data and internal temperature data). These are not required on the same scale, but ideally several hundred would be available from a range of housing types. However, we have been unable to identify any suitable datasets with known HTC values.

The three key data sources that have been identified as being most suitable for Alpha are:

1. **Hildebrand smart meter data** – 9-12k homes with >2 years of gas smart meter data (plus weather), accessible for decarbonisation-related research with a small fee (several thousand pounds) to cover processing costs. This will be our primary source of gas smart meter data. A small number of these

homes (~100) may also have internal temperature available and be used for developing the physics-based model.

2. **Smart Energy Research Laboratory (SERL)** - 13k homes with >2 years of gas smart meter data (plus weather and property information), accessible for public good research but with a complex and lengthy process for access (and the requirement for an academic partner). This will be our backup source of gas smart meter data.
3. **Living Lab data** – 17 homes with gas and internal temperature data. These will be our backup source for developing the physics-based model.

We have yet to identify a suitable enduring data source for comparison-based methods that would provide ongoing access to smart meter data for the purposes of comparison. Further work on this will be required as part of Alpha.

We have also produced draft data sharing agreement documents which will need adapting further in the Alpha, to be used in the Beta phase of the project. These have been produced from existing documents used by Carbon Co-op for the purpose of collecting data from their trial participants.

2. Introduction

The previous "Metered Energy Savings" (MES) work evaluated existing methods for MES including CalTRACK and SENSEI. This found that the existing methods' performance was insufficient when applied to hourly energy data, this is mainly due to the high variability of energy use at this time resolution. However, it did confirm that the daily or monthly calculations may be sufficiently accurate for certain use cases. Therefore, the initial task in this SIF discovery phase was to establish which use cases specifically require sub-daily savings calculations, and which retrofit interventions could be assessed against daily or monthly baselines. To do this, we explored the common retrofit use cases in more detail and identified five key scenarios of UK-residential retrofits which the model(s) would need to cover.

In this report we repeatedly use the concept of a counterfactual, so we take the oportunity to explain what we mean by this and why it is necessary. In this project, the counterfactual attempts to answer the question "What would the energy use have been had no retrofit been performed?". This question is hard to answer as there are many factors that impact home energy use the major one being weather because of its impact on heating load. The counterfactual can be thought of as similar to a forecast – although it is backward rather than forward looking. When performing a retrofit on a home we will collect data to allow us to modelled the counterfactual energy usage. In this case, things like metered energy use (both electricity and gas where appropriate) and weather data both before and after the retrofit has been performed.

3. WP1 – D1 - Data Requirements and Sample Size

3.1 Summary of MES Algorithm Requirements

A key assumption made in the use case exploration was that all retrofits targeted under the RetroMeter project would relate to heating rather than electricity generation (i.e. building fabric, heating system or both, but not solar PV). This has resulted in the five use cases listed below. Each associated diagram illustrates a property's gas (blue) and electricity (yellow) consumption prior to a retrofit (first column of the diagram). The retrofit interventions applied in each scenario are listed under "Retrofit Actions" against the fuel which they impact on, depending on the property's prior heating system. Underneath "Other Changes", those events which impact on energy consumption but are not related to the retrofit itself, are listed. The last column illustrates the property's energy consumption, again split by fuel, following the retrofit. In this column, the transparent boxes relating to each fuel and energy consumption type represent the counterfactuals that would have to be developed to calculate the Avoided Energy Use (AEU):

1 – Gas heated home: building fabric upgrades only

Use case one assumes that the household's electricity consumption remains stable and irrelevant to the model. As only building fabric retrofit interventions are taking place, the heating gas consumption will shrink whereas the non-heating gas consumption should remain the same. Therefore, the counterfactual needs to estimate both the heating gas consumption had the property not been retrofit, as well the non-heating gas consumption (though these will typically be modelled together). A daily or even monthly counterfactual will suffice in this case as gas usage is not Time of Use (ToU)-sensitive.

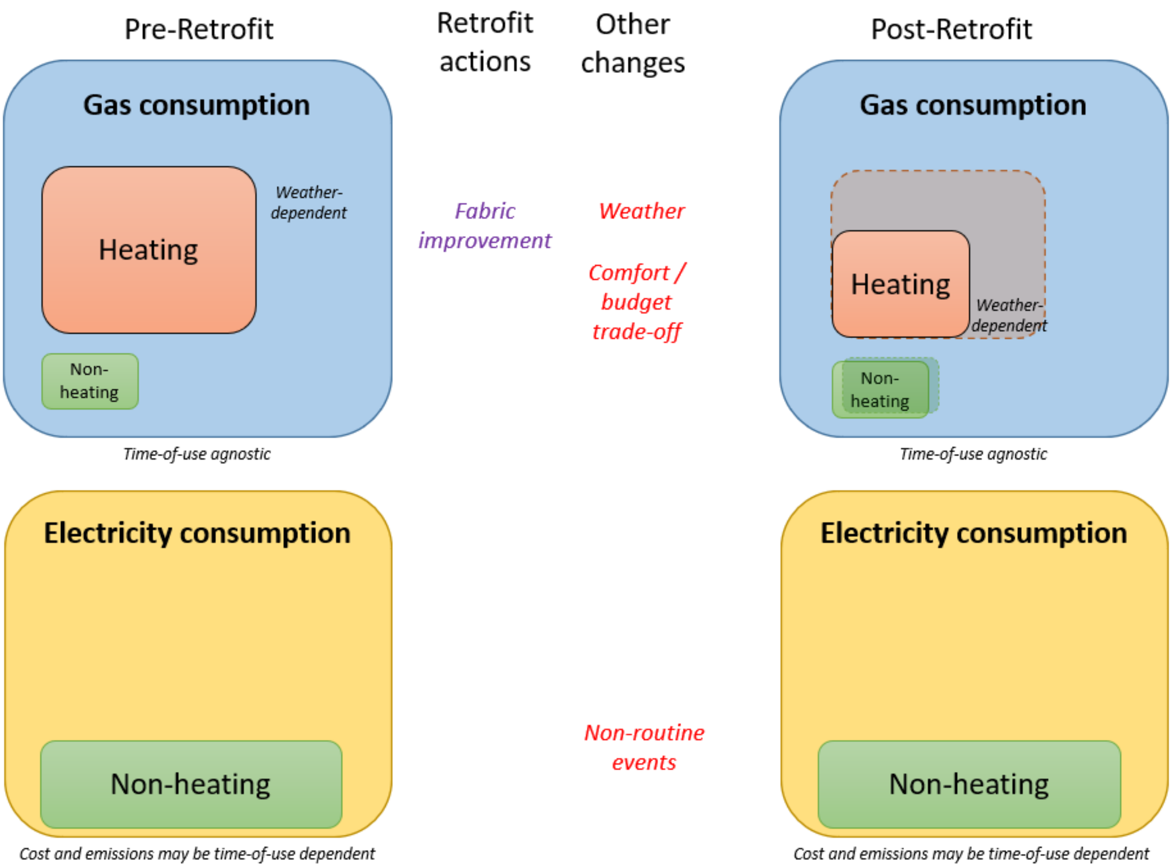

Figure 1: Counterfactuals required for gas-heated homes which receive fabric upgrades only.

In order to calculate savings in this use case we would have the counterfactual total gas usage (for heating and non-heating) and subtract the actual total gas usage (from the meter).

2 – Electrically heated home without ToU tariff: building fabric upgrades only

In use case two, gas data will not be required as all energy consumption is already electrified (or gas usage is not targeted as a retrofit, e.g. for cooking). To measure the impact of the building fabric retrofit in this scenario, two counterfactuals need to be developed:

- one for the non-heating related electricity consumption and
- one for the weather-dependent, heating electricity consumption.

Provided the electricity consumed by the household is ToU agnostic, daily or monthly counterfactuals will suffice for both types of energy consumption.

