DEFENDER : Modelling pre-/post retrofit demand (D1.1-5)

DE08 – Model design

V1.0

carbontrust.com

+44 (0) 20 7170 7000

Published in the UK: 2022

Revision History

Name	Notes	Author	Date
0.1	Model code and results	Joshua Cooper	10 November 2022
0.2	Notes added by Carbon Trust	Laura Glover	12 January 2023
0.3	Initial review Nick Devine	Nick Devine	3 January 2023
0.4	Intermediate draft incorporating feedback	Joshua Cooper	10 February 2023
0.5	Draft review by Laura Glover and Nick Devine	Joshua Cooper	28 February 2023
1.0	Additional peak demand analysis	Joshua Cooper	8 March 2023

Authors

Name	Position	Date
Joshua Cooper	Hildebrand	10 November 2022

Approvals

Name	Position	Date
Nick Devine	NGED – Innovation Engineer	14/03/23

Table of Contents

Abbreviations

Term	Description
DFES	Distribution Future Energy Scenario
DNO	Distribution Network Operator
DNOA	Distribution Network Options Assessment
EE	Energy efficiency
EHV	Extra High Voltage
ESA	Electricity Supply Area
EUI	Energy Use Intensity
HTC	Heat Transfer Coefficient
LV	Low Voltage
MDI	Maximum Demand Indicator
sFTP	Secure File Transfer Protocol
SMETS	Smart Metering Equipment Technical Specifications
SQL	Structured Query Language

1. Executive Summary

1.1. Purpose

It is well understood that an energy transition is required to achieve net zero targets, with the electrification of heat and transportation creating new demands on electricity networks. Demand modelling has traditionally been a **“top-down”** endeavor, where now with a growing body of smart metering data a **“bottom-up”** approach is possible. A bottom-up approach seeks to estimate energy demand by starting with detailed information on individual energy-using devices and build up a picture of total demand. A byproduct of the bottom-up approach are models that generate granular **synthetic data** of future household energy demands.

Hildebrand Technology Limited has access to a large national data set of half hourly electricity and gas consumption. Novel techniques were used to create models at a household level that were then aggregated to distribution network level. The household models can be scaled all the way up to national level analysis given a list of housing stock.

The household and network models are an ensemble and fit into a larger toolset. The models and tools are used within the DEFENDER project to predict the dynamics of energy demand based on energy efficiency measures and the electrification of heat. Figure 1 shows the Hildebrand scope of delivery. Carbon Trust contributed the transition scenarios with GHD Consulting using the output of CSV exports for Sincal network asset analysis. Generally the outputs of DEFENDER could be used for future scenario planning and analysis.

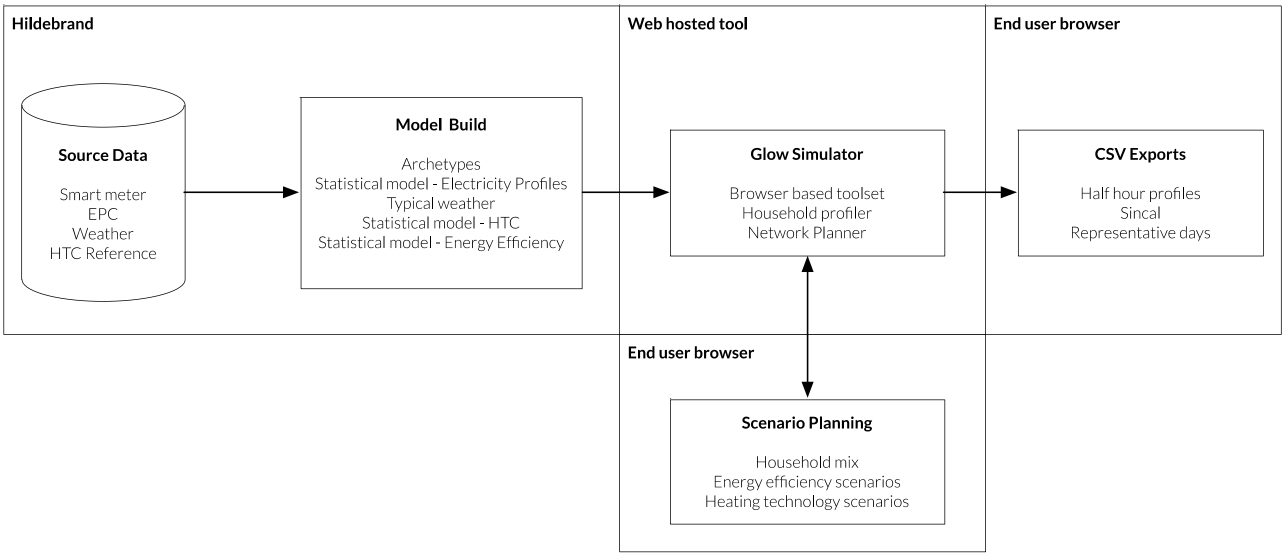


Figure 1 Overall system components with this report detailing the model build and validation from both intermediate results and end user comparison to expected results

1.2. Models

The novel aspects of the modelling are in the use of Bayesian techniques to represent the demand profiles as distribution functions. Figure 2 shows how the model elements can then generate electricity demand data from local weather conditions and the selection of a house archetype.

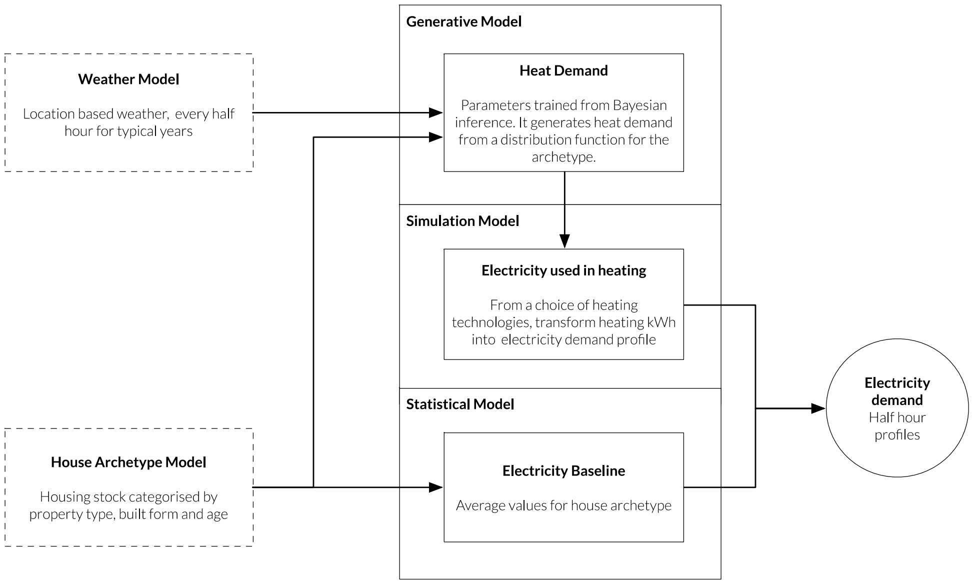


Figure 2 Details of the model elements as they are used to generate data; the data sources and model parameters are in the code running on an end user's browser

Heat demand is characterized by a heat transfer co-efficient (**HTC**), a measure of heat loss for a building. It is used to calculate the energy required to heat to a desired indoor temperature given the outside temperature.

For a given set of temperature profiles, **heating demand** is determined from simply multiplying the HTC times the difference of the indoor and outdoor temperature and the amount of time that the heating is turned on.

For every property there is a distinct HTC based on the building size and fabric features. A Bayesian methodology has been adopted that fits a probability density function of the HTC to the archetype categories. This means the HTC for a property type falls within a range treated as a random variable. That density function can be sampled to generate data with randomness equivalent to what is found in the real world to provide diversity.

Gas consumption is used to calculate the HTC. Checks on the source data showed that mid-floor flats were not sufficiently represented meaning the model will not perform well if scenarios contain a lot of mid-floor flat properties.

Comparatively the Bayesian approach outperforms a state of the art top down approach for a single household where actual performance is known.

Network planning relies on the construction of scenarios defined by a mix of housing stock. Data is generated by the models for the individual houses and then combined as a network model. The Network Planner should be able to generate realistic feeder and substation level data given the same conditions.

Network planning level outputs are formatted at the tool level to be compatible with SINCAL. The model simply generates data for each of the households, with the tool tagging, grouping and aggregating the results for feeder and substation levels. A SINCAL compatible file is simply a comma separated values (CSV) file of readings and dates, therefore is easily translated to other network modelling tools in the future.

Three locations were selected in the National Grid Electricity Distribution licence areas where dataloggers captured network level electricity demand. Housing stock for the areas was compiled from network maps and coded into the Network Planner.

Backtesting for 5 test days within a year was done with a goal of replicating the shape and volume of electricity demand under different weather conditions.

The results have shown that model estimates are higher than the real data, although the shape and dynamics of the consumption look similar between model and real data. This is likely due to bias in the training data (generally higher users) or housing stock selection (over estimating the number of occupied homes connected to a feeder.)

1.3. Future work

Next steps for the models should be to make improvements and create a housing stock database for the whole of NGED's licence areas.

Improvements would rely on training the models with a larger data set. The data set is now at least 5 times larger than the one that was used. No major issues are foreseen in running the models with this new data, however there is considerable time that would go into extracting and preparing data for a new model run.

The execution of the model to generate data is done on the web browser. Moving the model execution to the server should give more control over sampling functions and allow for more detailed Bayesian calibration with subarchetype information.

The model only considered a single day in the profile generation. A study of the effects of longer periods such as prolonged cold weather should be done to understand the effects of preceding days weather.

Electricity baseload models were done using annualised averages. This could be improved. Some models to explore would be models that link half hours as random variables, such as a Hidden Markov Model; or try and establish a Bayesian model like was done for gas, based on occupancy or appliances as the hidden variable.

2. Introduction

It is well understood that an energy transition is required to achieve net zero targets, with the electrification of heat and transportation creating new demands on electricity networks. Demand modelling has traditionally been a **“top-down”** endeavor, where now with a growing body of smart metering data a **“bottom-up”** approach is possible. A bottom-up approach seeks to estimate energy demand by starting with detailed information on individual energy-using devices and build up to estimate total demand. A byproduct of the bottom-up approach are models that generate granular **synthetic data** of future household energy demands.

Top-down energy demand modeling involves estimating the overall energy demand for a region or country by starting with macroeconomic variables such as GDP and population, and then breaking down the total demand into subcategories such as residential, commercial, and industrial use. This approach is in contrast to bottom-up modeling, which starts with detailed information on individual energy-using devices and builds up to estimate total demand.

Top-down energy demand modeling can be less accurate than bottom-up modeling. This is because it relies on macroeconomic variables, which may not fully capture the nuances of energy use in different sectors or regions. Additionally, the relationships between macroeconomic variables and energy demand can be complex and may not be fully understood or captured by the model.

Another weakness is that top-down models may not account for detailed changes in energy efficiency or technology adoption, which can have a significant impact on actual energy demand. Furthermore, it may not take into account the local characteristics of a region, such as climate, culture, and social-economic factors that may affect energy demand.

Finally, top-down models can be less flexible, as they typically require large data sets and complex algorithms to estimate energy demand. This can make them less accessible to policy makers and other stakeholders who may not have the technical expertise or resources to use and interpret the results of the model.

There are several challenges associated with bottom-up energy demand modeling. One of the main challenges is data availability and quality. In order to create a detailed model of energy use, a large amount of data is required on individual metering points and their usage patterns.

Another challenge is the complexity of the models themselves. Bottom-up models often involve a large number of variables and interactions between different energy-using devices and sectors, which can make the models difficult to understand and use. This can also make it difficult to validate the model results and ensure that they are accurate.

Synthetic data is artificially generated data. It is created using algorithms and mathematical models rather than being collected from real-world sources.

Synthetic data is often used in situations where collecting real-world data is difficult, expensive, or impossible. For example, synthetic data can be used to train models for self-driving cars, where it is difficult to collect large amounts of data on real-world driving scenarios. Additionally, synthetic data can be used to protect sensitive data by generating data that is similar to the real data but does not contain any sensitive information.

Two types of synthetic data models have been used in this project:

- **Simulation**: Creating synthetic data by simulating a process or system, such as simulating the electrical energy input required for the output of heat in different scenarios.
- **Generative models**: Creating synthetic data using generative models which learn the underlying distribution of the data and generate new samples that are similar to the real data.

Bayesian inference is a method for creating probabilistic models that can be used to generate synthetic data. It involves using Bayes' theorem, which states that the probability of a hypothesis (such as the parameters of a model) given some data is proportional to the probability of the data given the hypothesis, multiplied by the prior probability of the hypothesis.

In Bayesian inference, we start with a prior distribution over the model parameters. This prior represents our initial beliefs about the parameters before we see any data. We then use data to update our beliefs by calculating the likelihood of the data given the parameters, and using Bayes' theorem to compute the posterior distribution over the parameters. This posterior distribution represents our updated beliefs about the parameters after we have seen the data.

Once we have the posterior distribution over the parameters, we can use it to generate synthetic data that is similar to the real data. This is done by sampling from the posterior distribution, which gives us a set of parameter values that are consistent with the data. We can then use these parameter values to generate new synthetic data points by running the model with these parameters.

This approach is particularly useful when the data set is small and limited, as it allows us to infer the underlying distribution that generated the data. Furthermore, Bayesian Inference allows us to incorporate prior knowledge and subjective information into the model.

In summary, Bayesian Inference provides a way to create probabilistic models that can be used to generate synthetic data by combining prior knowledge with observed data to infer the underlying distribution of the data, and then generating new data points that are similar to the observed data.

This report shows how the Hildebrand smart metering data has been used along with Bayesian inference to parameterise building fabric in relation to heat demand, including more basic simulation models that have been used to translate heat demand into electricity consumption. And with the addition of a statistical model, measuring existing electricity consumption, producing a total electrical energy profile at a household level.

Section 3 shows how the models fit within the larger DEFENDER toolset.

Section 4 describes the archetype structure which has large implications as they are the basis for the statistical elements as well as the possible inputs to build scenarios. The rationale of the archetype selection methodology will be shown as it relates to typical UK housing stock and statistical separability between archetypes.

The next three sections (5-7) show the methodology of creating the Bayesian and energy efficiency models. Validation is presented through the comparison to published profiles and using Bayesian model metrics to test robustness.

Section 8 details the heating technology and demand profiling assumptions as well as how those are implemented in code.

Section 9 is a sanity check of the model, selecting a random home that was not used in the training data. This was done to understand if the internal model units and transformation of model parameters into energy profiles was correct. This validation is purely to test the mechanical workings of the simulator.

Section 10 is the network validation using real distribution network data to measure error rates for different scenarios.

2.1. Abstract for larger project

Work package 1.1 of Workstream 1 of the DEFENDER project develops the capability for simulating historical and future power demands, taking into account different energy efficiency measures.

Smart meter data from UK homes was used to create a building fabric model for archetypical homes. The model simulates before and after building fabric retrofit heat demand and base line electricity demand for use in future electricity network scenarios.

2.2. Related work within the project

Document D0.1 Profiling tool specification and design identified the data sources and flows to be used in the profiling tool and underpins this detailed functional and technical specification.

That document has been updated as part of this workstream and is re-issued at D1.1-2 Revised scoping document and high-level tool specification.

Background documents are useful for context including the user interface specification and the tool called Glow Simulator.

2.3. Assumptions

The following are the subset of project assumptions relevant to the technical design at the present time. Assumption status:

- Identified – The assumption has been identified but has not been validated.
- Validated – The assumption has been validated. The validation source and validation date should be stated.
- Dismissed – The assumption is not valid or is no longer relevant.

ID	Description	Status
A-1	Data coverage is good over all for the archetypes	Validated in this document
A-2	Weather stations that are used are representative of the conditions experienced at the meter point	Validated in using a wide range of weather stations, and additional functionality for uploading specific weather conditions has been added
A-3	Time period 2019 – pre-covid; largely office working. Time period 2020 - 2021 – covid: largely home working Time period 2022 - Covid new normal: 3 days a week office working	Validated during project meeting in July 2022
A-4	Data coverage between EPC and smart meter properties is good	Validated. 28/04/22
A-5	Retrofit measures are those available today, there is no forecasted change in insulation or window technology	Validated with Carbon Trust, consistent with their assumptions
A-6	Buildings fabric will not regress once a retrofit modification is made, e.g. insulation lasts for the forecast window	Validated with Carbon Trust, consistent with their assumptions
A-7	Half hourly energy (kilowatt hours) accurately represents the power demand for that period (kilowatts)	Identified, with a conversion factor being used to go from half hourly recorded energy to average kilowatts for the half hour. This is a multiplication by 2 and detailed in 5.4
A-8	Cooling is not included in this model	Validated with Carbon Trust, out of scope
A-9	Retrofit measures have been installed to best practice standards	Validated as an assumption that is made in the model
A-10	Electric vehicles have not been considered in the model	Validated 13 Dec 2022
A-11	Worst case day was from 28 February 2018 at the time of the “Beast from the East”	Validated with Carbon Trust and NGED week of 27 November 2022

2.4. Dependencies

The following are the subset of project issues relevant to the technical design at the present time. Assumption status:

- Open – The issue has been identified
- Resolved – A dependency has been met

ID	Description	Impact	Status
D-1	Model to be stand-alone (NGED IT environment restrictions)	No server required	Resolved

2.5. Risks and Issues

The following are the subset of project issues relevant to the technical design at the present time.
Assumption status:

- Open – The issue has been identified
- Resolved – A resolution has been accepted for the issue

ID	Description	Status
I-1	Diversity can be captured when generating load profiles or a at later stages before the profile is loaded into scenario analysis. Design needs to be clear when diversity is considered.	Resolved, this has been communicated to users of the model outputs
I-2	Parameters are stochastic and therefore multiple running of the tool will not always give the same result.	Resolved, this is understood as a feature of the tool
I-3	Weather assumptions for long term predictions may need to consider changing climate	Resolved, accepted as an unavoidable limitation, however new temperature profiles and assumptions can be uploaded

2.6. Design Decisions

The following table lists design decisions and their rationale.

ID	Description	Impact	Rationale
DD-1	Single page interaction for web	UI framework does not need to consider mobile interactions	Used in browser environment to export files
DD-2	Where possible easy to use selection criteria with ranges are used for describing house parameters.	May lose some fine grain control of model parameters	Complicated technical input is probably difficult for most end users of the tool, assumptions will have to be used to translate from high level combination to the reference values.
DD-3	Results downloaded to files that are stored locally	Portable; sharing must be done with the files	No server, login or user management required

3. Model and Components

Models are created from historical half hour electricity and gas source data from Hildebrand's data set. The gas profiles are transformed into distribution functions that represent energy loss for a household archetype. Heat demand is generated using those distribution functions given a set of local weather conditions. A baseline half hourly electricity profile is added to the new electrified heating half hourly profile to produce a realistic or "**synthetic data"** house level electricity demand profile.

The distribution network model is simply the summation of all of the house level profiles given a list of housing stock. The end result is a granular **"bottom-up"** data set that can be scaled from houselevel up to larger network or geographic boundaries.

The overall system can be described a set of models and a web browser based run time tool for end users. Models are built offline using machine learning methods and are published into the tool. The models provide the reference data whereby scenarios can be loaded, run and exported as data for use in distribution network analysis, policy development and cost benefit analysis of energy efficiency and electrification of heat.

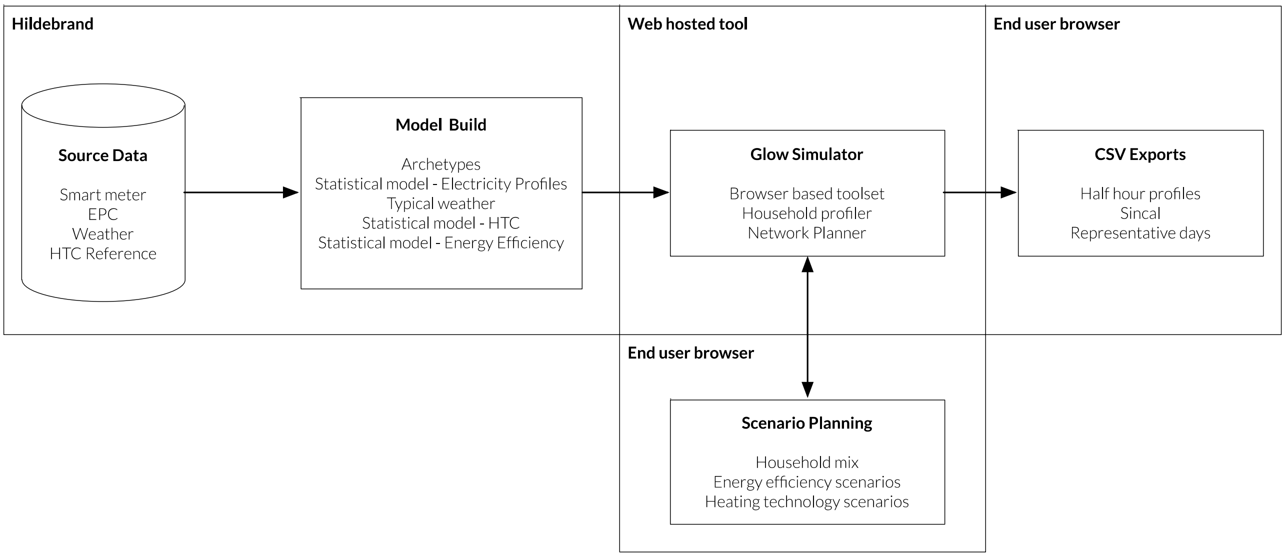


Figure 3 Overall system components with this report detailing the model build and validation from both intermediate results and end user comparison to expected results

The models and tool have been built in a modular way where future enhancements and improvements can be added without having to recreate the whole tool. The code structure follows the "bottom up" simulation paradigm where each home is simulated individually on a half hourly basis and then summed to represent the grid level view.

A number of convenient input and output formats have been accomodated, however these can be extended with developers being able to call the Glow Simulator directly, either to script and automate runs or embed directly in other tools. This report documents the interworkings of the models and the Glow Simulator whereas the end user tools are documented independently.

3.1. Source data and model build

Data collected for input into the model is from a dataset of approximately 6,200 homes from across the UK. This dataset consists of:

1. Measured energy consumption: Electricity and gas meter readings from UK SMETS meters.
2. Building properties: Descriptions of buildings from EPC for the properties being metered. This provides building size, fuel type and existence of low carbon technology at the time of assessment. Not all buildings have current EPCs and this reduced the household dataset to approximately 3,200 or a 50% reduction.
3. Weather data: An hourly weather file based on data recorded from the nearest weather station to the metering point with selection of weather location down to local authority level.

The homes in the dataset have been grouped into house archetypes based on the characteristics provided in the EPC data. For each group of houses two Bayesian calibrations were completed:

1. **HTC:** Bayesian calibration for HTC is accomplished through the selection of a prior probability distribution that we believe to represent HTCs that we would find across the building property categories. All 3,200 homes were used as input into this model. PyMC, Pandas and Numpy were used in a Jupiter Notebooks environment on a local workstation.

2. **Base electricity load profiles:** some homes were identified as having EV or night storage heaters. Processes were run to identify and remove these homes from the base electricity profiles resulting in approximately 2,800 homes being used as input. A similar approach to the HTC calibration utilising PyMC was used.

An Energy Efficiency model was developed from the Carbon Trust Options Analysis tool[1]. For all of the high level archetypes and detailed subarchetypes, a percentage change was applied to the baseline HTC of the property. These offsets were then applied when energy efficiency measures were selected. The total number of permutations that resulted was approximately 2,900. An Excel spreadsheet with reference values was translated into the code.

Weather data was coded as typical weather on a half-hourly profile of the day for each week of the year, to capture seasonality while not overloading the browser. This is to say there are 48 x 53 data points for each of 80 weather stations or 203,520 weather individual weather data points built into the tool.

3.2. Glow Simulator

The Glow Simulator is a package that generates energy profiles at the household level based on the statistical models of historic electrical energy consumption (Electricity Baseline) and also includes models for heating technology and heat demand profile. Heat demand is determined from weather and house archetype with that heat demand being used to generate equivalent electricity demand as shown in Figure 4.

[1] The Carbon Trust Options Analysis tool is a proprietry tool for conducting building assessments. It is built using published standards, including BS EN 12381 for heat loss, the most relevant assumption for use in the DEFENDER project. Further examples of its inputs and usage can be found at https://www.lambeth.gov.uk/environmental-services/climate-change-impact-plans/heat-decarbonisation-study/building-level-options-costs

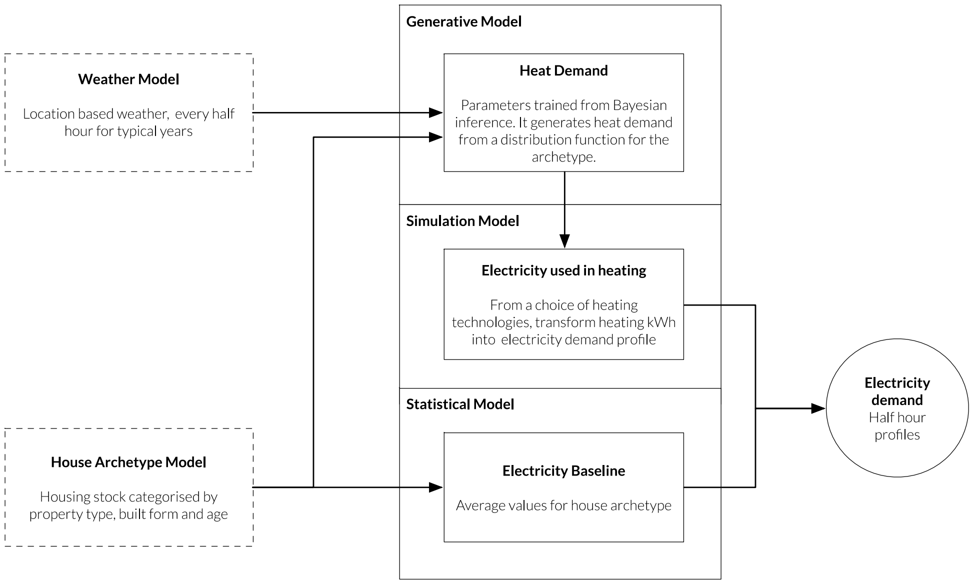


Figure 4 Details of the model elements as they are used to generate data; the data sources and model parameters are in the code running on an end user's browser

The separation of heat demand from appliance electrical loads means energy efficiency interventions can be applied to change heat demand. Likewise different heating technologies can be applied to meet that demand before then being combined with the baseline. This design of parameters and layers makes a wide range of scenarios possible to construct.

Weather input is done by selecting one of the local authorities, a data lookup is done to load the nearest weather station and from the time range selected a profile for a typical day in the season is used to generate heating load. There is a special selector for Week 4 of February that selects the historical weather from the "Beast from the East" for each weather station. There is also the facility to input a half hourly temperature profile without the selection of a preset weather station and run the simulator.

The historical "Beast from the East" of February 28, 2018 was chosen as it represents a one in 20 year prolonged cold period over a large section of the UK. Taking the historical data was simpler than forcing an artificial low temperature over the 80 weather stations that data was available from.

3.3. End user interfaces

The Glow Simulator was implemented in Javascript that runs within a standard web browser with a web based user interface. The Household Profiler and Network Planner both use the underlying Glow Simulator package to calculate energy profiles. The Household Profiler and Network Planner can output various comma separated value (CSV) file formats as well as save and load scenarios from an end user's local computer.

4. House Archetypes

House archetypes are core to the modelling approach. They are used to classify properties and the retrofit options available. Archetypes are a combination of first level categories and second level fabric features.

First level categories are defined by 3 fields (Property Type, Built Form and Construction Age Band) with a constrained set of options to select under each field. There are a total of 12 permutations or categories that can be uniquely selected. These are more or less permenant features of a property that can not change.

The second level considers fabric features of floor, roof, walls and windows. The number of permutations grows significantly and when combined with the first level categories expands to 2,904 distinct house archetypes.

EPC data at a license level was transformed with each record coded into a household archetype. The most common fabric features were catalogued by property category so that they could be used if an EPC was missing for a house.

Archetypes are classification models that represent the housing stock in the UK. An identifier is assigned to a selection of parameters representing the fabric characteristics. Characteristics can be obtained from energy performance certificates (EPCs) and then a matching archetype can be assigned to the house.

Archetypes reduce the dimensions of the overall model while still being expressive of size/shape and insultation qualities of the building. There are high level archetypes that describe fixed construction elements and sub-archetypes defined by fabric features that can be changed. Reconfiguration of the house sub-archetype features as model parameters is useful in what-if scenario building e.g. what if I change the windows in my house.

4.1. Archetypes – Modeling Demand

The high level archetypes provide good segmentation of heat load based on an assessment done by Carbon Trust using EPC analysis and SAP tools. A spreadsheet based analysis found the most common property types in the National Grid service area with average values taken for floor, roof, wall and window areas. The 68 high level archetypes and their rate of occurance is found in Section 12. A total of 861,872 properties that have current and completed EPC ratings for distribution network was used as input.

A Smart Meter Grouping label was applied to these 68 high level archetypes to group them based on the unique combinations for the three attributes that were used to key the Bayesian statistics. These are 12 labels, A-L, indicating the selection of Property type, Built form and Construction age band. The following choices can be made in each category:

- **Property type** – House or Flat
- **Construction age band** – pre 1930s or post 1930s construction
- **Built form**, if House – Detached, Semi-detached or Mid-terraced
- **Built form**, if Flat – Bottom floor, Mid floor, Top floor

Side Benefit

Although these are property categories, they also naturally correlate with some “behavioural” characteristics. “Behavioural” characteristics is a general label for the non-fabric related elements like household make up, lifestyle and appliances. The correlation isn't absolutely clear, but as a for instance, detached homes are usually larger than flats and therefore probably have more occupants. This makes the categories somewhat self organising in terms of electricity consumption.

The self organising characteristic means that these categories should provide good segmentation criteria for baseload electricity. When calculating mean electricity consumption for say a Semi-detached House, it will have good statistical separation from a Top floor Flat, that is to say as the mean value for a Semi-detached House is more statistically similar to other Semi-detached Houses than it is to Top floor flats. While the Smart Meter Grouping categories may not be the best indicator of electricity consumption, they are certainly practical as there is no need to gather additional data about each of the houses that would be used in the statistical calculations.

The 12 Smart Meter Grouping labels are fairly easily taken from EPC records. Where there are Property types and Built forms that do not fit the options, such as Maisonette, there is a translation to the best fit. The key point is that all of the data required for the Smart Meter Grouping is sourced from EPC records.

The second level of categorisation is floor, roof and wall situations that are fairly fixed for the property. For instance, cavity wall versus solid wall. For each of the 12 categories the floor, roof and wall possibilities were catalogued with assumptions made for Flats that to define the floor and roof situation there could be other premises above or below. In total there are then 68 possible high level archetypes.

4.2. Dominant Fabric Features

Dominant fabric feature analysis is the count of the most frequently found fabric features given each of the 68 high level archetypes. This analysis serves two purposes:

1. In running scenarios for a geographic or network area, if a house within the scenario does not have a complete EPC, the most common or average values will be taken from the dominant fabric features as a “best fit”
2. The actual consumption data will be assumed to come from these dominant types as the Smart Meter Groupings are only considering the 12 high level categories. This is important in establishing the base assumptions of insulation quality and size of properties such that changes in the insulation quality i.e. energy efficiency interventions can be applied as offsets.

All of the EPC records for the whole of the National Grid licence areas were transformed or cleaned to fit the semantics shown in Table 1. The dominant fabric features for each of the 68 archetypes were then catalogued. Where there are choices, such as Insulated, Partially insulated, No insulation – the most frequent occurence was taken as the dominant feature. Where numeric values are used, a mean value was calculated.

Characteristic	Choices
Dominant wall insulation	Insulated, Partially insulated, No insulation
Dominant roof insulation	Insulated, Partially insulated, No insulation
Dominant floor insulation	Insulated, Partially insulated, No insulation
Dominant window insulation	Single, Double, Triple Glazing
Average floor area (m^2)	Numeric to 2 decimal places
Average floor height (m)	Numeric to 2 decimal places
Glazed area (%)	Numeric to 2 decimal places
Average no. habitable rooms	Numeric to 2 decimal places
Average number of floors	Numeric to 2 decimal places
Front/back exposed walls	Numeric integer values
Exposed side walls	Numeric integer values
Total no. exposed walls	Numeric integer values
Building shape	Wide, Long, Square, Very Wide, Very Long
Exposed wall area (m^2)	Numeric to 2 decimal places
Glazed area (m^2)	Numeric to 2 decimal places
Exposed floor area (m^2)	Numeric to 2 decimal places
Exposed roof area (m^2)	Numeric to 2 decimal places
Airtightness factor (ACH)	Numeric to 3 decimal places
Thermal Bridge factor	Numeric to 1 decimal places

Table 1 Dominant features captured for each of the 68 archetypes, numeric values are averages from EPC or mode values for integers

The dominant feature list adds 19 columns to each of the 68 archetypes, providing the reference values that can be used to relate to measured performance. An example use case is taking the measured consumption of a Pre-1930s Semi-detached House, say an average of 50 kWh of gas per day and then assuming that is representative of a house that has no wall insulation, insulated roof, no floor insulation and double glazing, etc because that is what the dominant values are for that type of home.

4.3. Sub-archetypes - Modeling Energy Efficiency

Using the same feature set that was used for the dominant analysis in Table 1, sub-archetypes are created. This is essentially using the “choices” column and putting the numeric values into ranges or fitting them into categories. The effect is to take the 68 archetypes and expand those into the possible fabric features that could be chosen to apply to a property.

The sub-archetypes are made based on the characteristics that were shown to have the additional impact on energy efficiency, while keeping a managable level of options. Not all options apply for Houses and Flats, but the options are:

- **Wall type** – Solid wall, cavity wall
- **Wall insulation** – Insulated, Partially insulated, No insulation
- **Floor type** – Suspended, solid
- **Floor insulation** - Insulated, Partially insulated, No insulation, [Flats] Other premises below
- **Roof type** – Flat, pitched
- **Roof insulation** - Insulated, Partially insulated, No insulation, [Flats] Other premises above
- **Window glazing** – Single, double, triple glazing

From the 68 archetypes there are up to 54 sub-archetypes. The total number of valid combinations are 2,904. Some of the combinations are not found, for instance having single glazing and full insulation in walls, roof and floor is very unlikely; in looking at EPC records 1,017 combinations were found. This is not to say that more combinations would not emerge as homes are retrofitted in the future.

4.4. Heating technology

The heating technologies that are modelled in the cost benefit analysis of Work Package 1.5 are the ones that were considered. They are listed here to provide a reference of what can be selected in the tool for scenarios.

- **Fossil fuels** – the tool can model the heating required, but the electicity demand due to heat is set to zero;
- **Direct electric** – this is assumed to be direct electric raditors rather than electric boiler or anything with storage or losses
- **Night storage heaters** – Traditional night storage heaters that will charge overnight and release heat through a mechanically controlled vent; ceramic/brick heat storage that will disapate within a day
- **Air source heat pumps (ASHP)** – No particular size or storage tank, should be sized

Note: Ground source and electric boilers are user interface selections in the household profile tool, however from a model perspective they map to air source heat pump and direct electric respectively.

5. Electricity baseload

To represent the total electricity demand for a property, the non-heating related electricity half hourly profile is added to the new electricity demand profile due to the electrification of heat. Rather than taking an average profile for all homes, distributions have been calculated for each of the household archetypes for each half hour of a day as an independent random variable. Those normal distributions are sampled to mimic diversity, generating realistic profiles.

Smart meter data has been analysed to determine if an electric vehicle or night storage heater is present for a property and those properties removed from the calculation of archetype statistics.

A typical electricity profile for a house will vary depending on a number of factors, including the size of the house, the number of occupants, the appliances and devices being used, and the time of day. There is an assumption supported by visual analysis of average daily electricity use by archetype that archetypes provide some segmentation of the electricity, including slight differences in peak usage hours.

The term "electricity baseload" is defined here as the non-heating related loads for electricity. The goal is to both segment the electricity consumption on common criteria that will be used for driving the heat demand and decouple the electricity demand from any heating loads that might be found historically. This is to avoid double counting consumption when the new electrified heating energy is added and reducing the number of parameters that the end user would have to specify.

The assumption is that each half hour period is an independent random variable, meaning that a mean and variance is calculated for every half hour of the year for each archetype. This is a total of 17,520 distributions for each of the 12 permutations of house archetype.

A sample baseload can be generated from by selecting:

1. A house archetype,
2. Month (January – December)
3. Day of week (Monday – Sunday) and
4. Half hour of the day (1-48)

Resulting in a mean and variance for that half hour to be used as parameters into a normal distribution and randomly sampled to generate a data point. A full day profile is generated by stepping through the half hours of the day and generating a new sample.

The sample value will contain the diversity of fabric features (sub-archetypes) and lifestyle found in the source data, which may or may not be biased. Note this is different from the Bayesian distribution fitting of source data. Here the traditional frequentist method was used as the project did not have enough scope to establish a constraint function for electricity consumption that could be used for Bayesian calibration.

This section shows how the source data was transformed into the statistical distributions.

5.1. Source data

While the primary source data is from approximately 5,140 homes from across the UK, several criteria were applied as filters before the data was used in baseload calculations. If households met the criteria they were included, if not then they were removed. The criteria for inclusion was:

1. Having an EPC, this was to ensure that a household could be placed in an archetype category

2. Having both gas and electricity consumption data available to have more likelihood the property was heated with gas
3. Not having night storage or electric vehicles as they dramatically skew results

After applying the filter criteria, approximately 2,900 housholds remained.

In order to easily load and run aggregations by archetype, each half hourly electricity reading was tagged with the archetype information of the house. This results in a large file, but makes the group calculations easier within the Python analytic tool. The resulting number of unique half hourly electricity readings was 57,131,035 and just under 10GB as a comma separated value file.

The saved file, called filtered_elec_data.csv, has a head row and first entry example data as:

```
rowid,date,veid,ownerid,fuel,value,epc,epc_potential,property_type,built_form,local_authority,month,
day,dow,minutes,timeofday,dateonly
0,2017-11-13 10:30:00,ffccd4c4-be4b-4c64-8fc1-1ae66a76867b,64a74654-8e36-42b7-a4a4-
36fb31e99e09,0,57,E,D,House,Semi-Detached,E06000023,11,13,0,30,10:30:00,2017-11-13
```

Figure 5 Heading and single sample rows of the source data used for the creation of electricity baseload statistics

To inspect data coverage in the file, a count of readings per day was plotted on a calendar. As can be seen in Figure 6 most of the data is from 2021 and the early part of 2022. A further sanity check of the geographic spread of properties was done by geo-coding the postcode and plotting it on a map of Great Britain, shown in Figure 7. Although no formal metric was used, the data appears to have fairly good spread.

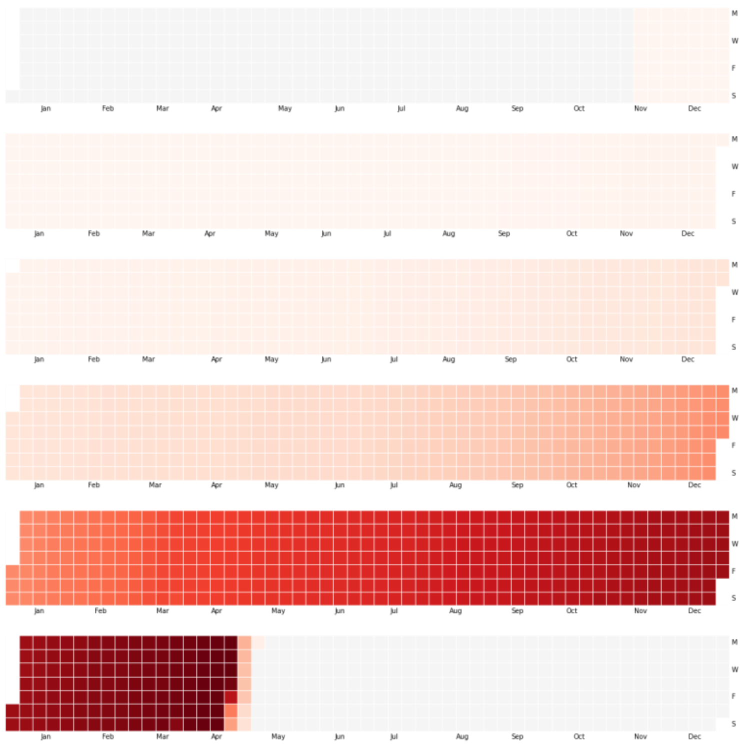


Figure 6 Dates covered by the source data set; darker red indicates more data

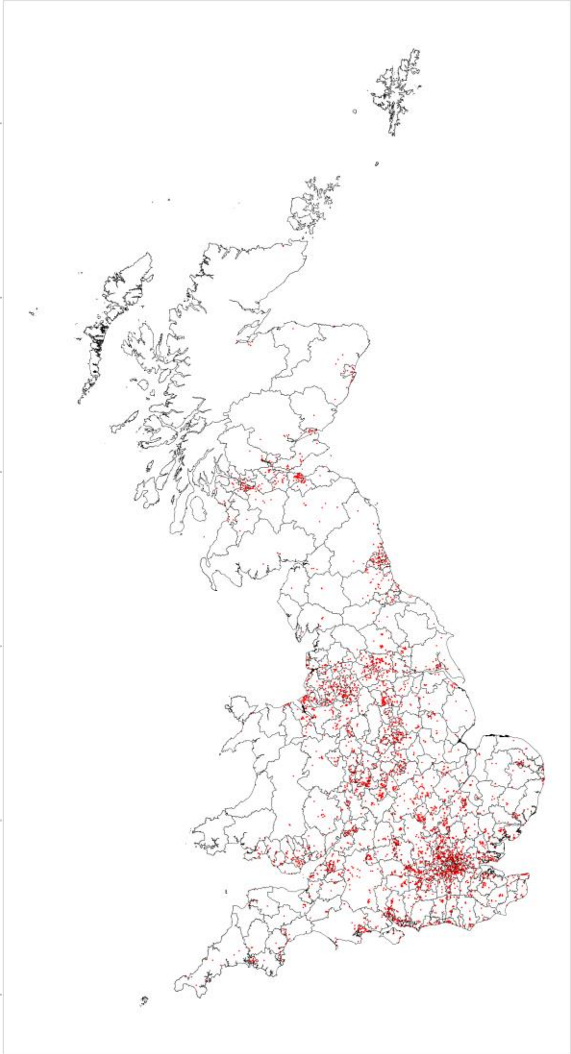


Figure 7 Geographic coverage of the source data set, with a spread across Great Britain.

5.2. Electric vehicle and night storage heater identification

Electric vehicle (EV) and night storage heaters have a distinct usage signature, with very large and disproportionate electricity demand overnight. This is due to low cost electricity at those times. As stated above, EV and storage heater properties were filtered out of the data set to avoid skewing the dataset. This section shows the method that was used to identify properties that has EV or night storage heaters.

The assumption was that overnight electricity demand would be disproportionately high as compared to the to day time usage. For each property, half hourly data was grouped as nighttime or daytime, and for each day a ratio of nighttime to total usage was calculated. Nighttime was defined as midnight to 4am, and to simplify, UTC time was used rather than transforming to local time. Furthermore, each household's average percentage of nighttime use was taken, resulting in a list of properties and their average use over a period.

To understand the appropriate cutoff value that would determine high use, a histogram of properties and their average overnight percentage was done. Figure 8 shows that most homes are below 30% nighttime consumption. To generally confirm that 30% is a good cut off, a few homes that were above and below the 30% threshold were inspected. Appendix B shows three examples to illustrate differences seen in overnight usage and how those patterns relate to EV and night storage heater detection.

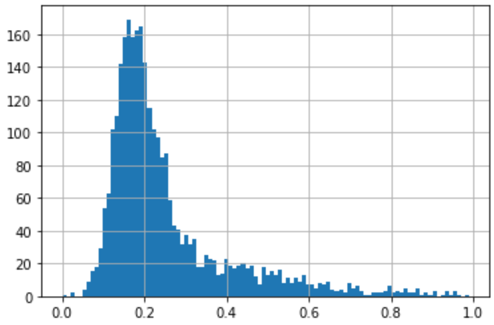


Figure 8 Histogram of ratio night to total electricity consumption; at approximately 30% there is a new cluster of properties

From the 30% cut off, a list of 648 properties were removed from analysis. Before and after average daily curves are shown in Figure 9 and Figure 10 showing the large overnight features removed.