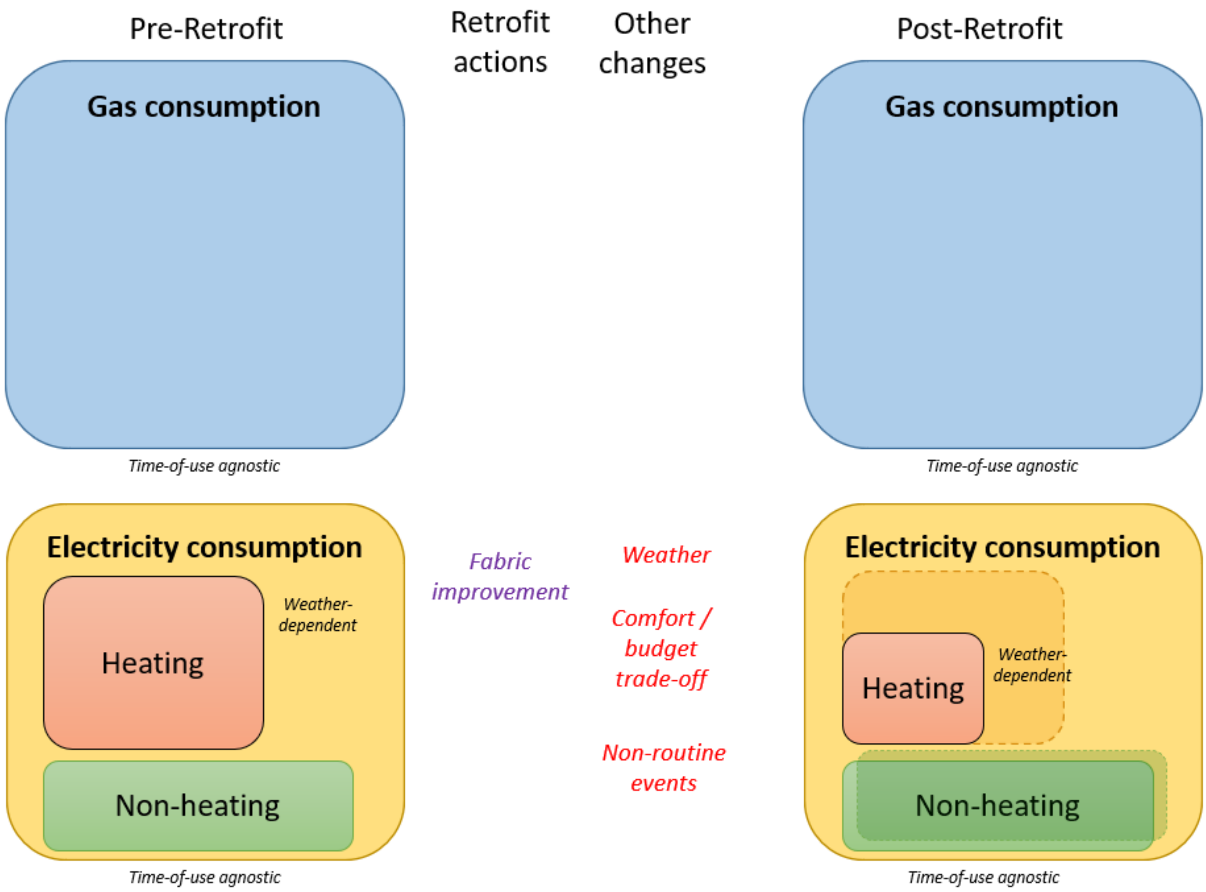

Figure 2: Counterfactuals required for electrically-heated homes which receive fabric upgrades only (ToU agnostic).

Savings for this use case would be calculated in a similar way to use case 1. Counterfactual total electricity usage (heating and non-heating) and subtract the actual total electricity usage (as metered) to give the energy saved.

3 – Electrically heated home with ToU tariff: building fabric upgrades only

Use case three is similar to use case two the only difference is that the household is on a ToU tariff. This means that the counterfactuals developed for this scenario must estimate the household's electricity consumption on a sub-daily (e.g. hourly) basis to take this into account.

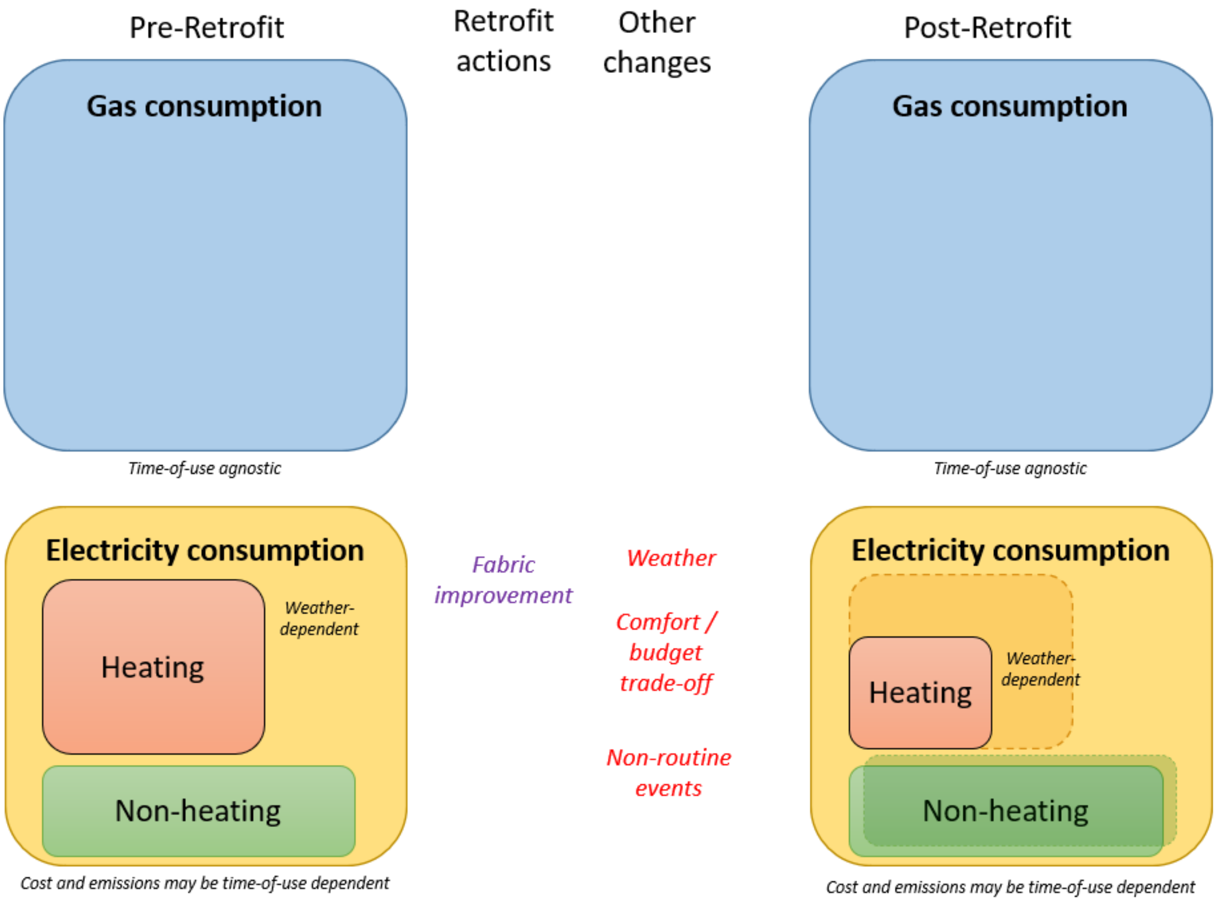

Figure 3: Counterfactuals required for electrically-heated homes which receive fabric upgrades only (ToU tariff).

Calculating savings for this use case would be the same as for use case 2.

4 – Gas heated home: building fabric upgrades combined with fuel switch to electric (ToU agnostic)

In use case four, fabric improvements are coupled with a fuel switch from gas to electricity for heating, e.g. the installation of a heat pump. To estimate the avoided energy use (AEU), a counterfactual for gas consumption must be calculated. Additionally, the non-heating electricity consumption should also have a counterfactual so that the post-retrofit heating energy consumption can be disaggregated from the non-heating energy consumption. As this scenario is ToU-agnostic, a daily or monthly counterfactual will suffice.

If submetering of the new electric heating (and DHW/cooking if that changes fuel too) was installed at the same time, the need for a non-heating electricity counterfactual would be removed as the new heating-systems' electricity consumption could easily be determined from the submetering.