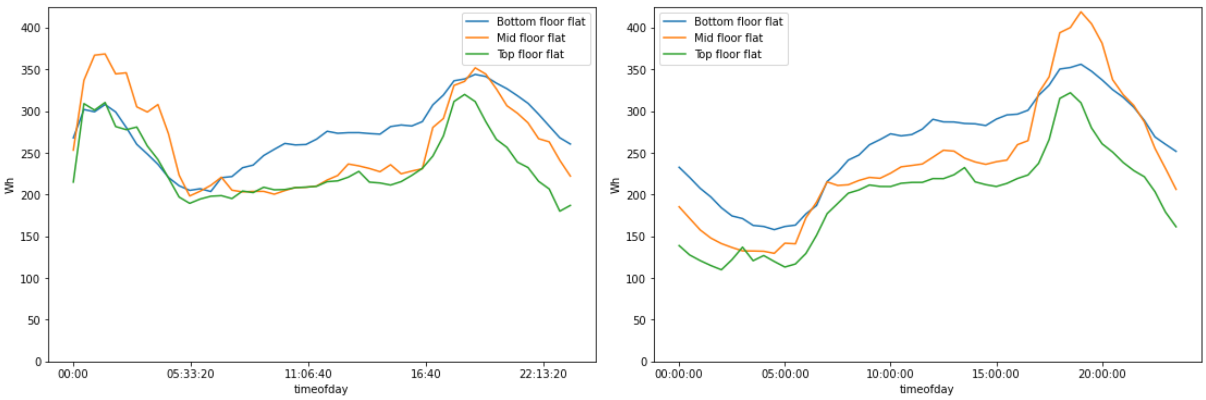


Figure 9 Flats before (left) and after (right) removal of EV and storage heater properties

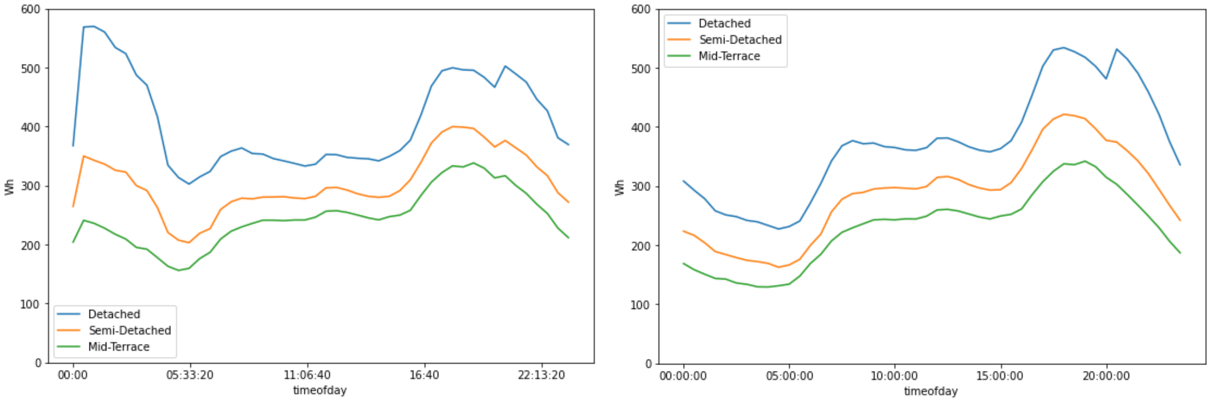


Figure 10 Houses before (left) and after (right) removal of EV and storage heater properties

5.3. Baseload statistics

From the remaining properties, detailed 30 minute mean and standard deviation statistical measures were generated for each Property Type, Built Form, day of the week and week of the year for all years. This is a total of 2 (Property types) x 3 (Built forms) x 53 weeks x 7 days of the week x 48 half hours in the day or 106,848 data points per statistical measure.

```
year_house_semidetached = pd.pivot_table(elec_df[(elec_df['minutes'] == 0) | (elec_df['minutes'] ==
30) & (elec_df['property_type'] == "House") & (elec_df['built_form'] == "Semi-Detached")],
index="woy", columns=["dow", "hh"], values="value", aggfunc=(np.mean, np.std))

year_house_semidetached.to_json("reference_house_semidetached.json")
```

Figure 11 Example baseload pivot table calculation of mean and standard deviation for a Semi-detached house with indexes on week of year (1-53), day of week (Mon-Sun) and half hour slot (0-47)

Each statistical measure is taken as an independent event and is used in a Gaussian sampling function to generate data.

For instance, given a Semi-detached house, analysis for Monday in the first week of January the first half hour of data for that day is generated by sampling a Gaussian function with the mean and standard deviation taken from those indexes.

An example from the Household Profiler shows the selected parameters and the resulting electricity consumption values in the graph. Fossil fuels has been selected to only show the electricity baseload.

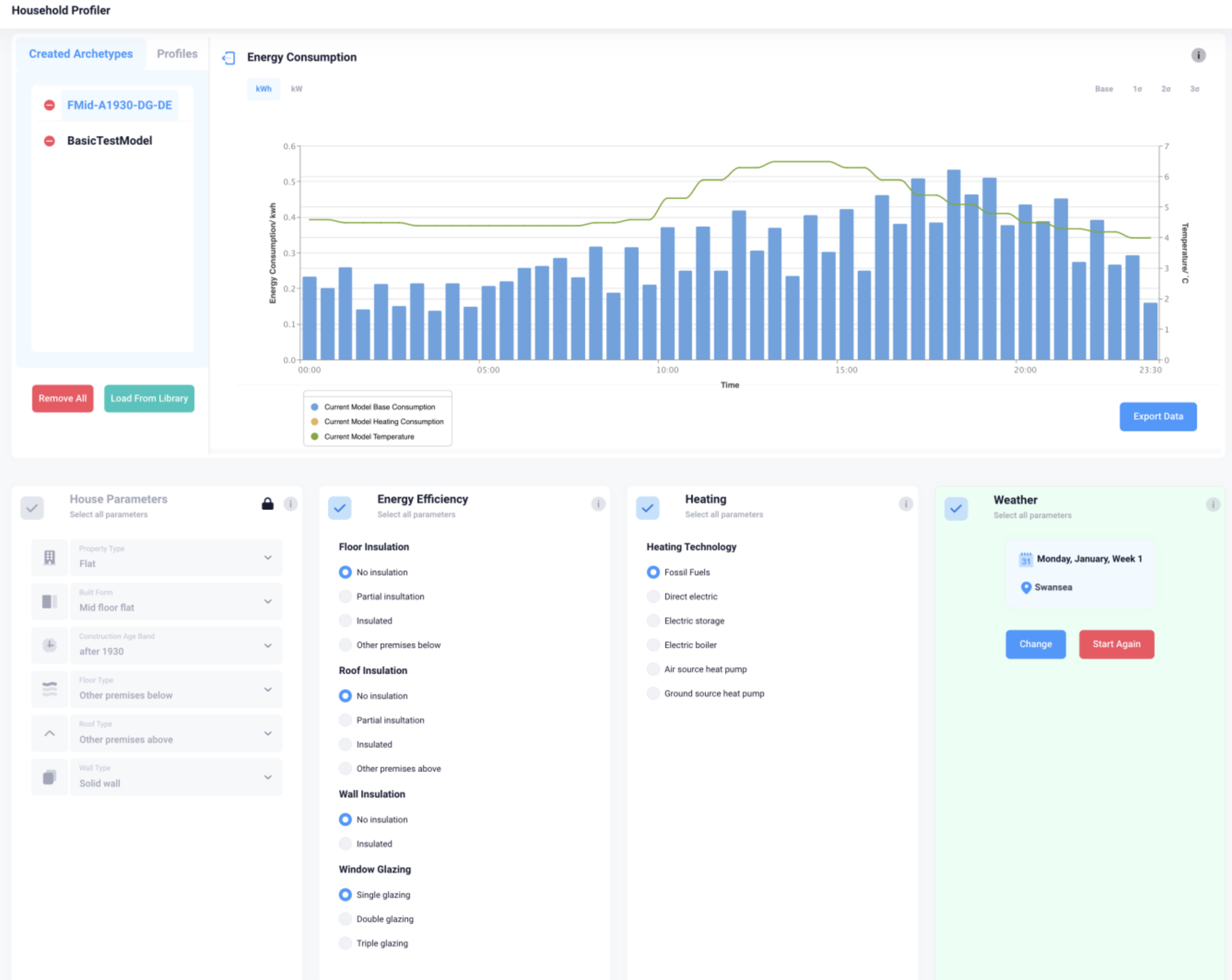


Figure 12 Example parameters selected to generate half hourly electricity profiles. The Property Type, Built Form and date parameters drive the generation of the profile.

5.4. Conversion factors

The input data from smart meters is recorded as energy in kilowatt hours (kWh) occurring within the half hour. To convert to power in kilowatts (kW), the kWh are multiplied by 2. This means the average power for the half hour is represented instead of the instantaneous peak during this time period.

Coordinated universal time (UTC) was used in the model creation and subsequent time stamps are stored as a UTC value. This corresponds to local time in the UK during the winter and for British Summer Time (BST) it is plus one hour.

6. HTC Calculation and Heating Demand

Heat transfer co-efficient (**HTC**) is a measure of heat loss for a building. It is used to calculate the energy required to heat to a desired indoor temperature given the outside temperature.

For a given set of temperature profiles, **heating demand** is determined from simply multiplying the HTC by the difference of the indoor and outdoor temperature and the amount of time that the heating is turned on.

For every property there is a distinct HTC based on the building size and fabric features. A Bayesian methodology has been adopted that fits a probability density function of the HTC to the archetype categories. This means the HTC for a property type falls within a range treated as a random variable. That density function can be sampled to generate data with randomness equivalent to what is found in the real world to provide diversity.

Gas consumption is used to calculate the HTC. Checks on the source data showed that mid-floor flats were not sufficiently represented meaning the model will not perform well if scenarios contain a lot of mid-floor flat properties.

Checking the Bayesian estimator against other methods for a single household where actual performance is known, shows a 22% over estimation of heat demand. Comparatively this outperforms a state of the art top down approach.

Further improvements to accuracy could be made by using a minimum threshold of temperature difference and the subarchetype information to improve predictive power for the Bayesian method. In addition, using reference values from standard assessment procedure (SAP) as the prior estimate in the Bayesian may increase accuracy.

HTC (also known as thermal transmittance) is a measure of the rate at which heat is transferred through a material or assembly, such as a building wall or window, per unit surface area and per unit temperature difference between the inside and outside of the building. Typically, it is expressed in units of $W/(m^2K)$ (watts per square meter per kelvin).

HTC is used to measure building performance in terms of its thermal efficiency, or how well it retains heat in the winter and keeps heat out in the summer. A lower HTC indicates better insulation and more efficient use of energy. This value can be used to compare the performance of different building materials, components, and systems, and to calculate the energy required to maintain a desired temperature inside the building.

A subtly different measure is heat loss coefficient (HLC) which simplifies the relationship of energy and change of temperature, removing surface area by measuring the heat loss directly from temperature and energy data. It should be said that we are using HTC and HLC as somewhat interchangeable terms as they are so closely related.

Some high level sense checks on the data set were done to check that quantity, bias and noise were within reasonable bounds. Gas data preparation was done in a similar way to the previous electricity data set to join UPRN (unique UK property reference) to EPC data. In addition the UPRN was used to find the nearest weather station to be used the reference for outdoor weather conditions.

Gas data was used to calculate the heating demand for buildings. All of the gas input was attributed to heating for the calculation of the HTC. It was reasoned that treatment of data to remove hot water and cooking usage may be complex, and that those smaller consumption values would not correlate with outdoor temperature. If there is no correlation with the observed variable of temperature different then the Bayesian inference will remove those effects as they appear the same as noise.

6.1. EPC analysis - data sense check

To understand the quality and coverage of the gas dataset as well as the quality of the house archetypes in separating fabric features, an EPC analysis for the house archetypes was run. By definition the EPC will correlate with the HTC, so in the later stage of HTC calculation the ordering of the archetypes can be compared to that of the EPC ordering.

For houses, generally pre 1930 construction were EPC B to E with post 1930 having on average better insulation, shown as B to D or better. The sample size for Houses was higher than for Flats, which is due to fewer gas smart meters being installed in multi-dwelling units.

	HOUSE CATEGORY	MEAN EPC
0	Detached pre 1930	46.9
1	Detached post 1930	63.8
2	Semidetached pre 1930	53.9
3	Semidetached post 1930	62.6
4	Midterrace pre 1930	58.1
5	Midterrace post 1930	66.5

Table 2 Mean EPC ratings for Property Type equal to House, with pre 1930 having worse performance than post 1930. Detached post 1930 performing better than semi-detached post 1930 was not expected.

The EPC rating is taken from the CURRENT_ENERGY_RATING field in the EPC dataset. Higher is better with an A rating starting at 92. Table 2 confirms the properties that are in the dataset are what would be expected with Detached being the worst performing (more exposed walls), Semi-detached somewhere in the middle and Mid-terrace the best with shared walls on either side of the property. Likewise pre 1930 construction is less efficient than more modern post 1930 construction.

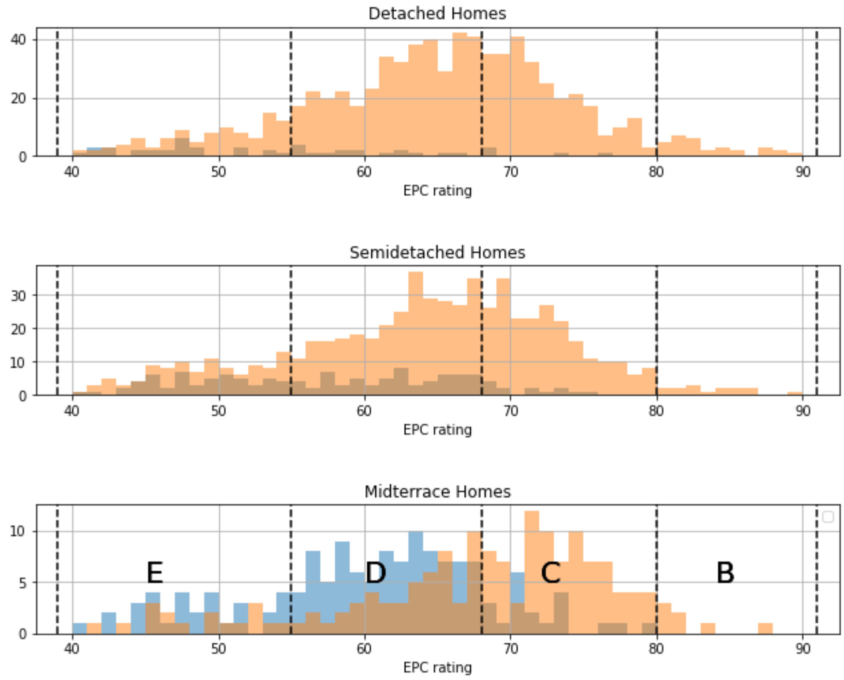


Figure 13 Histograms of post 1930 (orange) and pre 1930 (blue) construction Houses by built form. Generally more modern homes show better insulation.

Figure 13 places the properties into bands A-G with most occurring in the D band. Comparing the dataset to national experience[2], it is reported the D band is the median for England and Wales.

Similarly flats were analysed, with Mid-floor flats performing the best due to protection above and below. The one unexpected result was pre 1930 Mid-floor flat performing better than post 1930 Mid-floor flat. This seems to be due to the low sample size in the dataset. There are only 2 pre 1930 and 10 post 1930 Mid-floor flats, which shows low coverage in the data set for those types of dwellings, reducing reliability of any conclusions that can be drawn.

The lower sample numbers for Flats is to be expected as smart meter installs in multidwelling units is less common, as is the use of gas heating. The sample that was used for the EPC analysis was homes with gas, so this is reflected in the lower number of buildings that would have Mid-floor dwellings.

	FLAT CATEGORY	MEAN EPC
0	Bottom pre 1930	60.1
1	Bottom post 1930	67.7
2	Mid pre 1930	72.5
3	Mid post 1930	70.7
4	Top pre 1930	60.5
5	Top post 1930	68.7

Table 3 EPC for categories of flats within the data set used to create the heating models

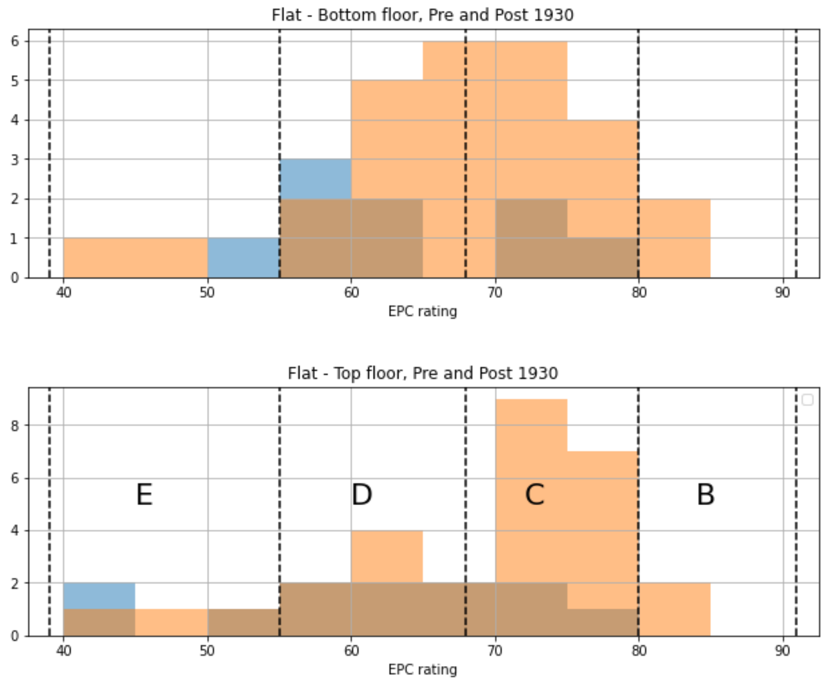


Figure 14 EPC rating histogram for Flats with rating E-B labelled on the occurances. Most common rating is C, consistent with known national averages for flats.

2 https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/energyefficiencyofhousinginenglandandwales/2022

6.2. Gas Analysis – data sense check

Direct gas consumption from the smart meters was first analysed to check that the archetypes would provide enough expression of the heating loads. This initial characterisation was also completed to reveal bias in the data set and show typical heating demands.

Each of the half hour periods for the heating season (beginning of October to end of March) were plotted on graphs for each archetype. The mean and standard deviations were calculated for all of the half hours to examine the trend and noise in the dataset. Mid-floor flats were not shown due to the small sample size.

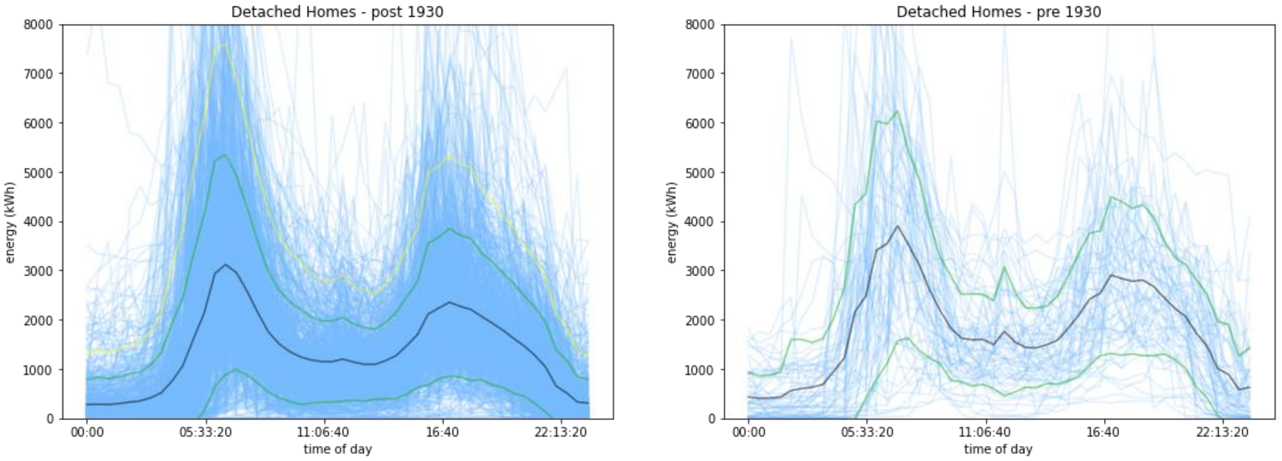


Figure 15 Gas usage for Detached Homes, pre and post 1930 construction by half hour for all days of the heating season. The black line is the mean with green one standard deviation above/below, yellow line is 2 standard deviations.

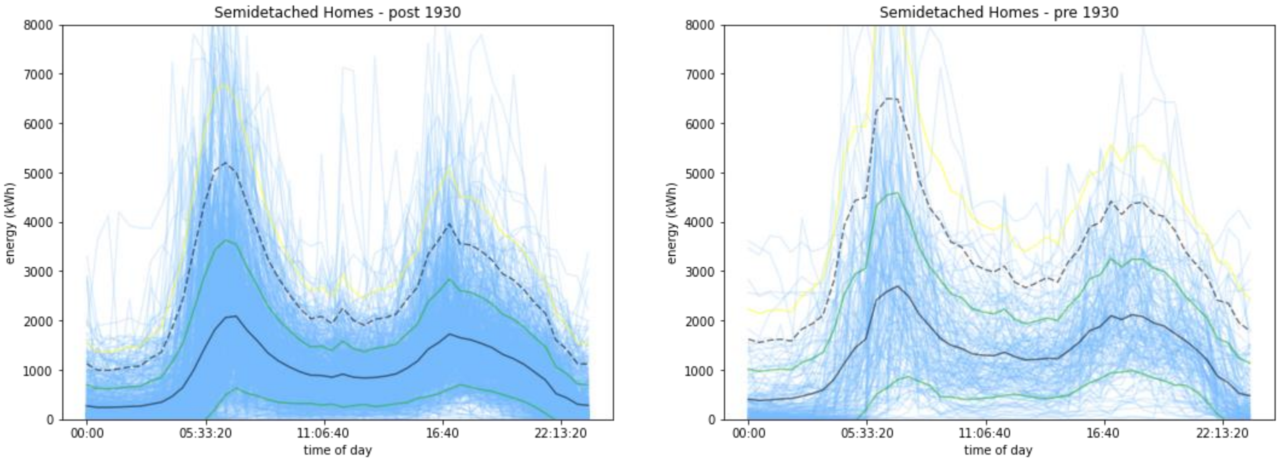


Figure 16 Gas usage for Semi-detached Homes, pre and post 1930 construction by half hour for all days of the heating season. The black line is the mean with green one standard deviation above/below, yellow line is 2 standard deviations.

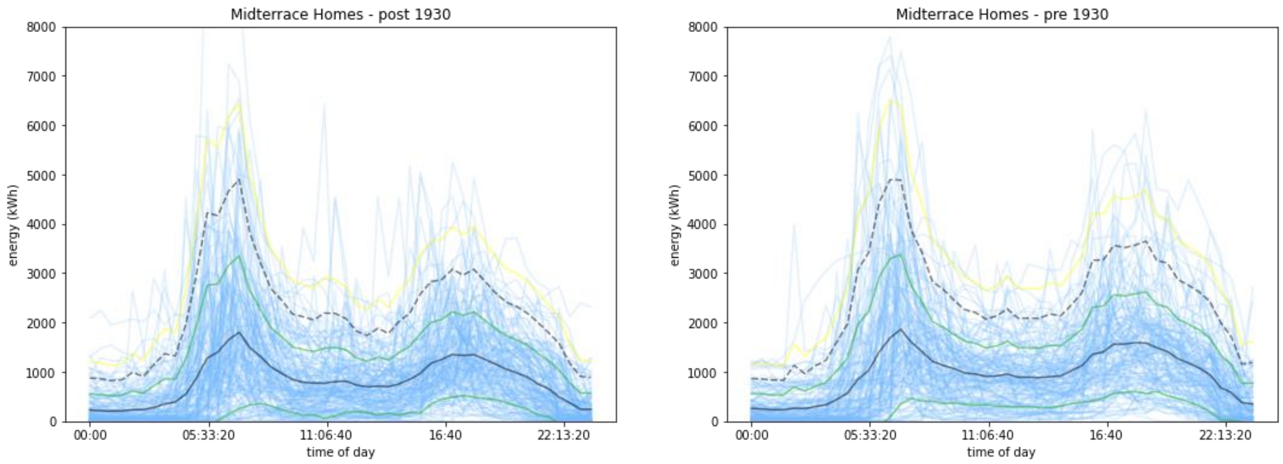


Figure 17 Gas usage for Midterrace Homes, pre and post 1930 construction by half hour for all days of the heating season. The black line is the mean with green one standard deviation above/below, yellow line is 2 standard deviations.

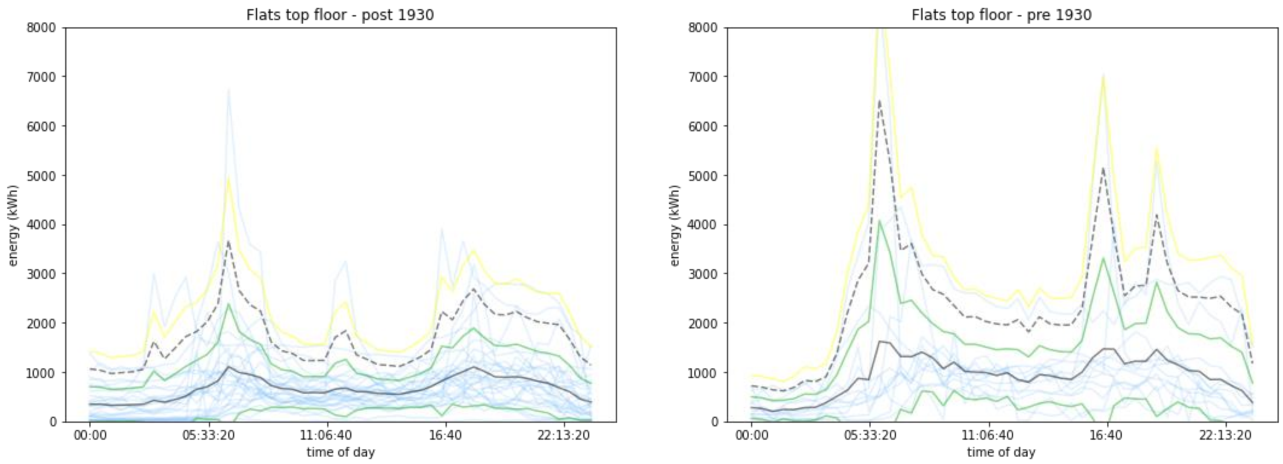


Figure 18 Gas usage for Top floor flats, pre and post 1930 construction by half hour for all days of the heating season. The black line is the mean with green one standard deviation above/below, yellow line is 2 standard deviations.

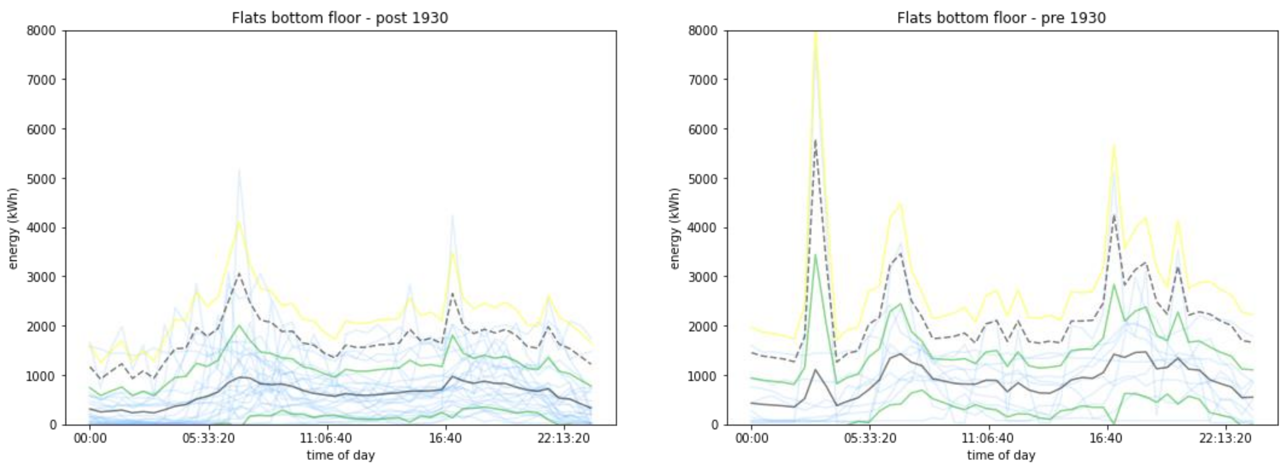


Figure 19 Gas usage for Bottom floor flats, pre and post 1930 consturction by half hour for all days of the heating season. The black line is the mean with green one standard deviation above/below, yellow line is 2 standard deviations.

When the sample sizes are larger, there is a good pattern that emerges from straightforward statistical analysis or frequentist methods. The highest consumers of energy tend to be the outliers, as subtracting 2 standard deviations would take the consumption to negative, which is not physically possible. In the case of flats with small sample sizes the patterns of use look like the individual flat rather than a population trend.

Bayesian analysis and frequentist analysis are two different statistical approaches to analyzing data and making inferences. The strengths of Bayesian analysis make it a good candidate for the model, including

1. Flexibility: Bayesian analysis allows for the incorporation of prior information and the combination of data from multiple sources, which can make the results more robust and generalizable.
2. Probabilistic reasoning: Bayesian analysis provides a natural framework for modeling uncertainty and quantifying the degree of belief in a particular hypothesis or parameter value.
3. Model comparison: Bayesian analysis allows for the comparison of multiple models and the selection of the most appropriate one based on the data.

Whereas a frequentist approach may have weaknesses in this model characterised by:

1. Frequentist analysis is based on the long-run behavior of repeated trials - therefore with different variations in day to day patterns, noise may dominate the statistics and it is difficult to repeat the trial under standardised conditions
2. Although frequentist analysis often requires fewer assumptions and is generally more straightforward and easier to apply than Bayesian analysis there is typically a need for much larger data sets
3. Frequentist analysis does not provide good insight into the structure of the physics

In summary, Bayesian analysis appears more suited for this model because it provides a probabilistic framework for making inferences based on known relationships between physical properties (energy and temperature) that can be constrainted.

6.3. Heat loss and transfer co-efficients

The HLC calculation is formally represented by the following:

$$hlc = \frac{Q_{day}}{\max(T_s - T_{ext}, 0)}$$

Equation 1 Relationship of gas energy to temperature with the heat loss co-efficient expressing the heat loss of the building fabric in units of kWh / K

From the equation there is no need for surface area or volumetric input, the resulting hlc is time invariate (assuming a stationary system) and does not require dimensions of the building. If there is no heat required, then the denominator creates a divide by zero error; to avoid this issue only days that require heat are included in the analysis.

The observed gas heating energy, Q_{day} is a function of the external temperature T_{ext} and the desired internal temperature, T_s (or set point for a thermostat). The external temperature T_{ext} is also an observed value sourced from historical values from the nearest weather station to where the energy and temperature are time aligned. The random variable hlc or heat loss co-efficent is the amount of energy required to raise the temperature by one degree therefore its units are kWh / K. T_s is set as a constant 19 degrees Celcuis.

Daily energy values and temperatures expressed as degree days were used to smooth out lag in the heating system. Degree days are the sum of temperature differences over time therefore the total degree days for the day are be used in the denominator in Equation 1. An enhanced degree day calculation method was used, referred to as the Integration Method[3].

Treating the HLC as an unobserved random variable means that for each building archetype, the heat loss co-efficient and temperature set point is a probability distribution function.

Determining the distribution function and its parameters is a computational problem, fitting the dataset to a function using Bayesian calibration. This provides noise reduction for the HLC, and usually works well for small or large data sets as compared to taking a statistical average that would only work on a large data set.

The HLC can be translated into an HTC by making an assumption of the heating hours. HTC is in units of kW / K, which is power required to raise the temperature by one degree. Standard assessment procedure (SAP)

[3] Further details on the Integration Method can be found at https://www.degreedays.net/introduction

produces U-values that are in units of W / m^2 K, so the HTC can be related to SAP values that are found in EPC records and are known for various building materials. For our purposes the heating hours is assumed to be 12 with the HLC to HTC transformation done on the input Q_{day} value before Bayesian calibration.

Heating hours set to 12 was somewhat arbitrary in order to match HTC values found in literature. In the review of the project, it has been seen that using a time divisor is not required. The reason is that the number of hours of "heating" is multiplied when calculating the energy in the Bayesian calculator – for which 12 was also used. This constant would simply cancel out. When making a comparison to a measured HTC a time divisor would have to be used.

An assumed base temperature of 19°C has been made where internal temperature is not directly measured. In some literature this is the "base temperature" or "set point".

6.3.1. HLC calculation

The smart meter energy and degree days have been recorded on a daily basis. They are in columns labeled energy and degreedays. Degree days come from precalculated look up tables, based on the weather station that is near the smart meter.

The HLC is computed by dividing the energy by the degree days.

Also to set up the calculation of HLC by property type, built form and construction age band, a category is created that uniquely combines all of these dimensions. An outcome will be a statistical description of HLC by these dimensions.

The gas data frame has one row for each Home + Day combination.

- Energy is reported in **Wh** (converted to kWh), so the summary of energy is all of the gas used in the day.
- Degree days is in **degrees C**
- HLC is **Wh / C**
- HTC is **W / C** - using 12 hours of heating demand for a day

```
gas_df['date'] = pd.to_datetime(gas_df['date'])
gas_df.set_index('date')
gas_df['month'] = gas_df['date'].dt.month
gas_df['hlc'] = gas_df['energy'] / gas_df['degreedays']
gas_df['htc'] = gas_df['hlc'] / 12
gas_df["category"] = gas_df["property_type"] + "-" + gas_df["built_form"] + "-" +  gas_df['construction_age_b
and']

winter_gas_df = gas_df[(gas_df['month'] > 9) | (gas_df['month'] < 4)]
```

Equation 2 Python code that converts the gas energy data into HLC with daily data

The data is filtered for winter days, using the month (for all years). Winter is defined as October to March.

	energy	degreedays	hlc	htc
count	223259.000000	223259.000000	223259.000000	223259.000000
mean	56.236544	10.926207	5.265034	0.438753
std	37.583918	3.294028	4.220238	0.351687
min	0.011022	0.100000	0.000759	0.000063
25%	29.760131	8.600000	3.010671	0.250889

	energy	degreedays	hlc	htc
50%	50.075272	11.000000	4.700735	0.391728
75%	74.985987	13.300000	6.758074	0.563173
max	533.091892	21.300000	488.152943	40.679412

Table 4 High level statistics for the result of the HLC and HTC calculations, data has not yet been run through Bayesian modelling.

There are 223,259 days represented. The energy and degree day values are expressed in days. To further characterise the HTC, histograms for each archetype were produced. This highlights cases where there are zero use days and extreme values probably due to low degree days that don't work well for this method. These low degree days are unlikely to have any impact on network operation and can be excluded from the analysis.

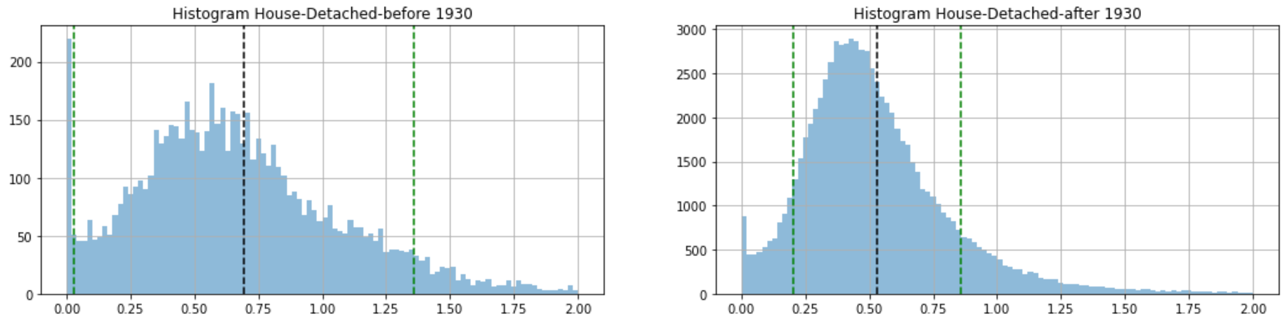


Figure 20 Detached houses before 1930 have a higher spread of values and overall worse energy efficiency

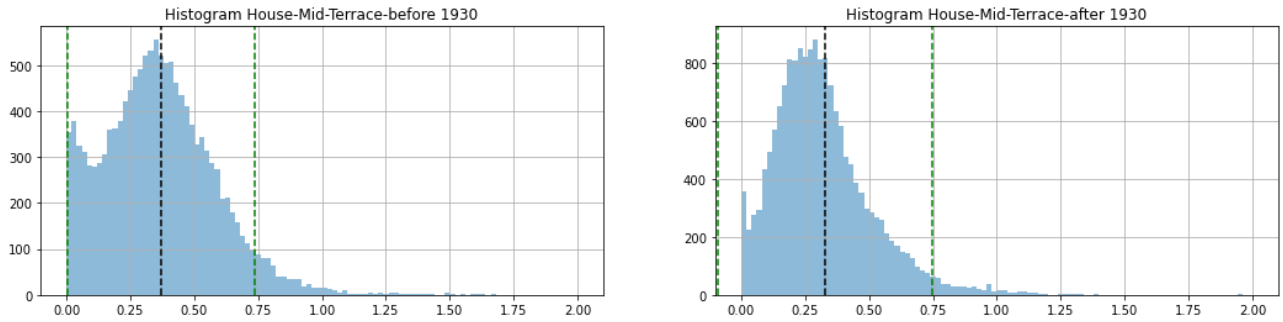


Figure 21 Mid-terraced houses before 1930 have an almost bimodal distribution, meaning that this categorisation may have a subgrouping that would be more expressive

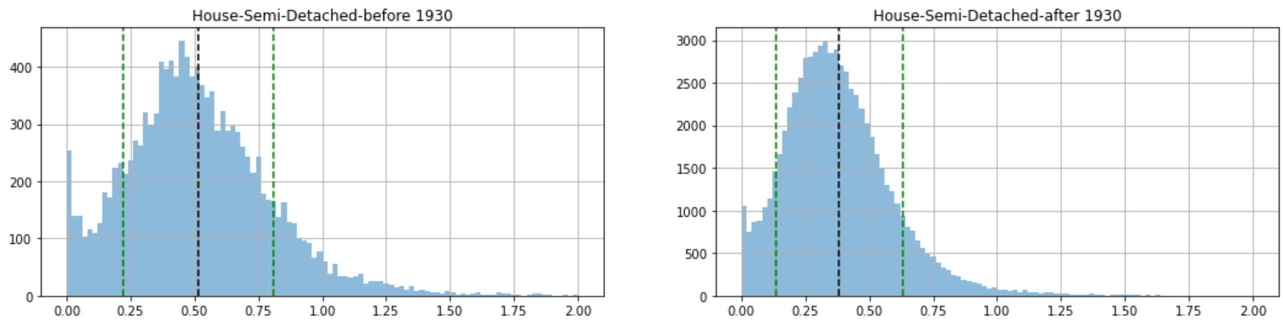
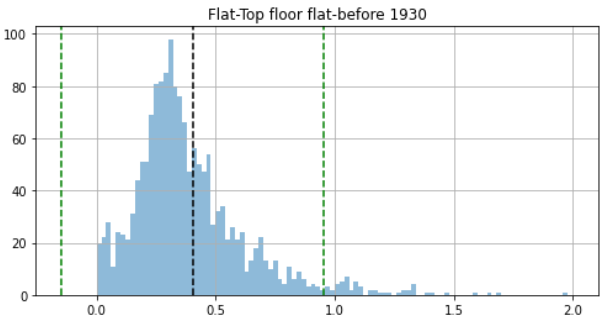


Figure 22 Semi-detached houses are well represented in the dataset, there are more samples than any other category



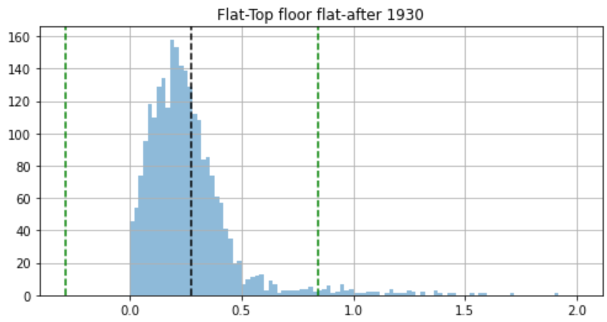


Figure 23 Top floor flats before 1930 have a mean that is not well aligned to the observed peak, in most cases this means the Bayesian method will outperform the use of simple mean values

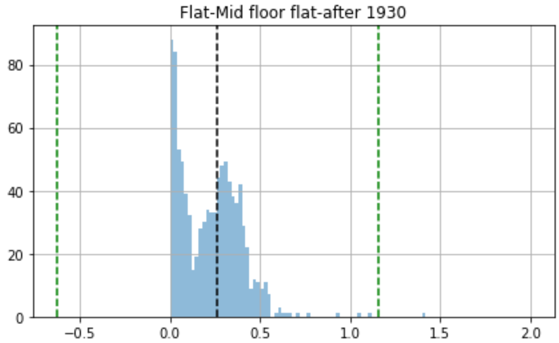
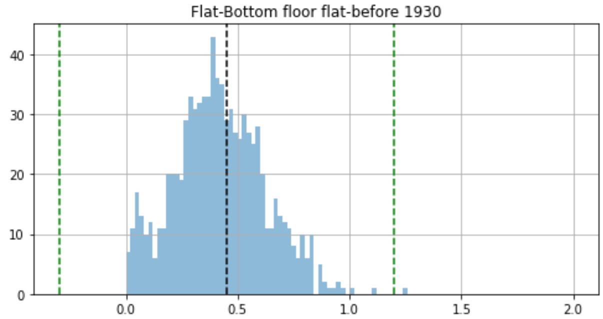


Figure 24 Mid-floor flats look like there is an additional sub-category; the data set for pre 1930 was very small and did not plot well so it has not been graphed, although remains in the model creation



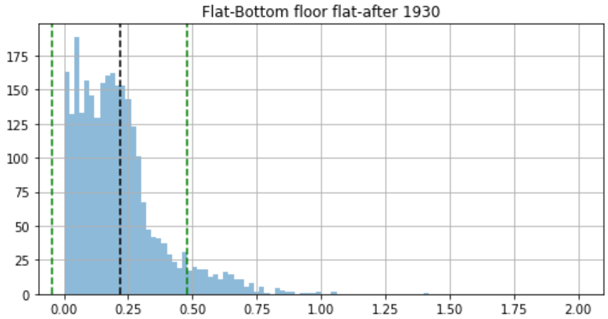


Figure 25 Bottom floor flats after 1930 look to have a wide range of HTCs, which indicates more hidden structure

The Bayesian method will attempt to replicate the shapes of the above histograms by curve fitting a hypothesized distribution function and then extracting the parameters that describe the function. In most cases above a gamma function shape looks like the best representation.