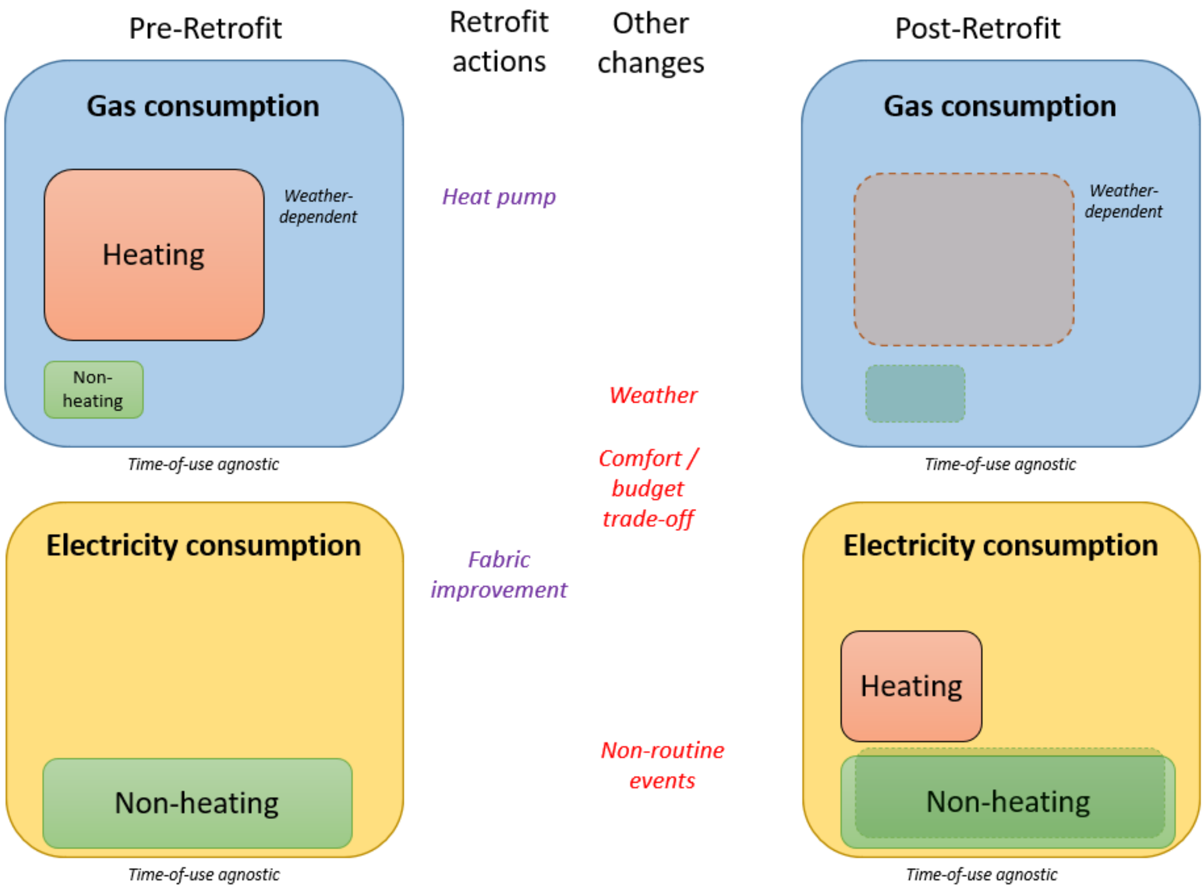

Figure 4: Counterfactuals required for gas-heated homes which receive fabric upgrades and switch to heat pump (ToU agnostic).

In this use case, savings are calculated by taking the counterfactual total gas usage (heating and non-heating combined) and subtract the sub-metered actual heating electricity usage. In this case, all quantities must be in the same units. If calculation of cost savings is required, each fuel will need to be converted to cost and then this equation can be applied. This can be done similarly for energy and carbon savings.

5 – Gas heated home: building fabric upgrades combined with fuel switch to electric (ToU tariff)

Use case five is use case four combined with ToU electricity tariffs. Therefore, the electricity counterfactual would need to be on a sub-daily or hourly basis to accurately estimate the non-heating electricity use counterfactual. As for use case four, if submetering of the new electric heating was installed at the same time, the need for a non-heating electricity counterfactual would be removed as the new heating-systems' electricity consumption could easily be determined from the submetering.

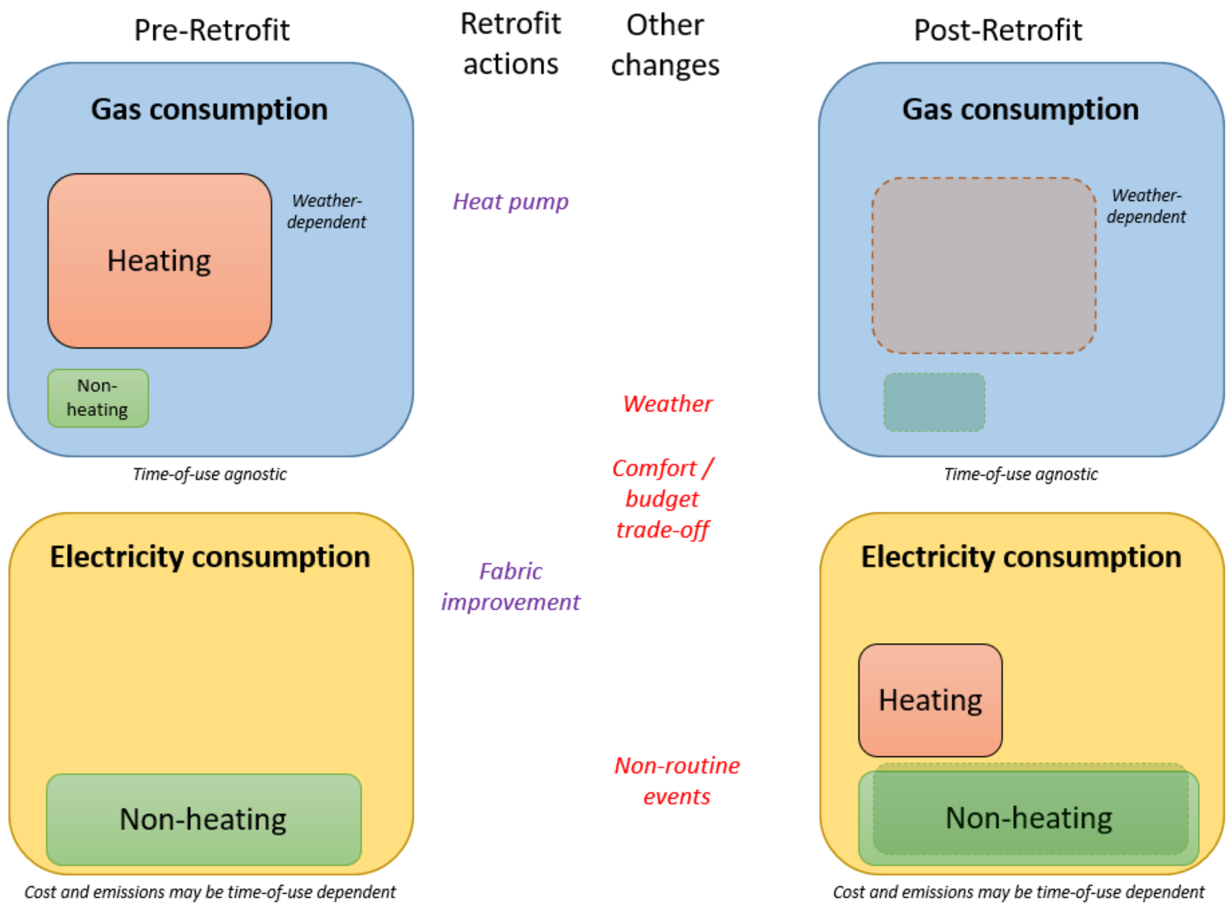

Figure 5: Counterfactuals required for gas-heated homes which receive fabric upgrades and switch to heat pump (ToU tariff).

For this use case, we can calculate savings in the same way as for use case 4.

3.2 Summary of use cases and required algorithms

In summary, the different use cases explored may require the following counterfactuals to be developed:

- Daily or monthly gas counterfactual for both heating and non-heating energy consumption combined.
- ToU agnostic: Daily or monthly electricity counterfactuals, split into:
    - Heating electricity consumption
    - Non-heating electricity consumption
- ToU aware: Sub-daily electricity counterfactuals, split into:
    - Heating electricity consumption
    - Non-heating electricity consumption

Non-heating electricity consumption counterfactuals, either on the daily, monthly or sub-daily granularity may be required for use cases four and five but could be pre-empted by the installation of submetering of the new heating system. The latter would provide a much more accurate picture of the heating and non-heating electricity consumption than an estimated counterfactual.

For clarity, the table below maps the required counterfactuals to the different use cases:

FUELS			INTERVENTIONS		REQUIRED COUNTERFACTUALS				
Use Case	**Original fuel**	**Target Fuel**	**Fabric retrofit**	**TOU**	**Gas daily/ monthly**	**Elec (heating), daily/ monthly**	**Elec (heating), sub-daily**	**Elec (other), daily/ monthly**	**Elec (other), sub-daily**
1	Gas	Gas	Y	N	Y				
2	Electric	Electric	Y	N		Y	Y		
3	Electric	Electric	Y	Y			Y		
4	Gas	Electric	Y	N	Y			Y	Y
5	Gas	Electric	Y	Y	Y				Y

Data from the English Housing Survey (for 2020 and 2022) shows, 90.2% of households had a central heating system with wet radiators in 2020 whereas 86.3% of households were connected to the mains gas supply. However, according to the Office for National Statistics, these figures vary considerably depending on local areas with four of the ten neighbourhoods most reliant on electric-only heating being located in the Greater Manchester area. Based on these statistics and further discovery with Carbon Coop it is anticipated that the majority of, if not all homes, which will be retrofitted as part of this project currently have gas central heating and will either fall under use cases one, four or five.