6.3.2. Bayesian calibration

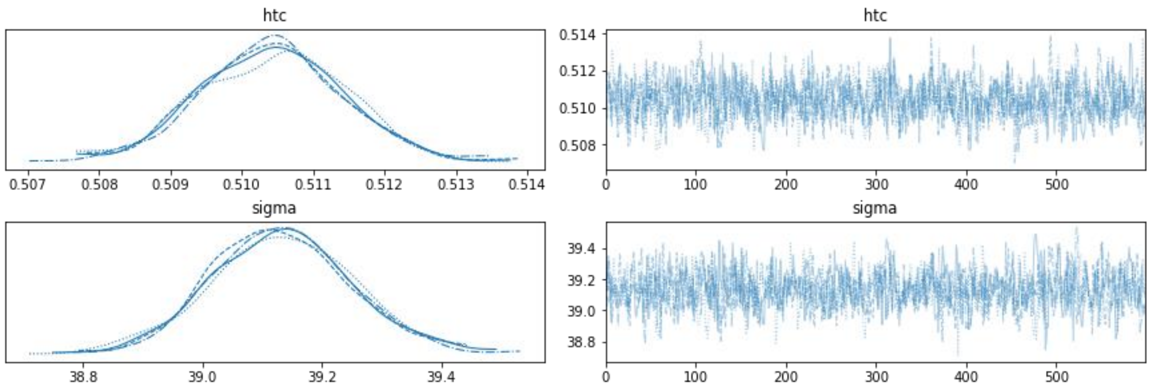
Each of the HTC values is then run through a Bayesian calibration routine to fit the data to a distribution function. The PyMC library is used in Python with data segmented on the archetypes. The Python code is shown to help illustrate the process:

```
obs = winter_gas_df[winter_gas_df['category'] == 'House-Detached-after 1930']

T = obs['degreedays']
Y = obs['energy']

glow_model = pm.Model()

with glow_model:
```

```
    htc = pm.Gamma("htc", mu=1, sigma=1)
    sigma = pm.HalfNormal("sigma", sigma=1)

    E = htc * T * 12

    Y_obs = pm.Gamma("Y_obs", mu=E, sigma=sigma, observed=Y)


with glow_model:

    pbar = Progress()

    trace = pm.sample(target_accept=0.85, return_inferencedata=False, init="advi", chains=4, draws=600, tune=
200, cores=4, callback=pbar.do_update, random_seed=42)

    az.plot_trace(trace);
```

Equation 3 Temperatures as degree days and the daily energy consumption which are observed values are fit to a Gamma distribution

Tests are performed on the model outputs using MCSE and r_hat measurements. The results for each archetype are found in Section 16 Appendix D.



	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.510	0.001	0.508	0.512	0.000	0.000	976.0	1047.0	1.0
sigma	39.133	0.118	38.899	39.352	0.004	0.003	970.0	1102.0	1.0

Figure 26 Output from the Bayesian calibration. The four lines show the four independent training runs to fit the model. Variational inference was used, which is more efficient for larger datasets. R-hat = 1 is a strong indicator of fit along with Monte Carlo Standard Error (MCSE) = 0.

6.3.3. HTC intuitative check

In isolation the HTC is somewhat difficult to interpret. The Bayesian calibration will refine the HTC and reduce the noise, however it is good to check that a known HTC would actually provide good predictive power. A single household case was used to demonstrate the model assumptions work.

First the gas consumption data for the whole year was extracted directly from the output of Equation 2 for a single home identified by unique property reference number (UPRN). The date range was from May 2021 until July 2022. The output contains the total energy for each day and calculated HTC.

Plotting these values shows the summer months having very little gas demand and also small degree day values. Where the HTC is shown in the winter months, there is a visble trend in the data with very noisy summer months when low degree days can occur, but very little gas consumption is present.

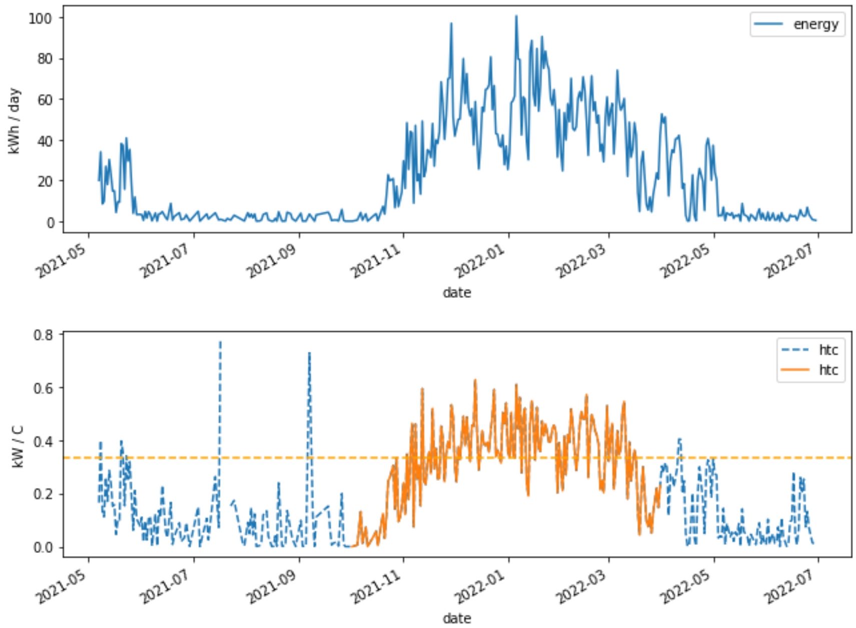


Figure 27 Daily gas energy consumption (top chart) and day by day HTC (bottom chart), the highlighted region is the October – March dates that were used as a filter

	count	mean	std	min	25%	50%	75%	max
energy	170.0	43.389315	22.679311	0.022441	27.148507	45.686158	58.572274	100.656211
degreedays	170.0	10.468824	2.973895	2.700000	8.425000	10.450000	12.675000	16.000000
hlc	170.0	3.979182	1.748055	0.003206	2.925546	4.232489	5.309700	7.514976
htc	170.0	0.331598	0.145671	0.000267	0.243795	0.352707	0.442475	0.626248

Table 5 Summary statistics for sample household; HTC assumes 12 hours of heating per day

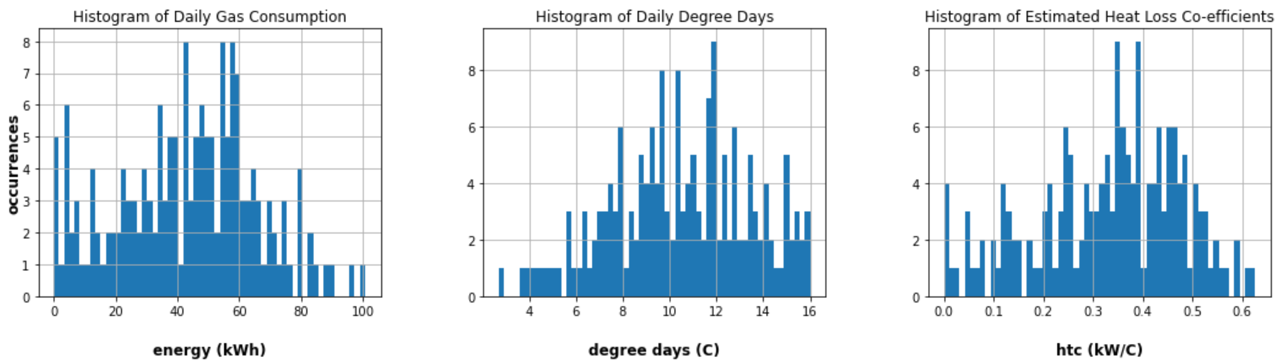


Figure 28 Historgrams of energy use, degree days and HTC for the sample house

Although it may be difficult to read, the EPC record is provided with the address details removed in order to show the context of the property and also classify the property into the archetype and subarchetypes from above.

EPC Field Name	Value
CURRENT_ENERGY_RATING	D
POTENTIAL_ENERGY_RATING	B
CURRENT_ENERGY_EFFICIENCY	68

POTENTIAL_ENERGY_EFFICIENCY	81
PROPERTY_TYPE	House
BUILT_FORM	Detached
INSPECTION_DATE	30/09/2015
LOCAL_AUTHORITY	E06000023
CONSTITUENCY	E14000600
COUNTY	
LODGEMENT_DATE	30/09/2015
TRANSACTION_TYPE	assessment for green deal
ENVIRONMENT_IMPACT_CURRENT	57
ENVIRONMENT_IMPACT_POTENTIAL	75
ENERGY_CONSUMPTION_CURRENT	224
ENERGY_CONSUMPTION_POTENTIAL	118
CO2_EMISSIONS_CURRENT	4
CO2_EMISS_CURR_PER_FLOOR_AREA	40
CO2_EMISSIONS_POTENTIAL	2.2
LIGHTING_COST_CURRENT	62
LIGHTING_COST_POTENTIAL	62
HEATING_COST_CURRENT	964
HEATING_COST_POTENTIAL	656
HOT_WATER_COST_CURRENT	136
HOT_WATER_COST_POTENTIAL	74
TOTAL_FLOOR_AREA	101
ENERGY_TARIFF	Single
MAINS_GAS_FLAG	Y
FLOOR_LEVEL	NODATA!
FLAT_TOP_STOREY	
FLAT_STOREY_COUNT	\N
MAIN_HEATING_CONTROLS	2102
MULTI_GLAZE_PROPORTION	100
GLAZED_TYPE	double glazing, unknown install date
GLAZED_AREA	Normal
EXTENSION_COUNT	1
NUMBER_HABITABLE_ROOMS	6
NUMBER_HEATED_ROOMS	6
LOW_ENERGY_LIGHTING	100
NUMBER_OPEN_FIREPLACES	0
HOTWATER_DESCRIPTION	From main system
HOT_WATER_ENERGY_EFF	Good
HOT_WATER_ENV_EFF	Good
FLOOR_DESCRIPTION	Solid, no insulation (assumed)
FLOOR_ENERGY_EFF	NO DATA!
FLOOR_ENV_EFF	N/A
WINDOWS_DESCRIPTION	Fully double glazed
WINDOWS_ENERGY_EFF	Average
WINDOWS_ENV_EFF	Average
WALLS_DESCRIPTION	Cavity wall, filled cavity
WALLS_ENERGY_EFF	Good
WALLS_ENV_EFF	Good
SECONDHEAT_DESCRIPTION	Room heaters, mains gas
SHEATING_ENERGY_EFF	N/A
SHEATING_ENV_EFF	N/A
ROOF_DESCRIPTION	Pitched, 250 mm loft insulation
ROOF_ENERGY_EFF	Good
ROOF_ENV_EFF	Good
MAINHEAT_DESCRIPTION	Boiler and radiators, mains gas
MAINHEAT_ENERGY_EFF	Good
MAINHEAT_ENV_EFF	Good
MAINHEATCONT_DESCRIPTION	Programmer, no room thermostat
MAINHEATC_ENERGY_EFF	Very Poor
MAINHEATC_ENV_EFF	Very Poor
LIGHTING_DESCRIPTION	Low energy lighting in all fixed outlets

LIGHTING_ENERGY_EFF	Very Good
LIGHTING_ENV_EFF	Very Good
MAIN_FUEL	mains gas (not community)
WIND_TURBINE_COUNT	0
HEAT_LOSS_CORRIDOR	NO DATA!
UNHEATED_CORRIDOR_LENGTH	\N
FLOOR_HEIGHT	\N
PHOTO_SUPPLY	\N
SOLAR_WATER_HEATING_FLAG	N
MECHANICAL_VENTILATION	natural
LOCAL_AUTHORITY_LABEL	Bristol, City of
CONSTITUENCY_LABEL	Bristol North West
POSTTOWN	BRISTOL
CONSTRUCTION_AGE_BAND	England and Wales: 1976-1982
LODGEMENT_DATETIME	30/09/2015 18:04
TENURE	owner-occupied
FIXED_LIGHTING_OUTLETS_COUNT	\N
LOW_ENERGY_FIXED_LIGHT_COUNT	\N
UPRN_SOURCE	Address Matched

Figure 29. EPC values for the house that was used to verify model fit and sense check model estimate

The sample EPC in Figure 29 is interpreted as Archetype 46 (a detached house built after 1930 with cavity walls, solid floors and a pitched roof) from the archetype reference values. The subarchetype is found by matching additional EPC attributes (rood, window, floor insulation) places the property into subarchetype 45. Using the archetype and sub-archetype the HTC is 379 W/K from the archetype reference spreadsheet calculator[4]. This is shown as Method 1 on the comparison table.

Similarly, the Building Research Establishment (BRE) also has published reference data under the EU Tabula programme. TABULA (https://webtool.building-typology.eu/#bm) typical housing stock for this type of home uses 137 kWh/m^2 per annum. The EPC for the property reports 101 square meters which would be 101 meters squared * 137 kWh = 13,837 kWh per year of heating energy use. This is shown as Method 4 on the comparison table.

Taking a year of data from the actual metered gas reports (Start date: 2021-05-07 End date: 2022-06-30) the total use is 8,923.71 kWh including any hot water or other gas used for cooking.

	METHOD	ESTIMATED ANNUAL ENERGY (KWH)
0	Actual measured gas	8,923
1	HTC of 379 W/K [5]	8,086
2	HTC of 331 W/K	7,063
3	HTC of 510 W/K	10,881
4	TABULA	13,837

Table 6 Comparison of different estimation methods, where the EPC is known and translated into a sub-archetype it appears to be a very good estimator of the annual heating. Note Method 3 only takes the archetype, not the sub-archetype as well.

- Method 1 is using all of the SAP values for the fabric found in the EPC record (which has been consolidated on the Carbon Trust tool), providing the best result to actual.
- Method 2 using 331 W/K is from Table 5 above, which is the average Archetype 46 using a frequentist statistic.
- Method 3 is the Bayesian expected value according to an archetype 46, (detached house, post 1930s) model which is derived in the next section.

[4] Carbon Trust – Archetype reference values V3 spreadsheet
[5] HTC * 1,778 degree days for the heating year * 12 hours of heating

6.4. Interpretation of results

The methods for estimating the HTC have different levels of information available. They are presented in order of most information (0 being the actual consumption) to least information (4 being an estimated annual consumption [EAC] from a national reference). It is reassuring that the order of accuracy corresponds to the order of information available to the estimator.

Method 1, summing the SAP fabric values from an EPC provides the most accurate HTC calculation with 9.4% difference from actual. It is probably even closer to reality as it would be obvious that the estimate should be lower than the actual given the actual also contains gas use due to cooking and domestic hot water. And in the case that EPCs are known and accurate this may not be a surprise. If the sub-archetype attributes were not used, the HTC from the spreadsheet tool has a wide range. Furthermore the spreadsheet reference was done for the National Grid license area, so the housing stock assumptions are refined to very specific region, increasing the certainty of SAP values.

Method 2, being a simple average of all daily HTC values during the Winter months calculated for this particular house is underestimating by 21%. Looking at the gold region on Figure 27 and the months of October and November appear to be warm. If those dates were discarded the average would be much improved. A criteria like a minimum of 12 degree days for the day would limit the HTC calculation to a more robust region. Note, this is for a particular house with known energy consumption, but it illustrates the sensitivity to either unusal gas use on a warm day or low gas use on a mild day.

Method 3, the Bayesian method has a maximum likelihood that is *over*-estimating by 22%, which is a similar error to Method 2. The input for training the model is the national population of dwellings that are in the top level Archetype. They will be influenced by a wider set of conditions, and also the same house level sensitivity seen in Method 2.

Method 4 is the equivalent of using a national level estimated annual consumption (EAC) for a dwelling that meets criteria similar to the top level Archetype. The 55% over estimate is an illustration of why top down energy modelling methods can be wildly inaccurate.

Although not tested, it is fairly clear that using the subarchetype information and a minimum threshold for degree days would improve the Bayesian estimation method. The other advantage of the Bayesian method is that *prior distributions* could be initialised with reference values and the *posterior* update done only with the days that meet the minimum degree day criteria.

7. Energy efficiency model

Energy efficiency in the model is expressed through changes to building insulation. From the fabric features that are available in EPC data sets, a simplified set of selection choices are used to indicate levels of insulation.

These changes result in a change in U-value for the property, which is closely related to the HTC. Rather than adding or subtracting the U-value directly, a percentage offset technique was created so that it could be used when the HTC is synthetically generated from the distribution function. If the U-value was used directly, it may have over or under influenced the resulting HTC as it is not normalised for the same square meter areas of building fabric.

Results of the energy efficiency model are internally consistent, and appear to be an accurate reflection of the percentage offset values that have been programmed. The consistency and accuracy are shown to be maintained when used in network scenarios.

The efficiency model is driven from a set of building fabric parameters, that when combined express the overall HTC of the building. This model allows for some of the parameters to be changed to better match sub-archetypes. The following parameters can be changed:

- **Wall insulation** – Insulated, Partially insulated, No insulation
- **Floor insulation** - Insulated, Partially insulated, No insulation, [Flats] Other premises below
- **Roof insulation** - Insulated, Partially insulated, No insulation, [Flats] Other premises above
- **Window glazing** – Single, double, triple glazing

The starting HTC is taken from the most frequently occurring fabric feature for a top level archetype in the National Grid licence areas. For instance, most semi-detached houses post 1930s construction have double glazing. That starting HTC is then refined by shifting the HTC up or down based on better or worse insulation choices.

7.1. Data source and encoding

The assumption for the base HTC is the distribution function defined by mu and sigma for each top level archetype, found as a result of the Bayesian calibration. The default efficiency parameters are then found by looking up the most frequently occurring parameter selection. That parameter selection then has a label as a subarchetype.

From the possible parameter choices above there is a translation into a percentage change in U-value to the base HTC for each fabric element. A percentage is used in order to make the changes independent of surface areas. The percentage difference was calculated from a set of typical U value changes for each archetype. These are encoded in the energy efficiency module of the Glow Simulator as a relative change to a baseline. For example:

```
"House": {
    "before 1930": {
      "Detached": {
        "mu": 0.575,
        "sigma": 0.004,
        "base": {
          "archetypeId": "34-25",
          "floorInsulation": "No insulation",
          "roofInsulation": "Insulated",
          "wallInsulation": "No insulation",
```

```
            "windowGlazing": "Double glazing",
            "startMask": [ 0, 0, -4.41, 0, -5.13, -24.85, -9.59 ],
            "efficiencyMask": [ 4.3, 4.49, 0.83, 4.87, 19.8, 17.4, 11.62 ]
        }
    },
```

Table 7 Encoding of base HTC and energy efficiency offsets to be applied when changing state

The efficiencyMask values are percentage changes from one state to another with each position in the array representing the insulation parameters and glazing. They were taken as offsets from the archetypeId (in this example 34-25) which has a starting state described by the insulation attributes in the base element.

The Carbon Trust Options Appraisal Tool supplied the HTC differences for the energy efficiency interventions. The assumption was made to use the average change for the sub-meter archetypes (A-L), for example archetype 34 above when creating the efficiency mask values. The mask is constructed as:

```
"efficiencyMask" : [ Roof State 1, Roof State 2, Floor State 1, Floor State 2, Wall State, Window Glazing State 1, Window Glazing State 2 ]
```

As the states are toggled the percentages are summed, either as positive or negative additions to the energy efficiency. Then at the end, the percentage is applied to the mu value that is fed into the Gamma function for simulation. For instance, going from the base case of Floor Insulation = “No insulation” to Floor Insultation = “Insultated” would first apply 0.83% for No to Paritial and then 4.87% for Partial to Insulated for a total of 5.7% improvement in HTC = 0.542.

7.2. Evaluation

Part of the evaluation of the retrofit efficiency model is to check for consistency within the energy transition scenarios constructed by Carbon Trust. The assumption is that improved efficiency will result in better HTC values and lower energy consumption.

Accuracy is difficult to assess without a known true value. A bounded assessment will be made to determine if the accuracy is better or worse than other estimation techniques using the transition scenarios as compared to current Elexon averages.

7.2.1. Consistency

Three scenarios built by the Carbon Trust were used to validate that the percentage offsets were being applied correctly to complex configurations. In all cases the expected ordering emerged.

For example, for Withycombe Raleigh the Consumer Transformation scenario illustrates the stratification clearly. The highest three lines represent the Extreme weather case. The top line is Low efficiency takeup, which is the highest energy, the middle line Medium efficiency takeup and the bottom line High takeup of energy efficiency all for year 2050. Aggregated as a population, the expected clusters and ordering emerge.

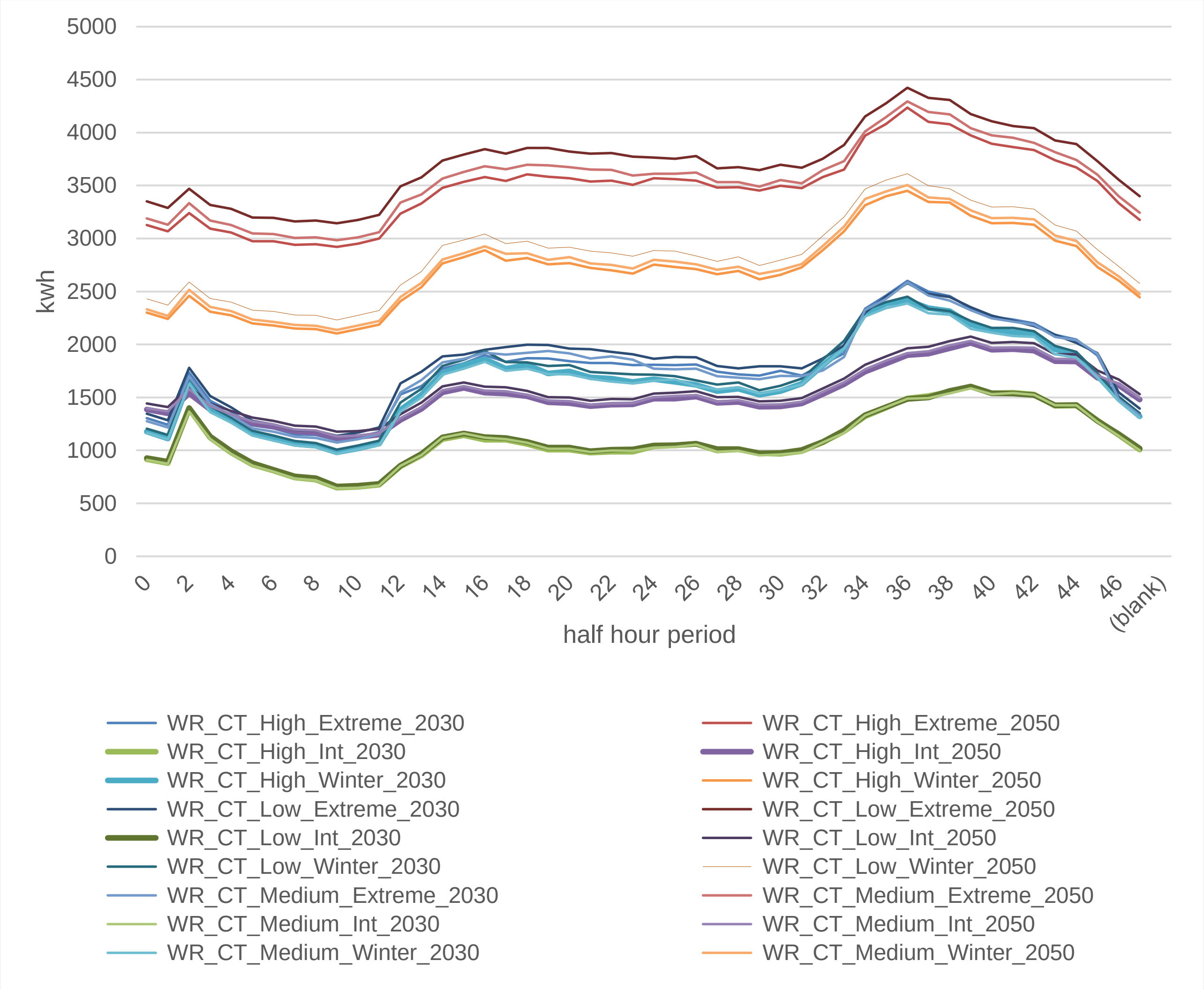


Figure 30 Energy profiles for heating simulated for the whole of WithycombeRaleigh (approximately 1,843 homes) and takeup of heat pumps using percentage offsets to the HTC random variable, results are in the expected order

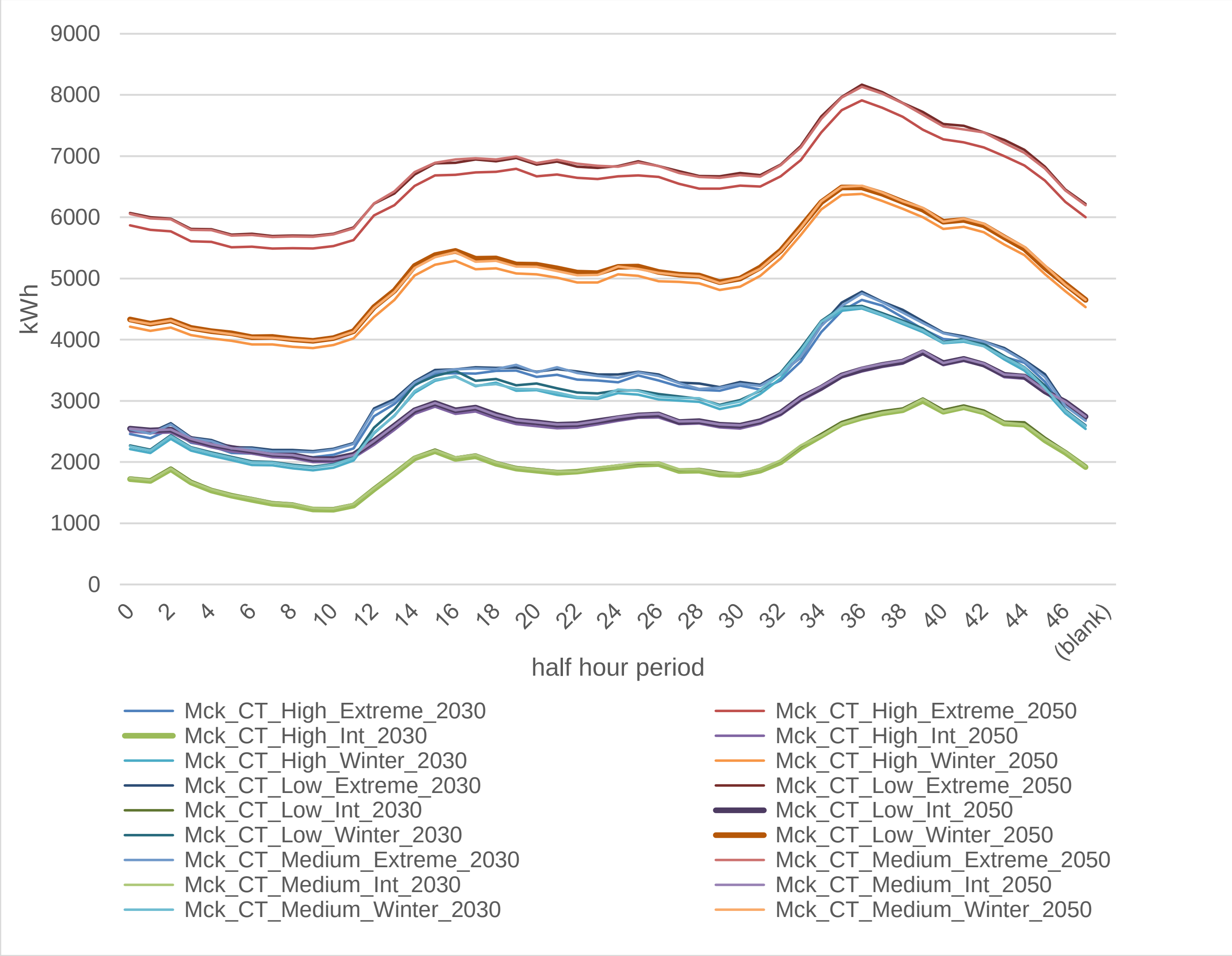


Figure 31 Energy profiles for heating simulated for the whole of Mackworth (approximately 3,363 homes) and takeup of heat pumps using percentage offsets to the HTC random variable, results are in the expected order

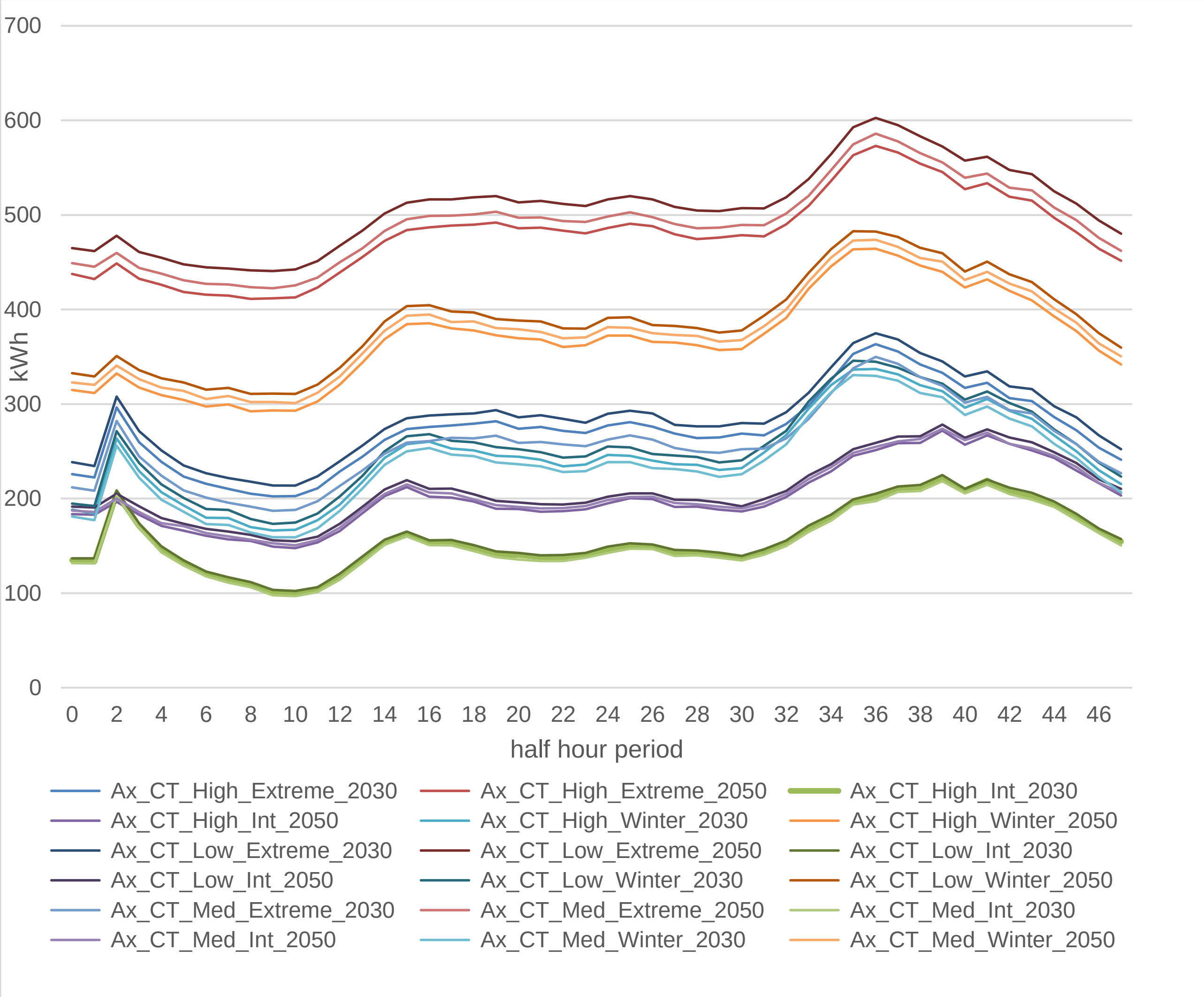


Figure 32 Energy profiles for heating simulated for the whole of Axbridge (approximately 252 homes) and takeup of heat pumps using percentage offsets to the HTC random variable, results are in the expected order

There are differences in heating technology assumptions for the scenarios as well as the change in energy efficiency measures. The very cold, extreme day is a good example to illustrate the percentage difference in the population. For instance the Ax_CT_Low_Extreme_2050 being the worst performing is 6% higher than Ax_CT_High_Extreme_2050, for the total energy used across the day. Clearly the half hour patterns visually match. Likewise for Withycombe Raleigh there is a 6.5% difference. Examining the percentages that are programmed in the energy efficiency model they range from 4% - 20%, so on a mix of properties the 6% seems realistic when considering energy efficiency impact on heat demand.

Note the overall 4% - 20% differences between scenarios also include adoption of heat pumps. The efficiency of the heat pump is modelled as a constant, however more heat pumps are present in the High adoption and 2050 scenarios as higher takeup is an underlying assumption. Again, the rank order of Low to High and 2030 to 2050 is consistent for all of the weather patterns, and more distinctive for the higher heating demand days.

7.2.2. Accuracy

A boundary comparison was made against the industry standard[6] profiles supplied by Elexon. These are average half hour consumption figures using 10 years of historic data for different profile classes and seasons of the year. The seasonal days for Profile Class 1 and a Winter day for Profile Class 2 were

[6] https://www.elexon.co.uk/wp-content/uploads/2012/01/Average_Profiling_data_201314_evaluated@10yearNET_v1.0.xlsx

calculated by multiplying the total number of dwellings in the scenario by the half hour consumption for the respective profile class.

For this comparison the Medium energy efficiency scenario was used, which means some efficiency will be implemented beyond what is there today. The anticipated result is lower consumption with better energy efficiency, however with higher heat pump take up, the base consumption will increase for Winter days compared to current levels. The Profile Class 2 reference was required for WithycombeRaleigh and Axbridge areas because they contain a reasonable amount of night storage heaters.

The Winter months are used in the comparison because it would be unlikely to see the effects of energy efficiency in the warmer Summer months.

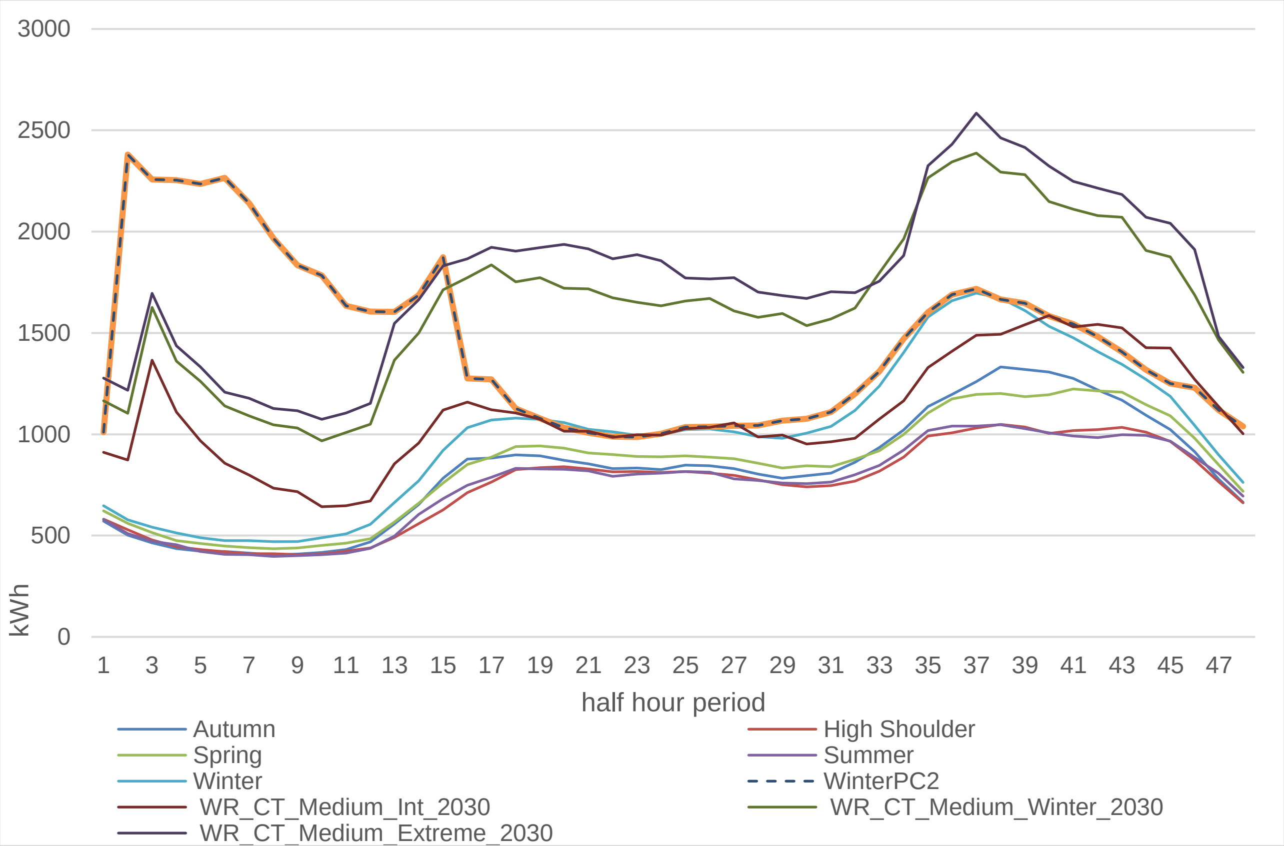


Figure 33 Energy profiles comparing Elexon Profile Class 1 half hour for the whole of WithycombeRaleigh (approximately 1,843 homes). Winter PC2 applies a Elexon Profile Class 2 (night storage heater) to all of the homes. Scenario labels are the same as above, reduced number to be able to fit on to graph.

WithycombeRaleigh has a mix of heating types, so the very pronounced peak due to night storage heaters is over inflated. The scenarios (Medium) for two different seasonal days (Extreme and Average Winter) are much higher than the industry profiles, this is believed to be due to the transition to heat pumps in the Medium scenario for 2030, compared to the Elexon profiles, which are based on historic data. The Intermediate scenario day starts to track with industry data, supporting the explanation that heat pumps are contributing to a baseline shift on cold days. The shapes are generally consistent, with one notable time alignment issue. The times seem to be off by 30 minutes which appears to be difference in how the industry profiles label data versus the recorded data – for this analysis it is noted, but the extra data manipulation has not been done.

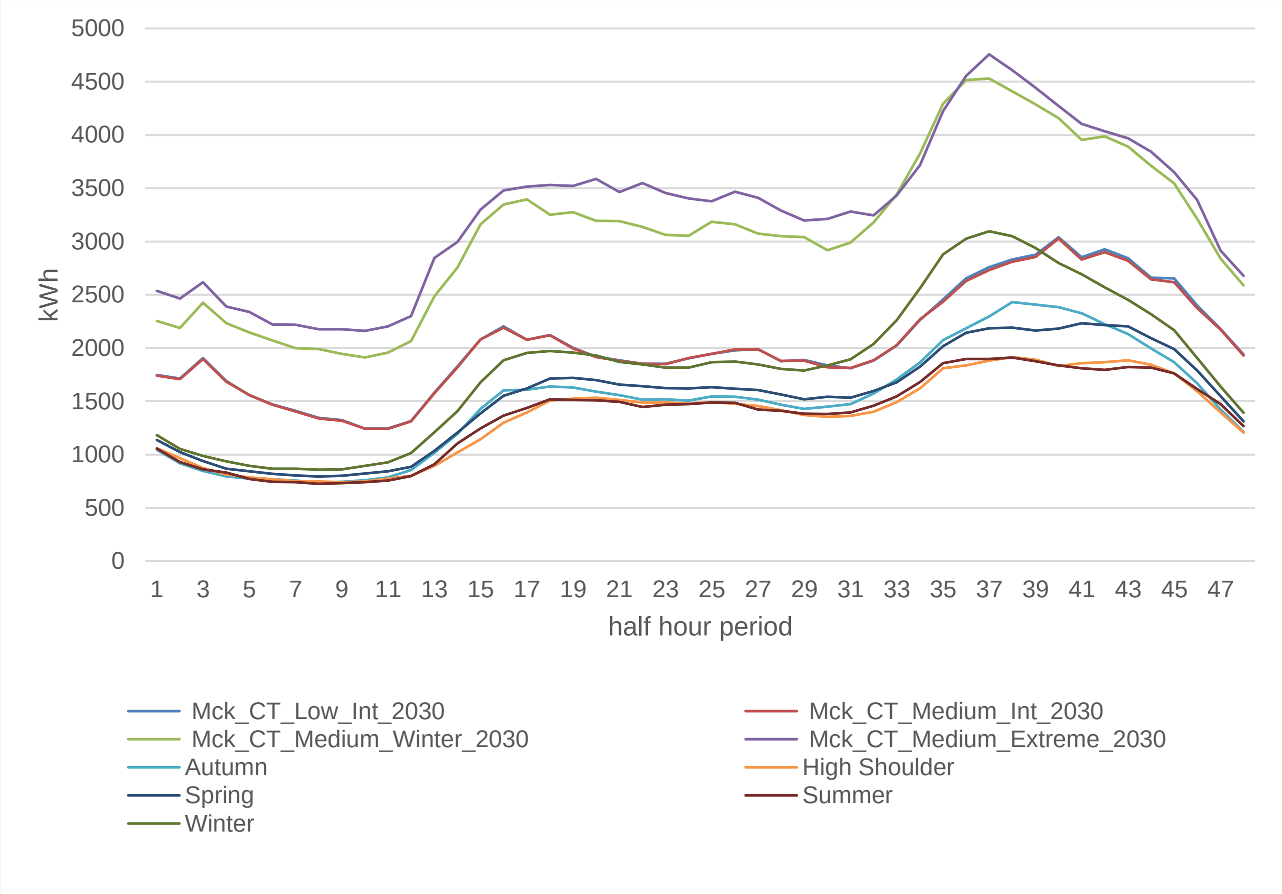


Figure 34 Energy profiles comparing Elexon Profile Class 1 half hour by season for the whole of Mackworth (approximately 3,363 homes). Scenario labels are the same as above, reduced number to be able to fit on to graph.

Mackworth does not have many night storage heaters, so only Profile Class 1 industry profiles are calculated. The scenarios (Medium) for two different seasonal days (Extreme and Winter) are much higher than the industry profiles. This is believed to be due to the transition to heat pumps in the Medium 2300 scenario. The Intermediate scenario day starts to track with industry data, supporting the explanation that heat pumps are contributing to a baseline shift on cold days, however the scenario estimates do generally look higher. The shapes are generally consistent, notwithstanding the time alignment issue described above.

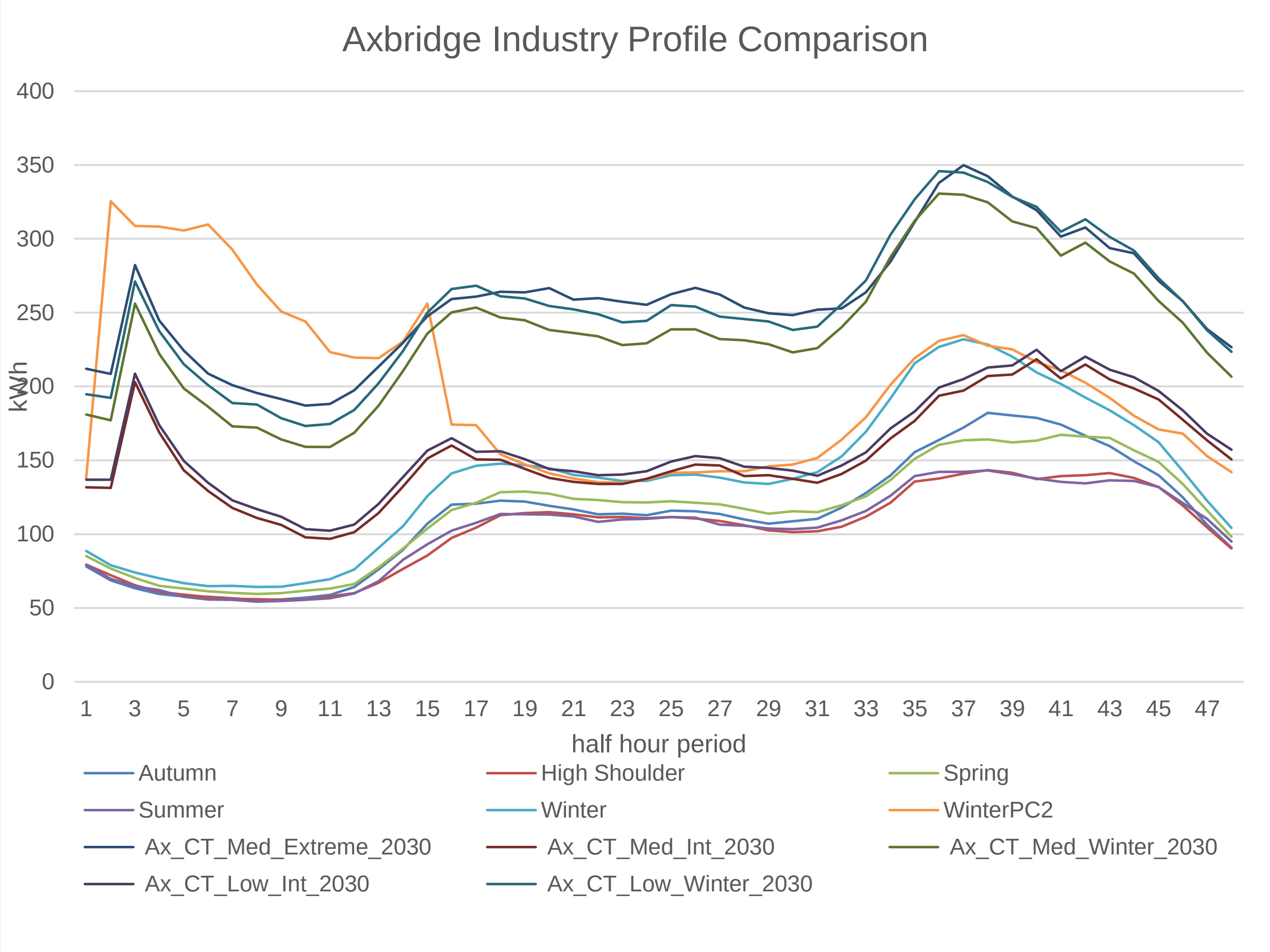


Figure 35 Energy profiles comparing Elexon Profile Class 1 half hour by season for the whole of Axbridge (approximately 252 homes). Winter PC2 applies a Elexon Profile Class 2 (night storage heater) to all of the homes. Scenario labels are the same as above, reduced number to be able to fit on to graph.