3.3 Data requirements

To build accurate counterfactuals for the priority use cases identified (#1, #4 and #5), we anticipate that the following data will be required:

Data type	Priority	Relevant Use Cases	Description/Details	History
Gas consumption data	Must have	1, 4, 5	Monthly or Daily. Note: as of December 2022, 50% of all households in Great Britain had smart	24 months (12 months training and 12 months testing data)

			meters installed, 13% of which operated in pre-payment mode.[1]	
Electricity consumption data	Should have	2, 3, 4, 5	Monthly, daily or half-hourly, depending on use case.	24 months (12 months training and 12 months testing data)
Energy tariff data	Could have (Alpha) Must have (Beta)	All	Tariff data at the same resolution as consumption data for both electricity and gas. Required post-retrofit in Beta for calculation of monetary savings. For the Alpha we can used assumed tariff if this data isn't available.	12 months post-retrofit
Internal temperature	Should have	All	Where metered, ideally half-hourly or hourly. Where this is not available default values could be assumed or requested from the home occupier in a survey.	24 months
Building outer postcode	Must have	All	Required to map household to external weather data. Alternatively, datasets with weather data relevant to the household could be used.	N/A
Weather data	Must have	All	External temperature but also wind speed, precipitation and solar irradiance can impact on heating demand.	24 months
Home usage data	Could have	All	Additional data about door and window opening and closing or building occupancy may be beneficial to understand where changes in behaviour are impacting energy	24 months

1 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1143890/Q4_2022_Smart_Meters_Statistics_Report.pdf

			usage (e.g. increased window opening after insulation upgrades).	
Building Fabric and Contextual Data	Could have	All	Data about the existing building type, fabric and previous retrofits might allow the algorithm to better train for different building archetypes.	At start of trial period.

Although gas usage is non-ToU dependent, we expect that it will be required to use smart gas data regardless. This is because only smart meters would record gas usage daily, or accurately on a monthly basis. However, by being able to estimate the gas counterfactual on a daily or monthly basis, we would expect better model performance as daily demand peaks and troughs identified in previous MES work will be removed from the data through aggregation (the "double penalty" effect).

For model development during the Alpha phase, homes selected should not have had any retrofits implemented during the 24-month observation period: At least 12 months of data will be used to train the algorithm ("baseline period") whereas a further 12 months of data will be used to test the performance of the MES predictions against. Homes without interventions are chosen for this phase to calculate the evaluation metrics comparing the model's predictions with the real-world data.
However, in the beta phase, the model would then be applied to homes which have been retrofitted to calculate the AEU following the retrofit.

3.4 Determination of Sample Size

Due to data privacy concerns, access to large, representative samples of daily or sub-daily smart meter data remains challenging, as discussed in previous work (Teng *et al.*, 2022). In addition to this, as highlighted by previous MES work, the datasets selected for training purposes should ideally have been collected from UK households (or households which experience similar climatic conditions to the UK and have a similar housing stock). This is a requirement as external temperature will be an important model input considering its impact on building heating demand. As a result of this, the sample size for this study will be impacted by data availability issues.

Determining an appropriate sample size for developing a new modelling approach is non-trivial and will vary greatly depending on the methodology chosen. Traditional statistical methods may be used to determine the sample size for experimental or survey-based studies. However, the smart meter and sensor-based data that will be used in this project does not fit this approach. Additionally, many modern machine-learning approaches such as the probabilistic models discussed in the literature review and methodology section require large sample sizes to train effectively and avoid overfitting. In absence of robust statistical methods for determining sample size, it is customary to evaluate related academic literature what sample sizes related studies have worked with. The following paragraphs outline some of the previous related research done in this field and the sample sizes similar studies used to develop their models.

A lot of existing research using gas smart meter data is predominantly concerned with the prediction of building thermal performance rather than energy consumption (Chambers and Oreszczyn, 2019; Brown *et al.*, 2020; Few *et al.*, 2023) or the detection of specific population characteristics (e.g. those in fuel poverty) based on energy consumption (Hurst *et al.*, 2020; Hurst, Montanez and Shone, 2020). With regards to forecasting of gas consumption data, a lot of the literature focuses on forecasts at the city (Wei *et al.*, 2019; Anagnostis, Papageorgiou and Bochtis, 2020) or neighbourhood level (Hribar *et al.*, 2019), rather than on individual properties.

When reviewing the existing literature using granular gas consumption data, only two relevant studies could be identified:

- Using data from the Smart Energy Research Lab (SERL), which collects energy data from over 13,000 households in the UK, previous research used linear mixed effects modelling to investigate the variables associated with variation in daily household electricity and gas consumption (McKenna *et al.*, 2022). They evaluated whether the addition of different levels of contextual data, or differences in sample size, led to statistically different results in their model. Focusing on pre-Covid 19 data only, they arrived at comparatively small subsamples of 1,418 and 682 homes. Although doubling the sample size in their research did increase the accuracy of their model slightly, the effects were nonetheless modest. However, this does not preclude that a considerably larger sample size (multiple thousands rather than hundreds), were this available, could deliver better modelling results.
- Others have developed a novel forecasting model ("genetic-algorithm-optimized regression wavelet neural network") for week-ahead forecasting of daily gas consumption (Hošovský *et al.*, 2021). However, their sample only covered three different buildings over a period of 4-5 years, only one of which was considered a building with some characteristics of a residential building (albeit being a hotel).

The literature using electricity smart meter data for demand and consumption forecasting is more abundant, some of which is summarised below, albeit the retrieved studies were not solely focusing on the UK:

- Further research has developed a forecasting model to predict the monthly electricity consumption of households over the coming year based on a few months of data from the previous year (Botman *et al.*, 2023). Using a sample of 3,248

households provided as part of a forecasting competition, they managed to develop a fast and simple forecasting algorithm which predicted households' monthly electricity consumption with a relative monthly error of 0.9802. The model itself consisted of three steps: clustering the data using k -means, prediction using an ensemble of forecasts based on the historical median distribution among similar households and smoothing the predictions to remove weather-dependent patterns.
- Using a sample of 500 households in Spain, researchers have developed a similar stepped approach whereby households were first clustered to identify different behaviours (Lazzari *et al.*, 2022). The distinct groups were then passed to an XGBoost model to predict day-ahead behaviour patterns and then fed to an Artificial Neural Network to forecast electricity consumption.
- When applying a hybrid CNN-LSTM for short-term individual household load forecasting, other work has selected a sample of 69 households from a dataset of over 10,000 Australian households (Alhussein, Aurangzeb and Haider, 2020). However, this selection is predominantly justified by the requirement to select households with a hot-water system and the aim to compare their research results to a similar study.

The review of the existing literature with a view to sample size definition highlights that the forecasting of individual household gas consumption is extremely limited at this stage. Evaluating research that is concerned with individual property electricity demand forecasting as a proxy, it is evident that sample size selection tends to be driven by data availability rather than thorough evaluation of model requirements with sample sizes varying from single digits to several thousands of households.

Realistically, based on research identifying suitable training datasets, we would expect to be able to gather the following sample sizes of data for each of the use cases:
Daily or monthly gas consumption (use case 1): a sample of 42 properties was already identified by the previous MES work which could be reused. If access to ESC's Living Lab data was granted for the purposes of this project, this could increase the sample size to several 100s of homes, whereas access to the SERL data would increase the data to up to 13,000 (although not all of these would be expected to have suitable data available). There is also the potential of procuring data from or partnering with a commercial data supplier, the most promising of these so far has been Hildebrand who may have data for around 15,000 homes.

Half-hourly, daily or monthly electricity consumption data: samples of several 1000s of UK households are available and should be used for developing the monthly, daily or sub-daily counterfactuals for non-heating electricity consumption (use cases 4 and 5).
For households where a split between heating and non-heating electricity consumption is required (use cases 2 and 3), we would expect this sample to be

much smaller: for example in the SERL survey, 1.98% (N=257) of respondents indicated that they had electric storage heaters, 4.23% (N=549) had electric radiators and 1.67% (N=217) had other electric heaters. As participants were able to select multiple options, it is possible that the total percentage of households with at least one of these heating types will be lower than the sum of these values.

Daily or monthly gas consumption (use cases 4 and 5 with sub metered heat pump): our research has not identified any suitable training data sets for the Alpha phase with this data; however, if RetroMeter homes were fitted with heat pumps as part of the Alpha and beta phase, we would hope that a small datasets (10s of households) could be built on which this methodology could be tested. Alternatively, it might be possible to build a synthetic dataset that combines data from the "Electrification of Heat" dataset (which contains detailed heat pump monitoring data) with gas and electricity smart meter data.

Non-linear algorithms often require larger datasets than parametric models (1000s of records) and tend to continue increasing in skill when additional training data is supplied. When developing novel algorithms where insufficient data about the required sample size is available, one common approach would be to train the model with different amounts of training data and plot its skill against sample size, i.e. to plot a learning curve, similar to the sample size comparison (Explaining daily energy demand in British housing using linked smart meter and socio-technical data in a bottom-up statistical model, 2022). By analysing the learning curve, it will be possible to identify when adding additional training data to the models leads to diminishing returns in terms of model skill.

Additionally, the existing literature suggest that it might be beneficial to apply unsupervised learning techniques to identify clusters within the energy consumption data (e.g. groups of households or specific day-type clusters). Not only would this provide further insights into the types of households which the pay for performance retrofit interventions may be applied to, but it would also enable a comparison of model performance between separate groups to check whether some groups are more suited to such a retrofit model, e.g. due to more predictable/uniform behaviour. The targeting of different types of customers with retrofit interventions has been trialed and findings suggest that energy savings could be multiplied by a factor of 2 to 4.5 depending on the type of intervention and targeting strategy (Energy Efficiency Program, 2022).

Any implementation of a comparison group-based methodology (as discussed in WP2 findings) would require access to very large datasets. This is because this type of analysis is sensitive to both the number of homes for which data is available and way in which they are grouped (which would be limited if number of homes in sample groups is insufficient). This emphasizes the need for access to data from as large a number of homes as possible.