In Axbridge, night storage heating is a larger percentage of the overall heating technologies present. The pronounced peak due to night storage heaters is therefore comparatively higher, but this comparison shows the run time of real storage heaters is longer. Clearly Profile Class 2 must be used over Profile Class 1 industry profiles. To see the shift in energy efficiency where night storage is present, the Low versus Medium take up of energy efficiency was added, the efficiency gains are still reflected in the generated profile. Like the other scenarios above, additional heat pump load and time misalignment are present.

8. Application of HTC and heating technology

From a set of HTC values that are determined in the Bayesian training, heat demand profiles are generated using a sampling from the distribution functions that were created for each archetype. Sampling is computationally lightweight with only 24 values required to represent the archetype distributions.

The heat demand is turned into an energy profile using a basic heating technology simulator. The simulator requires the selection of a heating technology with the heat demand provided as input.

Three basic simulation models are coded; 1) an air source heat pump, 2) direct electric heat and 3) night storage heating.

The air source heat pump has a co-efficient of performance with an assumption that the heat pump will run continuously to meet the total heat demand for the day.

Direct electric heat has a weighted coin chance of being on or off for each half hour of the day, during the morning and evening hours there is a 70% biased chance that heat will be required, whereas during the day it is a 35% chance heat is required.

For night storage, the daily heat demand is charged into the heater between the hours of 01:00 and 05:00 with a peak power of 3kW and maximum heater capacity of 6kWh. A decay function is used to spread the load out over time.

More work could be done on making the heating technology models more realistic. Due to time limits within the project only basic simulation models were used. For heat pumps, the model is considered "ideal", although research shows small (less than 5%) morning and evening bumps in energy for heat pump users when the weather is cold.

Since the HTC is modelled as random variable within a Bayesian framework, the statistics describing the HTC by property type, built form and construction age band can be used to generate data. A gamma function is used to draw a sample from the statistics. The sample is made for the HTC for each of the homes in the housing stock or in the case of the Household profiler, a single sample represents the HTC for that home.

Once the HTC is generated, an energy efficiency co-efficient is applied to either increase or decrease the efficiency of the building fabric based on the efficiency measures selected. The degree days for the location are used and then the demand for the day is calculated.

The heating technology selection has been made and a function is called that makes an interpretation of the day's heating demand into half hourly periods.

8.1. Heat Demand

The heating demand is a function of the degree days, HTC and heating technology selected. The degree days are referenced from the location and date parameters set in the Weather object, with a constraint on the months that will be considered. The constraint on the months (Oct – April) is configured in the Weather object.

From the set of statistical data, the mean and variance of the HTC is loaded based on the household parameters given. The HTC that will be used for the calculation is a sample based on a Gamma distribution of those values. The statiscal sampling provides the variation that is seen in real data.

Any energy efficiency measures that have been put in place for the home will affect the HTC. A percentage offset is applied to change the HTC and then the final heating demand for the day is calculated. A number of heating hours are assumed inline with those that were used to calculate the HTC (12 hours in this case) and then a conversion to watt hours is done by multiplying values by 1000.

This heating demand is spread through half hourly periods by applying heating technology models (detailed below).

```
  generateEnergy() {

    let degreedays = this.weather.degreedays;
    let htcParams = HTC[this.household.propertyType][this.household.constructionAgeBand][this.household.builtForm];

    this.energyefficiency.household = this.household;

    this.htc = Gamma(htcParams.mu, htcParams.sigma);
    this.daydemand = this.energyefficiency.interventionCoefficient * degreedays * this._attributes.hoursHeatDemand *
1000;

    if(this.heatingTechnology === 'Air source heat pump') {
      this._airsourceHeatPump();
    } else if(this.heatingTechnology === 'Ground source heat pump') {
      this._groundsourceHeatPump();
    } else if(this.heatingTechnology === 'Direct electric') {
      this._directElectric();
    } else if(this.heatingTechnology === 'Electric storage') {
      this._storageElectric();
    } else if(this.heatingTechnology === 'Electric boiler') {
      this._electricBoiler();
    } else {
      this._lowTemperatureBoiler();
    }

    return this.energy;

  }
```

Figure 36 Implementation of the heating energy requirement, the degree days are found for the location and day, then an HTC is sampled for the house hold parameters; the total demand is then calculated with a profile applied by heating technology

8.2. Air Source Heat Pumps

Within the model, air source heat pumps have a COP parameter of 2.6, taken from Carbon Trust's option appraisal tool. This assumes a flow temperature of 50 degrees C. The demand is split evenly across the day, which is the ideal running pattern for a heat pump.[7]

```
  _airsourceHeatPump() {
    // even split across day
    let cop = 2.6; // Options appraisal tool 50 degree flow

    let hhvalue = this.daydemand / cop / 48;
```

[7] S. Caird, R. Roy, S. Potter, Domestic heat pumps in the UK; user behaviour satisfaction and performance, https://doi.org/10.1007/s12053-012-9146-x

```
    for(let i=0; i < 48; i++) {
        this.energy[i] = hhvalue;
    }
}
```

Figure 37 Implementation of the heat pump heating model

8.3. Direct Electric

Direct electric heating is modelled as the total heat demand from HTC and degree days, with a stochastic behaviour from occupancy and time of use. The stochastic properties have been determined from gas consumption analysis showing two time regions of use with moderate day time use.

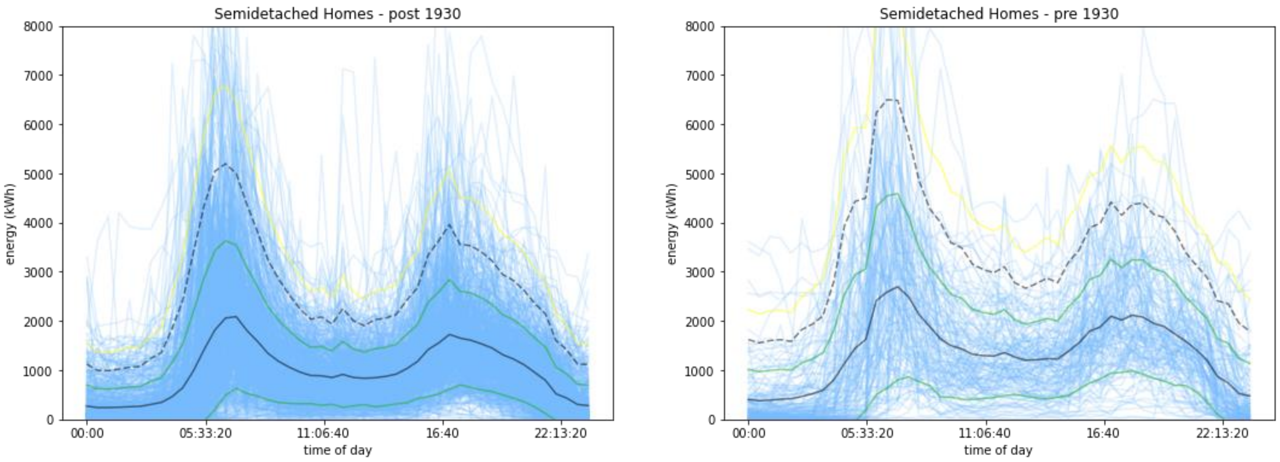


Figure 38 Typical gas use for Semi-detached homes. The dark black line is the mean consumption of all Semi-detached homes for all heating season days.

The implementation of the direct electric segments the day into three regions. The morning 06:00 to 12:00, then daytime from 12:00 to 17:00 and evening from 17:00 to 23:00. Statistical sampling functions are used to simulate 70% chance of using heating in the morning or evening and 35% using during the day. Each segment of the day is an independent sample. This should result in a mix of usage patterns when multiple houses are added together.

Random behaviour has further been built into the model with the amount of electric heating following a normal distribution within the segment of the day. In the morning and evening co-efficient of energy has a spread centered around 2 with the day time spread centered around 1. This co-efficient is multiplied by the heating demand that would be spread across a 12 hour period. The co-efficients are to obtain a shape, therefore are unitless, when used they are multiplied by the energy required in the half hour which then results in a unit of kWh.

Using 2 and 1 attempts to preserve the whole energy requirement for the day given a 70% chance and values that will be above or below 2 for a 12 hour period, while in the day there is a 35% chance above or below 1 for a 6 hour period. The demand is over the theoretical requirement (by 30%) if all of the segments of the day happen to occur, approximately equal if 2 of the 3 segments occur and under-estimated (by 30%) if one of the three occurs. This is consistent with some possibility that occupants overheat but are unlikely to underheat.

```
_directElectric() {

    let energy = [];
    // assume 12 hours of heating or 24 half hour periods
    let hhvalue = this.daydemand / 24;
```

```
    let morning = Bernoulli(0.7);
    let day = Bernoulli(0.35);
    let evening = Bernoulli(0.7);

    this._initEnergy();
    console.log("_directElectric " + morning + " " + day + " " + evening);

    for(let i=0; i < 48; i++) {
      // assume no heat unless in the time below
      energy[i] = 0;

      // assume 06:00 - 11:59 as first heating region
      if(i > 11 && i < 24) {
        if(morning) {
          energy[i] =  hhvalue * Math.abs(Gaussian(2, .5));
          //  energy[i] =  hhvalue;
        }
      }

      if(i > 23 && i < 34) {
        if(day) {
          energy[i] =  hhvalue * Math.abs(Gaussian(1, .5));
          //  energy[i] =  hhvalue;
        }
      }

      // assume 17:00 - 22:59 as second heating region
      if(i > 33 && i < 46) {
        if(evening) {
          energy[i] = hhvalue * Math.abs(Gaussian(2, .5));
          //  energy[i] =  hhvalue;
        }
      }
    }

    console.log(JSON.stringify(energy));
    return energy;

  }
```

Figure 39 Implementation of the direct electric heating model, including the stochastic elements during the 3 segements (morning, day and evening) of the day

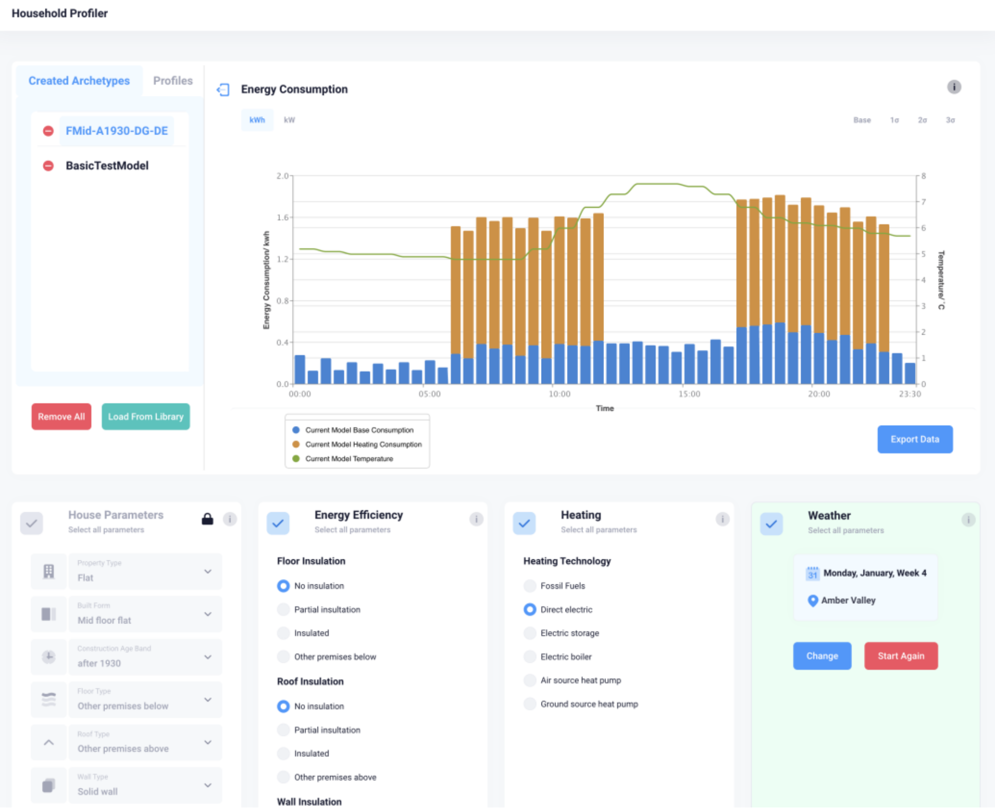


Figure 40 Example direct electric profile generated with randomness in time, notice the gap during the day and morning and evening activated with variation in the half hour demand

8.4. Storage Heaters

Storage heaters have been modelled to switch on in half hour 3, corresponding to what we are seeing in Axbridge preliminary substation aggregated data for Winter days in 2020. The historical analysis from Elexon[8] of Profile Class 2 meters has a similar pattern, albeit the start time is mixed time zone (GMT and BST) on the same chart.

Further discussions with end users of night storage heaters and research indicates that storage heaters have a behaviour of drawing power in a non-linear fashion and that the full heating demands of a property are often not met by the capacity of the storage heaters.[9]

[8] Elexon Average Profiling, https://assets.elexon.co.uk/wp-content/uploads/2012/01/28163748/Average_Profiling_data_201314_evaluated%4010yearNET_v1.0.xlsx
[9] S. Darby, Smart electric storage heating and potential for residential demand response, https://doi.org/10.1007/s12053-017-9550-3

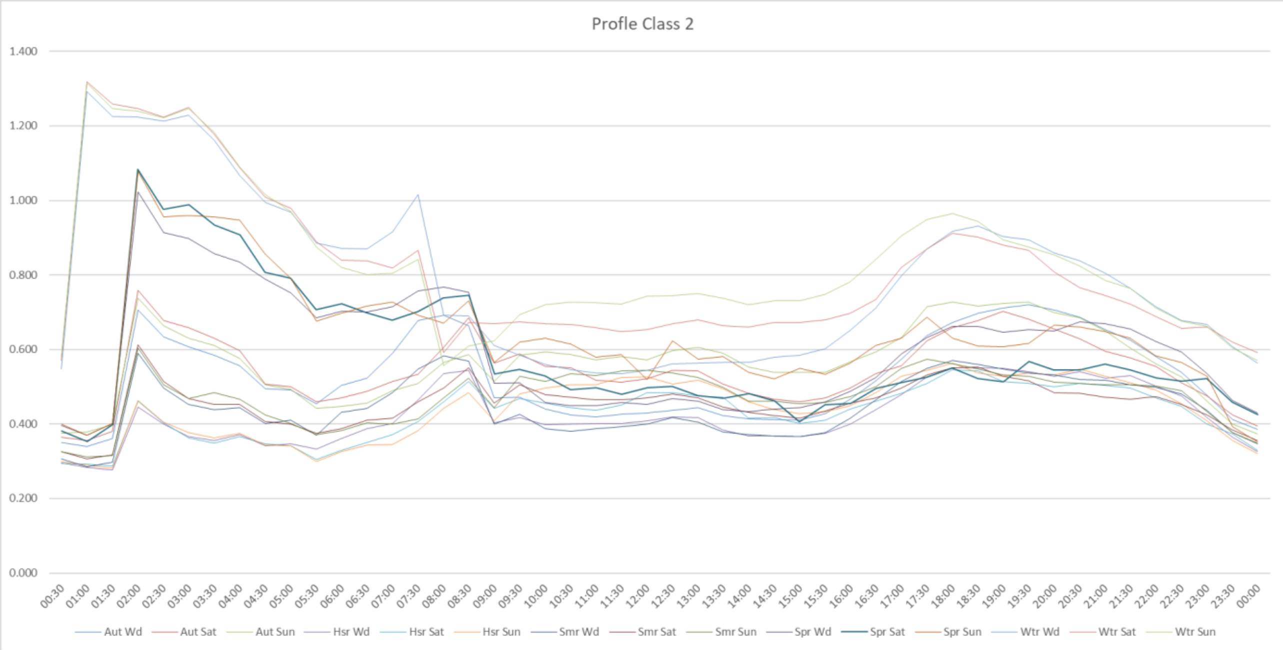


Figure 41 Average half hourly power consumption of Profile Class 2 meters, collected over a 10 year period by Elexon

This is most easily described in the implementation code where energy is allocated up to a maximum storage parameter, in this case 6kWh and a maximum draw of 3kWh for the half hour period. The 3kWh for the half hour corresponds to a 6kW load for the half hour period. This would be consistent with an average property having two storage heaters.[10]

The heating demand is then split over four hours or eight half hour periods. The natural logarithm is used to decay the charging rate, which approximates the non-linear function seen in the real data.

```
_storageElectric() {

  let maxStorage = 6000; // maximum storage is 6kWh
  let maxWatts = 3000; // maximum watt draw
  let decayRate = 1/8;
  let integratorCheck = 0;

  // apply decay
  let hhvalue = this.daydemand / 4;

  for(let i=0; i < 48; i++) {
    if(i > 1 && i < 9) {

      if(hhvalue > maxWatts) hhvalue = maxWatts;
      if(integratorCheck > maxStorage) hhvalue = 0;

      this.energy[i] = hhvalue+hhvalue*Math.log(.7);
      hhvalue = this.energy[i];
      integratorCheck += hhvalue;

      console.log(`Storage heater ${integratorCheck} / ${this.daydemand}`);

    } else {
      this.energy[i] = 0;
```

[10] https://www.esru.strath.ac.uk/Documents/MSc_2013/Becerril.pdf

```
        }
      }
    }
```

Equation 4 Code for applying energy to storage heaters with a decay function and constraints for maximum charge rates and capacity

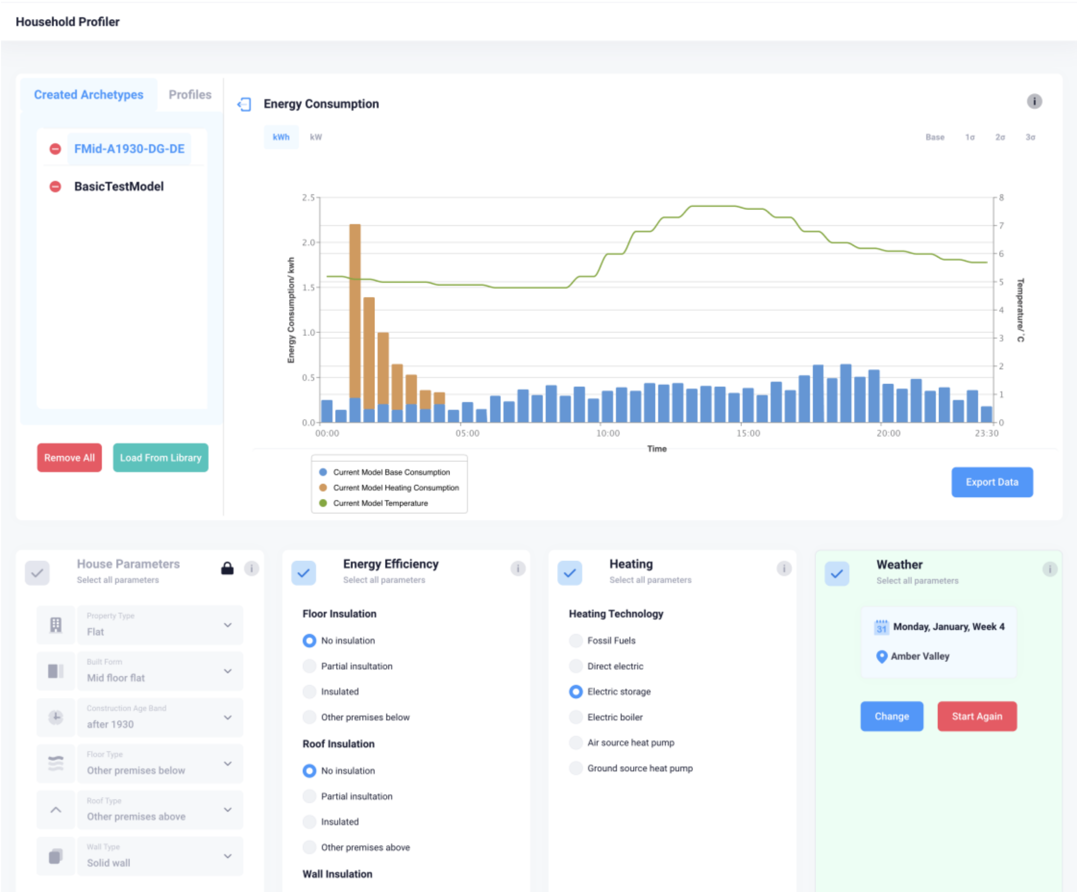


Figure 42 Result of applying electric storage as a heating type to a Mid-floor flat; this is to illustrate the charging decay function and how it is applied

An aspect of the storage heater model that could be considered is an expectation of seasonal differences if users have “intelligent” storage heaters or are using the “input” settings. Input settings instruct the storage heater to charge less as there is an anticipation that the weather will be mild for the upcoming day. There are output settings as well, however these will not play a strong role as compared to the input setting.