For gas-heating data, the SERL dataset or data from a commercial supplier currently look to be the most promising sources of data for the counterfactual model development in the Alpha phase. However because of the restricted accessibility of the SERL dataset, usage in the context of the project needs to be established as soon as possible so that potential onboarding training and processes could commence as soon as feasible. Furthermore, the restricted research environment in through which it could be accessed requires further exploration as this might limit the tools available for model development and deployment. This makes working with a commercial supplier more attractive. The other datasets identified this far are likely to significantly limit the sophistication of modelling techniques that can be applied due to the small sample size.

With regards to electric heating, sufficient datasets are publicly available for the development of non-heating electricity use counterfactuals. However, most of these do not distinguish whether the household uses electric heating. Even within the largest datasets identified (SERL or commercial suppliers), the number of households with electric heating is limited which would make the development of an accurate electric heating counterfactual challenging. If use cases 2 and 3 were deemed of high importance to the project, further targeted data collection will be required to expand the available sample size.

3.5 Summary points for deliverable

Further definition of the P4P use cases will help guide the counterfactual model design and may lead to reduced complexity of the developed algorithms.

The installation of submetering and other sensors as part of the retrofit intervention(s) provides the opportunity to record more accurate data and could remove the need for certain counterfactuals.

The most complex counterfactual models are anticipated in relation to homes which currently already have electric heating systems and have a ToU tariff (e.g. storage heaters). However, these represent only a small percentage of homes within the UK. Model development for these types of homes would also be challenging in terms of the data currently available about these types of homes.

The sample size for training the monthly and daily gas-consumption models will be dictated by the availability of data and could be limited to 100s of samples. Additional methods to understand the skill of the developed model based on the available sample size would be the creation of a learning curve based on varying sample sizes or the clustering of households within the sample to compare model skill across different clusters. For use cases 2 and 3, even the largest datasets (SERL or commercial supplier) are unlikely to contain sufficient data for robust model development.

Sufficient volumes of UK electricity smart meter data should already be available for model training; however, the need for developing this counterfactual will be determined by the retrofit types targeted by the project.

4. WP1 – D2 & D5 – Dataset Access and Use Feasibility Assessment

Data requirements for the Alpha phase are summarised in section 1.3.
This data will still be required for the Beta phase, along with the same measurements from homes which have undergone a retrofit. In addition, where heating has been electrified optional sub-metering of the electrified heating should be considered.

4.1 Potential datasets

The requirements for two years of UK domestic gas or electricity usage along with local weather data significantly reduces available datasets. Many publicly available home energy datasets exist but none of these come close to providing the data required for this project.
Other datasets exist which are not publicly available but may provide some or all of the data required for this work. It is not yet clear whether access to these datasets will be possible. A summary of relevant data available from these datasets is provided here:

SERL (Smart Energy Research Lab)
This dataset consists of electricity and gas data from more than 13,000 homes at 30 minute time resolution for a duration longer than 2 years. Along with this monitoring data, SERL can also provide weather data, basic location data for each home, EPC data and occupant reported survey data.
Access requirements are very strict. To gain approval to access the data training, ethical approval and an application are all required and the project lead must be a researcher at a UK Higher Education Institute. This process is expected to take a minimum of 3 months. Access is then only through a secure research environment which could provide technical barriers to model development. More detail on the steps required to access the data is given in section 5 "WP1 – D4 - Data Access Plan".

Living Lab (run by Energy Systems Catapult)
Published data from the Living Lab is available on USMART and relates to monitoring years 2018, 2019 and 2020. The data was used as part of the MES project where it was found that only 10s of homes had the required data. Since then, the Living Lab has grown significantly with many more participants giving access to their Smart Meter data.
As of April 2023, the Living Lab has:
474 homes with electricity smart meters, of which 179 have internal temperature monitoring
343 homes with gas smart meters of which 141 have internal temperature and gas monitoring

342 homes with electricity and gas smart meters, of which 140 have internal temperature, electricity and gas monitoring

Looking at the duration of available data, we found 64 homes have at least 2 years of data from at least one smart meter, and 52 of these are homes where they have both meters connected. This number falls to 17 homes if internal temperature data is also required for at least a year. As the location of these homes is known, weather data for them can be added.

Hildebrand

During this project we spoke to Hildebrand, and they would be able to provide anonymised gas smart meter data for 9-12 thousand homes, covering a span of 2 years up until October (when the Alpha phase would start) with data for both electricity and gas. Around 60,000 homes will have 13 months of data. This is available for use with innovation projects that support decarbonisation, with a small fee (several thousand pounds) to cover the processing costs of providing the data. Their data also includes weather data which is sourced from 80 airport locations across the UK. We could obtain data from a source closer to the home if we have a partial postcode or similar location data.

In terms of additional information, local authority information can be provided for these homes and internal temperature data is available for approximately 100 homes. The contextual data that we can access includes property type, built form, and age of the property, which is categorized as pre-1930 or post-1930. There is some data on heating type, though this doesn't have 100% coverage (but we may be able to infer gas heating usage from the data). There is ongoing work to enhance and expand this contextual data offering.

Heat Transfer Coefficient (HTC) data sources

In WP2 we highlight physics-based methods as one of our selected modelling methods to take forward into the Alpha phase of the project. This type of method would involve calculating the HTC for a home using the available data. We therefore ideally require a dataset containing smart meter data and the measured HTC for some homes. There is simulated data of this type available in a dataset made available by the SMETER project. We were unable to identify anydatasets containing large scale real-world HTC measurements with accompanying smart meter data.

Other potential data sources

We have also approached the following potential commercial sources of data: n3rgy, Perse Energy, Carbon Laces, EnergieSprong. We found that there was varying data availability but Hildebrand was the only one able to offer what we needed.

In addition, we approached Sero to enquire about access to data from their Optimised Retrofit Program but did not receive a response before the end of Discovery phase.

4.2 Ethics in data use

Much of the data we require for this project is classed as personal data, and as such is protected by the UK General Data Protection Regulation (UK GDPR). In order to protect the data subjects (the people whose data we are processing) we must follow strict legal and ethical principles.

Clearly, ensuring ethical usage of consumer data is essential and the project, both for data during Alpha and Beta phase and for an enduring solution. This will therefore form a key thread throughout Alpha and Beta.

Access to the SERL dataset has strict ethical requirements because of the sensitive nature of the data involved. While the strict requirements to access the SERL data pose potential hurdles and delays to using the data, they also provide a tried and tested framework for approaching the ethical and legal difficulties of accessing such data.

Our intention is to apply for access to SERL and if we find it beneficial to access data through a commercial data source, we will use the principles applied during the SERL application process, to the commercial dataset.

5. WP1 – D3 - Data Sharing Protocol

For the Alpha phase we will need to agree data sharing agreements with the data provider (when this is finalised). If this is SERL we will need to follow their procedures for accessing the data. If this is a commercial provider we will need to agree a data sharing agreement that works for all parties. It is therefore not possible to produce such an agreement currently before we have identified our data source.

For the Beta phase we will need data sharing agreements with the trial participants which allow for collection, sharing and processing of the data such that we are able to perform the analysis. Drafts of these agreements are given below and are based on existing documents used by Carbon Co-op (as they will be running the trial and communicating with the participants). We will finalise these documents at the start for the start of the Beta phase.

5.1 Carbon Co-op Smart Meter Service - Individual Data Agreement.

The Society for the Reduction of Carbon Ltd. (trading as Carbon Co-op) requests your consent to access, store and process your energy consumption, generation, and tariff information. This will be done in line with our privacy policy: https://carbon.coop/privacy.

Carbon Co-op requests your consent to process this information in order to provide you with the following services:

- Visualise your gas and electricity consumption and tariff data and enable you to download this data in an open format.
- Processing your gas and electricity consumption data to generate insights to help you reduce your energy usage and carbon / green house gas emissions.
- Processing your electricity consumption data to produce recommendations and schedules for your use or for the purpose of automatic control and dispatch of energy smart appliances which enable you to optimise your electricity usage subject to your tariff, grid carbon intensity, and/or the requirements of any demand side response programmes you may be enrolled in.
- Share your anonymised gas and electricity consumption, generation and tariff data with project partners in the RetroMeter project. This project is piloting methods of measuring energy savings from energy efficiency interventions.

.

If you agree, the following information will be requested from your smart meter on a periodic basis and stored on our servers to be further processed for the purposes outlined above.

- Energy consumption & generation information for both gas and electricity (where available).
- Energy tariff information containing information of the price you are charged for energy consumed.
- Information identifying the energy meter (number, location, type).

.

Carbon Co-op will share anonymised data with project partners in the RetroMeter project for the development of methods for measuring savings from energy efficiency interventions. Carbon Co-op will not disclose or sell your personal data to any third-party without your specific and explicit consent.

n3rgy data limited (https://data.n3rgy.com) requests your consent to access, store and process your energy consumption, generation and tariff information. If you agree, the following information will be requested from your smart meter on a daily or more frequent basis.

- Energy consumption & generation information for both gas and electricity (where available).
- Energy tariff information containing information of the price you are charged for energy consumed.
- Information identifying the energy meter (number, location, type).

n3rgy data limited provides an energy data gathering service for many businesses. Each time a company requests your data, you will be presented with this notice and requested to give consent to that specific company . The data will be gathered every 30-minutes.

n3rgy data limited does not collect personal contact information and therefore relies on our Customers (the business you are currently consenting to) to agree with you the periods of time for which your data will be held and processed for.

n3rgy data limited provides full visibility and control of the information you have shared with these businesses via a consumer portal (https://data.n3rgy.com/consumer) where you can:

- Review which organisations you have granted access to.
- Download a copy of your data held by n3rgy data limited.
- Withdraw consent for any or all of the organisations listed.*

* if consent is withdrawn from all organisations, n3rgy data limited will also remove all of your identifiable data from our platforms.

All consumption, generation and tariff information gathered by n3rgy data limited will also be cleared of all personal identifiable data (anonymised) and pooled within a data set which is made available to all organisations who use the n3rgy data limited service. This enables us together with our customers to better understand the way we use, generate and pay for energy, with the intent to further improve our nation's energy efficiency and carbon reduction. Due to the nature of anonymisation used, it is not possible to identify your data from this wider set, and therefore to remove your contribution.

n3rgy data limited will use the Smart Energy Code (https://smartenergycodecompany.co.uk/) Party credentials and Party ID of its shareholder, N3RGY LIMITED, incorporated and registered in England and Wales with the company number 11203504 whose registered office is at 4 Ovington Drive, Fleet, United Kingdom, GU51 1DF.

Declaration of Consent

Introduction

The RetroMeter project will provide and demonstrate a consistent methodology to accurately measure the energy and cost savings of retrofit energy efficiency measures, unlocking pay-for-performance financing, increasing uptake and leading to reduced costs for consumers and additional flexible services for the DNO.

0 ***What***? RetroMeter is an innovative project aiming to enhance energy efficiency by implementing Pay-for-Performance (P4P) financing models based on metered energy savings calculations, utilizing smart meter data and external parameters. The project seeks to develop an effective and credible methodology to validate energy savings, enabling the first metered savings-financed retrofit programme in the UK.

1 ***Why***? Traditional residential retrofit programmes have struggled to validate energy savings accurately and incentivize mass adoption, leading to inefficiencies and overstated savings. RetroMeter addresses this challenge, offering a solution that could substantially reduce residential emissions, lower energy bills, stimulate private financing for retrofitting, and contribute significantly to the broader goals of decarbonisation and grid optimisation.

2 ***Where***? The RetroMeter solutions will be tested in the UK

3 ***When?*** The RetroMeter project started in April 2023 and will run until insert end date here

4 ***Who?*** RetroMeter is led by Electricity North West working with project partners Energy Systems Catapult, Energypro Ltd, Carbon Co-op and Manchester City Council. Add any new partners

More information on the project's website:

https://www.enwl.co.uk/go-net-zero/innovation/strategic-innovation-fund/retrometer/

Project participation agreement

The conditions for participating in the project are described below. By signing this form you agree with the following information.

Duration: Participants are expected to be available until the end of the project, planned for insert end date here.

Activities: Following activities are planned in the course of the project, participants agree with these activities in mutual agreement with pilot site coordinators:

- *List activities here when known*

Devices property: The devices installed in the dwelling of the pilot participants remain the property of the pilot manager cooperative and can be removed at its own discretion and at the cost of the pilot manager cooperative.

Liability: (*tbd*)

Early termination: The activities of the project are planned to end at the end of insert end date here. Any prior termination should be notified 14 days in advance to allow for the deletion of private data.

Image right: Pictures of my house and of the installed devices may be taken and used by the RetroMeter project. The picture will be used without mentioning names or locations and consent will be asked again at the time of taking photos.

O I agree

O I disagree

Private data processing agreement

RetroMeter is a research project respecting privacy according to the UK law (UK General Data Protection Regulation).

Juridical basis: Add information here

Participants rights: All research participants have the right to quit the RetroMeter experiments at any moment, have the right to leave the project and require the deletion of their data, without any consequences and have the right to be informed by the Data Controller about the type of processing their data. Also they have the right to access, correct and delete his/her data at any moment. You can also obtain information and ask for rectifying it. If you decide to exercise your rights, including the withdrawal from the project, please contact the persons below ("contact persons"). Note that data access, correction or deletion may take up to 14 days.

Privacy and confidentiality: The project partners will collect user-related data for the installation, the running and fine-tuning of the RetroMeter tools. This data includes responses you may give in questionnaires, as well as your consumption data and other appliances usage related information. Recorded data will not include any personal identification.

The consortium will comply with all national legislation relevant to the country where the data collections are taking place. To this end:

0 Personal data managed by RetroMeter partners will be pseudonymised and stored in a form which does not permit identification of users without the complementary information kept separately.
1 Data processing will be done in respect to the purposes for which the data is being collected.

Your decision to participate and therefore to give your authorization for the use and diffusion of the information provided by you is completely voluntary.

A data privacy impact assessment has been realised prior to pilot implementation, making sure that all components of the RetroMeter tools, including hardware, software, networks and people are covered by security measures.

Data access and storage: Information will be held and used on a pseudonymised basis only for the purpose of the project. Consumption data and other appliance usage-related information will be held on servers hosted in the European Union or in the UK for processing. Your raw data will be kept confidential and not disclosed to third parties. Appropriate protection will be given for personal data stored 5 years after the end of the project for auditing purposes.

More details on data categories, collection purpose, recipient and storage location is provided in the annex.

Data treatment duration: The research activities last from insert start of Beta phase here to insert end date here. After the end of the project the data will be only accessible to Carbon Co-op for follow-up and auditing purposes only, until a 5 year period has passed. After this period, the data of all non-active users will be deleted. The authorization for the use and access to this information is valid until the end of the study unless the user cancels it before.

Contact persons: Your participation in the project is voluntary and implies that you agree on your personal data treatment. Consent can be refused, and withdrawal from the project is possible at any time per email to:

- *Name Surname (email: xxxx@xxxxx, phone +xxxxxxxx)* (data controller representative 1)*;*
- *Name Surname (email: xxxx@xxxxx, phone +xxxxxxxx)* (data controller representative 2)*;*

Declaration on consent: I hereby declare my consent my data may be conveyed and documented for the above stated purpose. I confirm that my participation is voluntary. I am aware that I may withdraw my consent for data processing, and therefore withdraw from the project, at any time.

Name of the pilot participant: **Location and date:**

Signature of the pilot participant:

Annex: Data category, collection purpose, recipient and storage location

Must be updated when monitoring, data collection and storage protocols are known.

Personal data	**Purpose**	**Component** and recipients	**Storage location**
Electricity and gas consumption, generation and tariff smart meter data Internal temperature data Building fabric information EPC and SAP data Partial postcode	Establishing, improving and validating methods for measuring the impact of energy efficiency retrofits. Monitoring the good functioning of monitoring equipment.	Smart meter More info to be added here when monitoring setup is known Pilot site managers (and any technical partner which is mandated by the pilot site managers. E.g. installer has access at least once) Developers (Carbon Co-op, Energy Systems Catapult and Energypro Ltd.)	**More info to be added here when data storage solution is known.** **Local device storage**

6. WP1 – D4 - Data Access Plan

As discussed in previous sections, it has not yet been possible to finalise our data source. For this reason, access requirements for the different datasets are summarised here:

SERL

We have spoken to Imperial College academics about this. In principle this is possible in the time before Alpha phase would start.
Required steps are (not necessarily in this order, and most can be done simultaneously):

- All team members need Accredited Researcher (AR) status
  - Safe researcher training
  - AR exam
  - Await exam result
- Identify an academic partner
  - Academic partners are encouraged for access to SERL
- Get Assured Organisational Connectivity.
  - This is a requirement for using your own computer
  - Additional steps are required for home working
- Apply for University Ethics approval
- Make project application through UKDS (UK Data Service)
  - Before application, AR numbers for each team member must be provided.
  - Team members can be added after project approval.

The whole process is expected to take 3 months (at least). We will start the process in advance of knowing if we have been accepted to Alpha phase.

More detail of this process is available online here, here and here.

Commercial Data Sources

For any of the commercial data sources the process will be fairly similar and is likely to include the following steps (not necessarily in this order):

- Agreement of what data will be supplied, for what cost, when and in what form.
- Mutual preparation of data sharing agreement.
  - Will need to be checked by legal, us and an executive.
- Sign data sharing agreement.
- Procurement process for the data service.
- Design, procurement and setup of data storage /compute facilities.
- Data download (either bulk file or via API).
- Confirm deletion of data (after project)

For each of potential commercial data sources, we will add any more detail here.

Hildebrand
Access would come at a cost, and this would depend on exactly what data we wanted, for how many homes and how we accessed it. They indicated this would be on the order of a few thousands of pounds. We (ESC) already work with Hildebrand for the Living Lab, this might make procurement etc. simpler/quicker.

Other commerical data sources
Responses from other commerical data sources indicates that it wont be possible to access the required data in the short timescale dictated by the Alpha phase of the project.

Living Lab
Anonymised data is likely to be permitted. This would come as a bulk data file of anonymised data, either with weather data already or we could add the weather data using vague geographic data/ partial postcode.
In principle we may be able to have access. This may require additional consent to be given by the participants.

Sero
We have not heard back from Sero.

If we plan to use data from multiple sources, we will need to do some mapping of data between these sources. Ideally this would be done before the start of the Alpha phase in October.

8. WP1 – D6 Release Access Plan for Additional Parameters

8.1 Executive Summary

In planning the proposed pilot sites we have set out to cover each of the data points identified in across the methodology work package. In the first demonstration site we will work very closely with a small number of private householders as part of an area based scheme, combining a whole house fabric retrofit with a high concentration of in-home monitoring equipment and engagement work. In the second demonstrator we will work with a larger number of social housing tenants, as part of a fabric works scheme which will also see the replacement of gas boilers with heat pumps. The primary data collection source in the latter will be the Switchee smart thermostat and related add-ons.

Resolution specified below relates to data collection, this can be downscaled to the preferred resolution of the project team.

8.2 Demonstration 1: Area based scheme approach (private tenure)

Carbon Co-op is currently running phase 1 of an Area Based Scheme (ABS) in the Levenshulme area of Manchester. The project will provide a limited 'whole house retrofit' offer, i.e. 4-6 measures with an emphasis on external works, around £15-20K per property on 7-8 properties. The majority of houses will be 100% grant funding (fuel poor homes) with additional homes covered by a mixture of loans from Manchester City Council and direct householder payments. The homes in question are classic turn of the century, red-brick terraces, with problematic elements such as 'Manchester Windows' (a window design particular to the Manchester area).

Proposed measures

- External wall insulation
- Loft top-up
- Air tightness works
- Triple glazed windows / high performance doors
- Decentralised mechanical extract ventilation
- Insulation of solid floor external perimeter
- Advanced heating controls including smart-TRVs (adjustable Thermostatic radiator valves)
- Chimney cap and fill

8.2.1 Access to data on building fabric and Energy Efficiency measures

Detailed building fabric baseline and as built retrofit scenarios are generated by our team of retrofit advisors, using our open source Home Retrofit Planner (HRP) tool. HRP is adapted from the UK's national calculation methodology, SAP 10.2 (Standard Assessment Procedure). The tool calculates the space and water heating requirements for a home from a detailed breakdown of the building fabric: floor, walls, roof, windows etc. It uses U-values and areas to calculate building fabric heat loss rates, combined with calculated heat loss from infiltration and ventilation and heat gains from solar radiation, lighting, household appliances, cooking and occupants.

Results can either be based on standard SAP assumptions (around usage patterns) or on user survey data. Anonymised data exports of pre and post works fabric and survey based usage assessments are available for RetroMeter evaluation.

8.2.2 Data Access Agreements

Householders are aware of the pilot nature of the scheme and the requirement to gather sufficient data to demonstrate the efficacy not only of their individual home retrofit but of the area based approach. Several months prior to the commencement of works monitoring equipment will be installed and householders will enter into an agreement to retain monitoring for a period of at least one year post retrofit completion.

Users are provided with a real time dashboard (browser or phone app based), showing their home data in context and allowing heating system control. This data forms the basis of the retrofit evaluation which accompanies a householder satisfaction survey at the end of year one. In this sense the monitoring is provided for the benefit to the householder as much as the retrofit assessor.

8.2.3 Data Gathering and Storage

Carbon Co-op has developed a Home Energy Management System (HEMS), based on a fully open source software stack and currently operating in over 200 homes as part of our grid flexibility trials. This system will provide real time data gathering and storage of information within the property (per room temp/humidity, heating system activation and air quality). This will be combined with smart meter data provided by our PowerShaper service, to provide the full list of data project data requirements.

Electricity and gas usage

Measure	Metric	Resolution
Electricity consumption	x.xxxkWh	30 min
Gas consumption	x.xxxkWh	30 min

Data source

PowerShaper Monitor smart meter service - data via third party DCC user API, current connections are available using N3RGY and Hildebrand.

Data consistency

Data recorded are automatically checked for data gaps, with automated gap filling processes. Where data on the centralised smart meter data repository is inconsistant, installation of a Consumer Access Device (CAD) in the home has been shown to improve consistency. Our chosen Hildebrand CAD also makes real time data available locally for collection by our Home Energy Management System.

Data availability: Access to 13 months of historical data from date of consent (or smart meter installation date if less than 13 months). Provided as a csv file at 30 min resolution.

Data accuracy: Usage data as billed.

Heating system

Measure	**Metric**	**Resolution**
Thermostat - target temperature	x.x °C (.5 °C)	on state change
Thermostat - current temperature	x.x °C	5 min
Heating	binary	on state change
Boost	binary	on state change
Heating operation mode	-normal[Text Wrapping Break]-eco -comfort -away	on state change
Away mode target temperature	x °C	on state change
Away mode affects hot water	binary	on state change
Room x - target temperature (TRV)	x.x °C (.5 °C)	on state change
Room x - current temperature (TRV)**	x.x °C	5 min (calculated)
Room x - heat demand (TRV)	binary	on state change
Room x – open window detection (TRV)	binary	on detection

** TVR current temperature reading is based on an algorithm (to account for radiator proximity), we therefore install additional, independent room sensors for evaluation purposes (see below).

Data source

Drayton Wiser smart thermostat.

Data consistency

Data is forwarded to the HEMS unit via Zigbee, is stored on a local InfluxDB OSS server partition (providing caching in the case of loss of external data connection) and finally to our remote InfluxDB cloud installation. Any loss of data feeds to the cloud triggers an alert and remedial action is taken via the remote management interface.

Data availability: From date of installation and provided as csv files in the resolution requested by the project team.

Per Room environmental measurements

Measure	Metric	Resolution
Room x - current temperature (sensor)	x.x °C	5 min (average)
Room x - current humidity (sensor)	x.x%	5 min (average)
House temperature (across 5 rooms)	x.xx °C	Hourly (average)
House humidity (across 5 rooms)	x.xx %	Hourly (average)

Data source

Sensirion SHT3X Modules

Data consistency

Data is forwarded to the HEMS unit via Zigbee interface, is stored on a local InfluxDB OSS server partition (providing caching in the case of loss of external data connection) and finally to our remote InfluxDB cloud installation. Any loss of data feeds to the cloud triggers an alert and remedial action is taken via the remote management interface.

Data availability: From date of installation and provided as csv files in the resolution requested by the project team.

Data accuracy: Temperature sensors have a range of -20 to 105°C, with a typical accuracy of ±0.3°C at 25°C. RH 0 to 100% with a typical accuracy of ±2% over the range of 20% to 80%.

Air quality

Measure	Metric	Resolution
Carbon Dioxide (Senseair)	x ppm	5 min (average)
Particulate Matter • <1.0µm • <2.5µm • <10µm	x µg/m^3	5 min (averages)
Total volatile organic chemical	X VOC Index	5 min (average)

Data source

Aigradient Pro sensor in the main living area. This data can be provided as a csv format in the resolution requested from 5 minutes upwards from the date of installation.

Data consistency

Data is forwarded in high resolution both to Airgradient cloud servers, locally to the HEMS InfluxDB server (providing caching in the case of loss of external data connection) and finally to our remote InfluxDB cloud installation.

Data availability: from date of installation.

Data accuracy: CO2 sensor is of primary interest for both occupancy and ventilation - accuracy range ±40ppm and ±3%.

External weather condition

Measure	Metric	Resolution
Temperature	x.xx °C	Hourly average
Humidity	x %	Hourly average
Cloud coverage	x %	Hourly average
Wind speed	x m/s	Hourly average
Wind bearing	x °	Hourly average
Snow	mm	Hourly accumulation
Rain	mm	Hourly accumulation

Data source

OpenWeatherMap API based on full postcode. This data can be provided as a csv format in the resolution requested from hourly upwards.

Data consistency

OpenWeatherMap data is assembled from a range of public and commercial weather station sources, so even in the event of an outage data can be retrospectively retrieved.

Data availability: full historical data available as per project requirements.

Data accuracy: a review of nowcasting demonstrated an error below ±0.5 degrees of ground truth.

8.3 Demonstration 2: SHDF heat pump retrofit (social tenure)

As part of a Social Housing Decarbonisation Fund sponsored project Manchester City Council (MCC) are carrying out retrofit works to 1,000 of their properties in North Manchester. As part of this work MCC will be installing a minimum of 700 heat pumps, at least 100 of these units will be combined with Switchee smart thermostat reporting usage data back to MCC.

8.3.1 Proposed measures:

- External wall insulation
- Replacement of gas combi boiler with low temperature ASHP
- New ventilation system
- Smart heating controls
- Low energy lighting

8.3.2 Access to data on building fabric and Energy Efficiency measures

MCC maintain an internal asset management database, containing information about property archetype, construction date and type, number of bedrooms. In addition to this they have a log of all energy efficiency measures installed to date (covering glazing, insulation etc).

The Switchee system gathers sensor data as often as every 10 minutes transmitted via GSM to Switchee's cloud server, but data is only exportable from the landlord portal as daily averages. The data below represents the minimum available, we are in discussions with Switchee about what higher resolution data may be available via the Switchee API. While the system (as currently specified) does not record overall energy usage, it does record heating system electricity consumption, total heating period and both target and average temperatures over 24hrs (which may be sufficient for use cases 4 & 5).

Additionally, the Tier 3 service (see fig.1) allows for the incorporation of smart meter data and calculation of an HTC. This service can be appended to the existing setup, with the consent of the end user.

Electricity and gas usage

Measure	Metric	Resolution
Electricity consumption	x.xxxkWh	30 min
Gas consumption	x.xxxkWh	30 min

Heating system & internal environmental information

Measure	Metric	Resolution
Max target temperature	x °C	Daily
Average temperature	x.xxxx °C	Daily (average)
Average relative humidity	x.xxxx %	Daily (average)
Heating system usage	x.xxxx hours	Daily (sum)
Hot water daily usage	x.xxxx hours	Daily (sum)
Heating system energy consumption	kWh	Daily (sum)

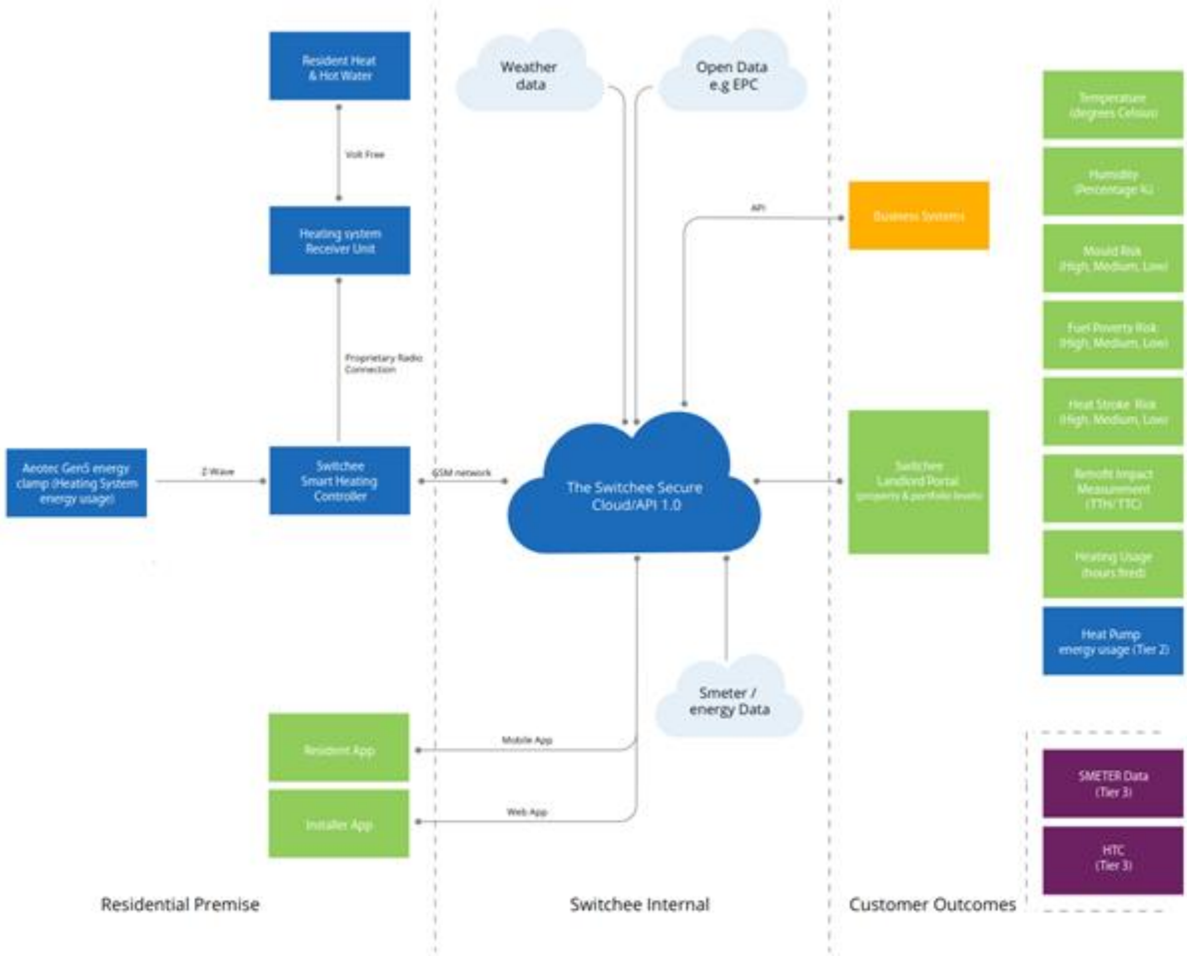

Fig.1 - showing system as specified - option 'Tier 3' meter connection is marked in purple.

9. Bibliography

Alhussein, M., Aurangzeb, K. and Haider, S.I. (2020) 'Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting', *IEEE Access*, 8, pp. 180544–180557. Available at: https://doi.org/10.1109/ACCESS.2020.3028281.

Anagnostis, A., Papageorgiou, E. and Bochtis, D. (2020) 'Application of Artificial Neural Networks for Natural Gas Consumption Forecasting', *Sustainability*, 12(16), p. 6409. Available at: https://doi.org/10.3390/su12166409.

Botman, L. *et al.* (2023) 'A Scalable Ensemble Approach to Forecast the Electricity Consumption of Households', *IEEE Transactions on Smart Grid*, 14(1), pp. 757–768. Available at: https://doi.org/10.1109/TSG.2022.3191399.

Brown, J. *et al.* (2020) 'SMITE', in *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. New York, NY, USA: ACM, pp. 21–30. Available at: https://doi.org/10.1145/3408308.3427609.

Chambers, J.D. and 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*, 183, pp. 443–453. Available at: https://doi.org/10.1016/j.enbuild.2018.11.016.

Energy Efficiency Program, C. (2022) 'Utilizing smart meter data to improve program cost-effectiveness and customer outcomes - executive summary'. Available at: https://assets.website-files.com/5cb0a177570549b5f11b9550/6406237d9d91624e275ec1aa_ComEd-Emerging-Technologies-Smart-Meter-Data Executive Summary.pdf (Accessed: 9 May 2023).

Few, J. *et al.* (2023) 'The over-prediction of energy use by EPCs in Great Britain: A comparison of EPC-modelled and metered primary energy use intensity', *Energy and Buildings*, 288, p. 113024. Available at: https://doi.org/10.1016/j.enbuild.2023.113024.

Hošovský, A. *et al.* (2021) 'Comparative study of week-ahead forecasting of daily gas consumption in buildings using regression ARMA/SARMA and genetic-algorithm-optimized regression wavelet neural network models', *Journal of Building Engineering*, 34, p. 101955. Available at: https://doi.org/10.1016/j.jobe.2020.101955.

Hribar, R. *et al.* (2019) 'A comparison of models for forecasting the residential natural gas demand of an urban area', *Energy*, 167, pp. 511–522. Available at: https://doi.org/10.1016/j.energy.2018.10.175.

Hurst, W. *et al.* (2020) 'An Ensemble Detection Model Using Multinomial Classification of Stochastic Gas Smart Meter Data to Improve Wellbeing Monitoring in Smart Cities', *IEEE Access*, 8, pp. 7877–7898. Available at: https://doi.org/10.1109/ACCESS.2020.2964885.

Hurst, W., Montanez, C.A.C. and Shone, N. (2020) 'Towards an Approach for Fuel Poverty Detection from Gas Smart Meter Data using Decision Tree Learning', in *Proceedings of the*

2020 3rd International Conference on Information Management and Management Science. New York, NY, USA: ACM, pp. 23–28. Available at: https://doi.org/10.1145/3416028.3416034.

Lazzari, F. *et al.* (2022) 'User behaviour models to forecast electricity consumption of residential customers based on smart metering data', *Energy Reports*, 8, pp. 3680–3691. Available at: https://doi.org/10.1016/j.egyr.2022.02.260.

McKenna, E. *et al.* (2022) 'Explaining daily energy demand in British housing using linked smart meter and socio-technical data in a bottom-up statistical model', *Energy and Buildings*, 258, p. 111845. Available at: https://doi.org/10.1016/j.enbuild.2022.111845.

Teng, F. *et al.* (2022) 'Balancing privacy and access to smart meter data: an Energy Futures Lab briefing paper', *64* [Preprint]. Available at: https://doi.org/10.25561/96974.

Wei, N. *et al.* (2019) 'Daily natural gas consumption forecasting via the application of a novel hybrid model', *Applied Energy*, 250, pp. 358–368. Available at: https://doi.org/10.1016/j.apenergy.2019.05.023.

10. Licence/Disclaimer

- 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
7th Floor, Cannon House
18 Priory Queensway
Birmingham
B4 6BS

es.catapult.org.uk