There is some natural consideration of this in the model, although it is not an explicit parameter. The natural consideration is the heat demand will be less on mild days and electricity consumption in the model only charges up to the required amount for the day. This would be the same as considering end users were perfectly setting their input and output of storage heaters. We know that in reality this is not the case, end users over and under charge their storage heaters.

9. Model validation – Household Profiler

A bottom up approach requires household profiles that can be combined for higher level analysis. The models that generate data should be a good match to real data. Given the models have parameters based on archetypes, validation of the models can be done with real home data selected by archetype.

If actual household level data was available it could be used to build up network level data. However, models would still be required to predict future scenarios.

Backtesting against a single real household provides an indication of suitability and demonstrates how this could be done at scale if more time was available to perform the error analysis.

For the random house that was selected, daily profiles for heating and electricity consumption are well matched.

To validate the model, it was first compared against a real property to sense check the units of measure and application of energy efficiency and heating technology. It is not expected that any individual home would exactly match the model as the model is representative of a population.

A second level of validation is done on an annual basis to check that no seasonal issues are present.

9.1. Random home selection

To protect the privacy of the home, the address will not be revealed, however it is located near the NGED service area in Wiltshire. This home is not in the data set that was used to train the model. The home has no solar panels, no EV and has gas heating.

A day of Thursday, 27 January 2022 was selected for comparison. The model selection parameters were selected as shown in Figure 43

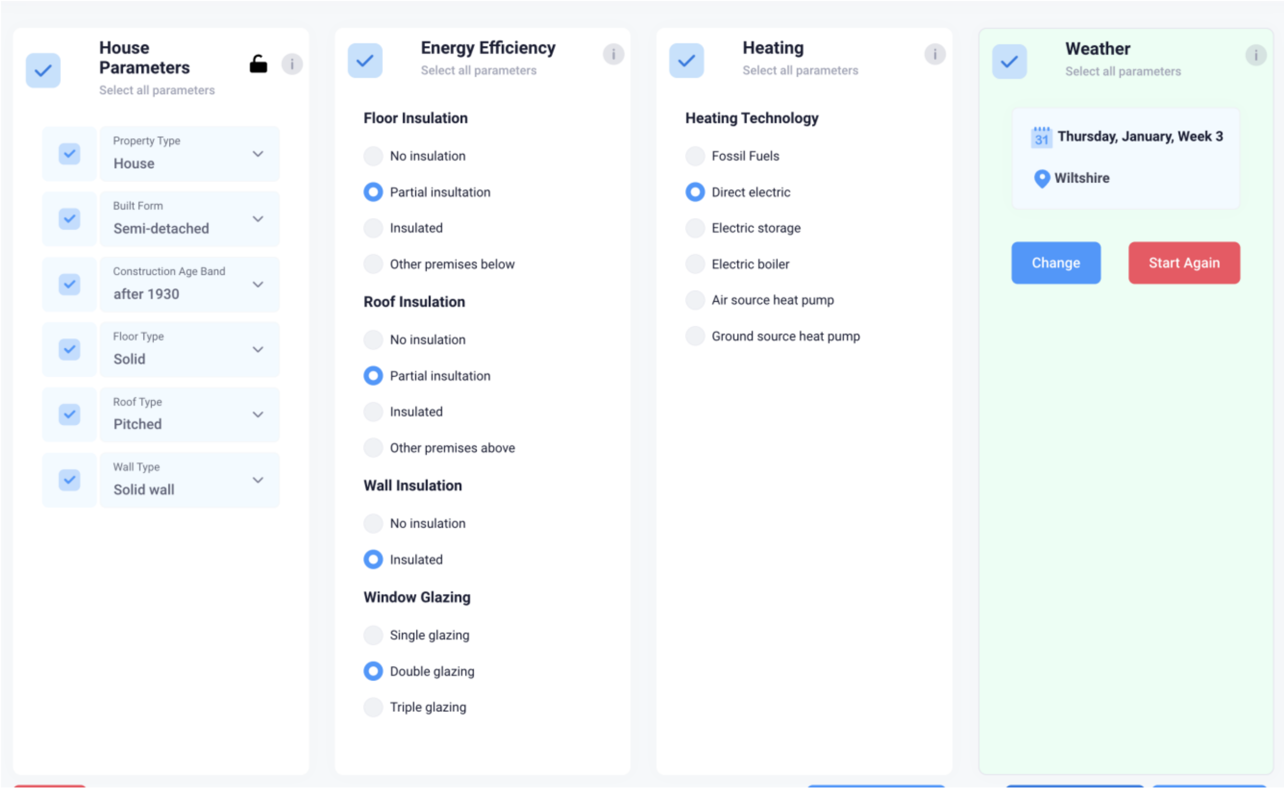


Figure 43 Model parameter selection, although direct electric is show in the figure, this is done to show the estimated heating energy demand from the model

9.2. Electricity comparison to model

There is a good match to the general pattern and scale of consumption between the observed home and what the model is generating.

There is an observed peak of electricity use at 14:30 of 1.923 kWh or 2.8kW running for that half hour. A few other days were examined and they too had a similar spike at around the same time in the afternoon. The household profiler is not expected to predict a single household with precision, so this outlier isn't a concern.

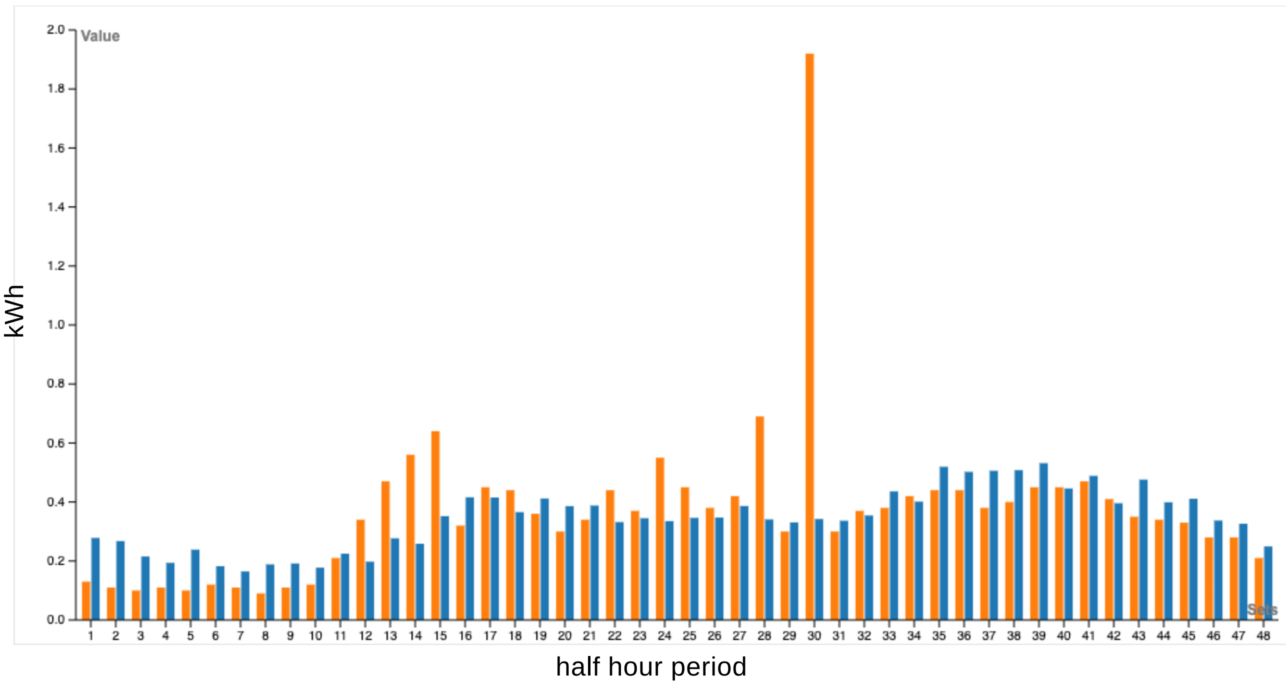


Figure 44 Real (orange) versus model (blue) electricity usage for Thursday, 27 January 2022

The model graph is shown as a screen shot in Figure 45 so that units and visual output can be seen. Note the units are in kWh for the half hour period and the temperature for the day is also on the graph.

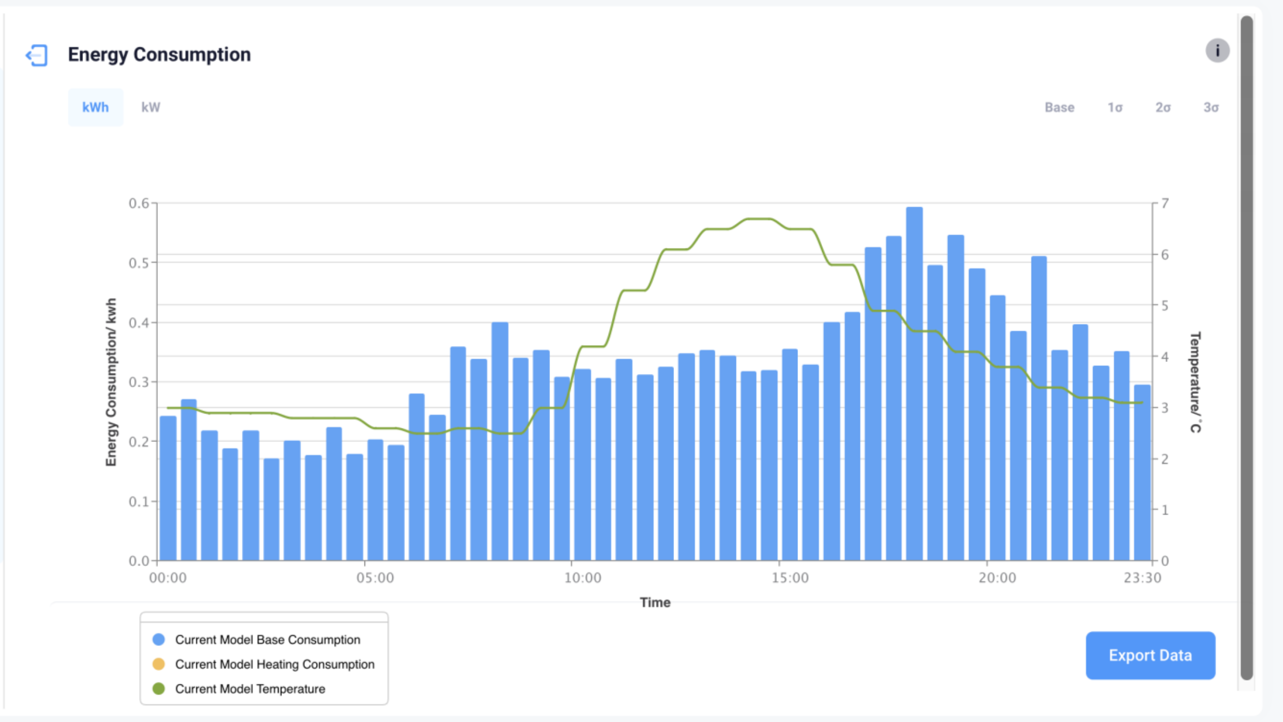


Figure 45 Model (blue) electricity usage for a Thursday, 3 week in January for a Semi-detached house, not the units are in kWh for the half hour

date,guid,name,consumption_type,units,00:00,00:30,01:00,01:30,02:00,02:30,03:00,03:30,04:00,04:30,05:00,05:30,06:00,06:30,07:00,07:30,08:00,08:30,09:00,09:30,10:00,10:30,11:00,11:30,12:00,12:30,13:00,13:30,14:00,14:30,15:00,15:30,16:00,16:30,17:00,17:30,18:00,18:30,19:00,19:30,20:00,20:30,21:00,21:30,22:00,22:30,23:00,23:30
Thursday/Week 3/January, undefined, Current Model, Base Consumption, kwh,
0.2781303536217646,0.2674990175904124,0.2152799613614906,0.19365892714513278,0.23846065796716712,0.18247934911531008,0.1645216123025242,0.1884075127179498,0.1914773913142574,0.17742045671130544,0.22479945624893946,0.19770305664410648,0.27684597036373704,0.2584186379215276,0.35206753210568914,0.4159544453582812,0.4152214697274664,0.3656291532314649,0.4117800188645516,0.38603926184292225,0.3878769686312421,0.33198753938866216,0.34494732765766783,0.3351477292327468,0.34609202149861934,0.3472570861664778,0.3863746063178234,0.3409133587043772,0.3307646807970424,0.34226930919542603,0.336453586257703,0.3543125751351941,0.43586561747689834,0.4012330593078076,0.5193509648792052,0.5022248491991582,0.5058734398471637,0.5079007036869791,0.5315023758881304,0.44578519436793385,0.48866459298787257,0.39578957844237717,0.4758257300948627,0.3989839707993199,0.4110946769599301,0.33723884043171437,0.3263713825464262,0.2493182222903138
Thursday/Week 3/January, undefined, Current Model, Heating Consumption, kwh,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
Thursday/Week 3/January, undefined, Current Model Temperature, Temperature, ˚C,
3,3,2.9,2.9,2.9,2.9,2.8,2.8,2.8,2.8,2.6,2.6,2.5,2.5,2.6,2.6,2.5,2.5,3,3,4.2,4.2,5.3,5.3,6.1,6.1,6.5,6.5,6.7,6.7,6.5,6.5,5.8,5.8,4.9,4.9,4.5,4.5,4.1,4.1,3.8,3.8,3.4,3.4,3.2,3.2,3.1,3.1

Figure 46 Export Data result of fossil fuel being selected

9.3. Heating demand comparison

The selected home had an estimated HTC of 345 W/K from the model. On the 27 of January 2022 the degree days were 15 degrees Celcius taken from a nearby weather station.

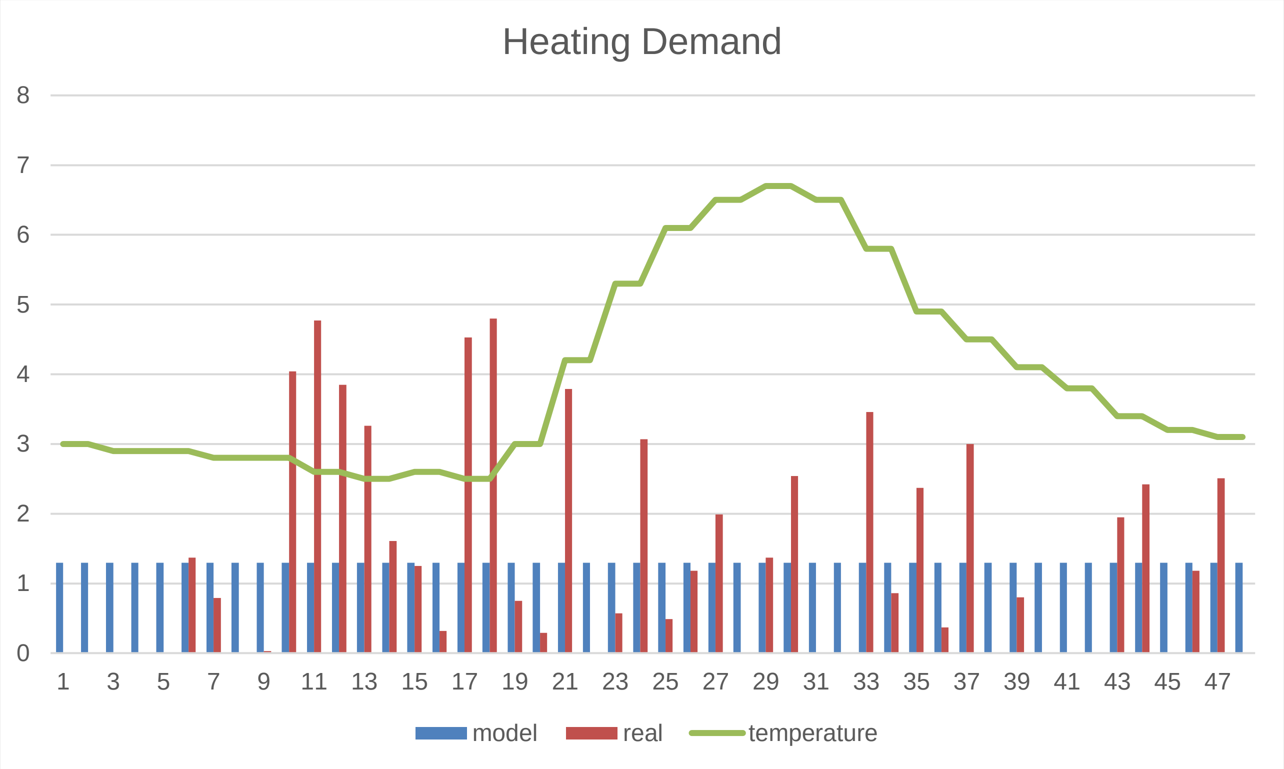


Figure 47 Model (blue) energy for heat and real gas usage (red) for a Thursday 27 Junary 2022, both in kWh units. Temperature is overlayed in green. A heat pump model is used for the simulated data.

The total actual energy used for heating on the day is 65.58 kWh with the modelled energy use as 62.16 kWh, or approximately 5% difference between the model and the real data. In Figure 47, the modelled energy consumption of the heating is a heat pump that is running continuously. The translation from gas to electrical heating assumes a co-efficient of performance for the heat pump and a change in operation profile.

9.4. Annual energy use comparison

The model was run for 365 days, to cover the period 1 November 2021 to 31 October 2022 for comparison to real data.

9.4.1. Base electricity consumption

Annual electricity consumption estimated by the model is 4,467.08 kWh and actual is 2,851.75 kWh. When looking at the model it is over estimating on a monthly basis in an even way. This would indicate that the population that we have trained the model with has generally higher consumption than this particular home although the trend and scale is within reason for seasons.

If comparing to UK averages for semi-detached homes, the model is also shown to be over estimating as ranges are from 2,000 to 4,200 kWh for annual electricity consumption where gas is the main heating source[11].

MONTH	ACTUAL ELECTRICITY (KWH)
NOV-21	219.20
DEC-21	256.85
JAN-22	247.92
FEB-22	206.86
MAR-22	198.83

[11] Energy Follow Up Survey: Household Energy Consumption & Affordability (2021), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1018725/efus-Household-Energy-Consumption-Affordability.pdf

APR-22	188.74
MAY-22	242.01
JUN-22	232.55
JUL-22	272.97
AUG-22	295.07
SEP-22	287.78
OCT-22	202.97
TOTAL	2,851.75

The peak power demand for this house on was 3.4kW for a short 30 minute burst, indicating some appliance use within the home. For every month this peak could be seen on many days as a stand out. The model is not very good at anticipating peak demands for a single household, so there is not an expectation that peaks such as this would reflect in data generated by the model. Peak power from the model is around 1.1kW. Given that there is a significant difference, it would be an enhancement to the model to add a peak power random variable and incorporate it into the model.

9.4.2. Heating

Annual heating energy calculated by the model is 10,118 kWh and actual for gas is 11,423 kWh which also includes gas for domestic hot water. The model is within an expected[11] range for gas consumption with annual figures cited between 8,000 and 16,600 kWh.

It is observed that shoulder months can be colder or warmer than expected. For instance, the model considers October to March, therefore a direct comparison of 9,166.91 (actual) and 10,118.283 (modelled) kWh would be more aligned.

MONTH	ACTUAL GAS (KWH)	ACUTAL GAS WINTER (KWH)
NOV-21	1,568.44	1,568.44
DEC-21	1,951.86	1,951.86
JAN-22	2,258.70	2,258.70
FEB-22	1,661.29	1,661.29
MAR-22	1,407.63	1,407.63
APR-22	900.01	
MAY-22	364.73	
JUN-22	282.26	
JUL-22	210.50	
AUG-22	220.53	
SEP-22	277.89	
OCT-22	318.99	318.99
TOTAL	11,422.83	9,166.91

10. Model validation – Network Planner

Network planning relies on the construction of scenarios defined by a mix of housing stock. Data is generated by the models for the individual houses and then combined as a network model. The Network Planner should be able to generate realistic feeder and substation level data given the same conditions.

Network planning level outputs are formatted at the tool level to be compatible with SINCAL. The model simply generates data for each of the households, with the tool tagging, grouping and aggregating the results for feeder and substation levels.

Three locations were selected in the National Grid licence areas where dataloggers captured network level electricity demand. Housing stock for the areas was compiled from network maps and coded into the Network Planner.

Backtesting for 5 test days within a year was done with a goal of replicating the shape and volume of electricity demand under different weather conditions.

The results have shown that model estimates are higher than the real data.

Five days were selected to validate the network planner model, with housing stock assumptions remaining the same across runs of the model. The housing stock was extracted by Carbon Trust using GIS data from NGED and then mapped to archetypes and subarchtypes. The input into the Glow Simulator was a list of archetypes and subarchetypes from a CSV file, the the Glow Simulator then compiling building counts.

Location	Building Count	Weather Station
Axbridge	252	Sedgemoor
WithycombeRaleigh	1,843	East Devon
Mackworth	3,363	Derby

Table 8 Reference data used for each of the scenarios supplied by NGED

The five days then have weather data that correspond to the location. The days were selected as Thursdays in order to make them comparable from a behavioural aspect. The December day was chosen as potentially a challenging day, knowing that close to Christmas and Covid cases meant many people would be home.

Label	Date	Average temperature across 3 sites	Notes
1	11 March 2021	8 C	Heating is factored in to this month
2	9 June 2021	NA (12.3C)	No heating month, therefore no heating demand, only base electricity
3	14 October 2021	13 C	Similar to an intermediate heating day.
4	23 Dec 2021	5.4 C	Challenge day
5	27 January 2022	4.5 C	Cold day

Table 9 Summary of five test days used

10.1. Results

Graphing the results of each day shows how well the shapes match and if there are any data quality issues. In all cases the Glow Simulator over estimated the energy.

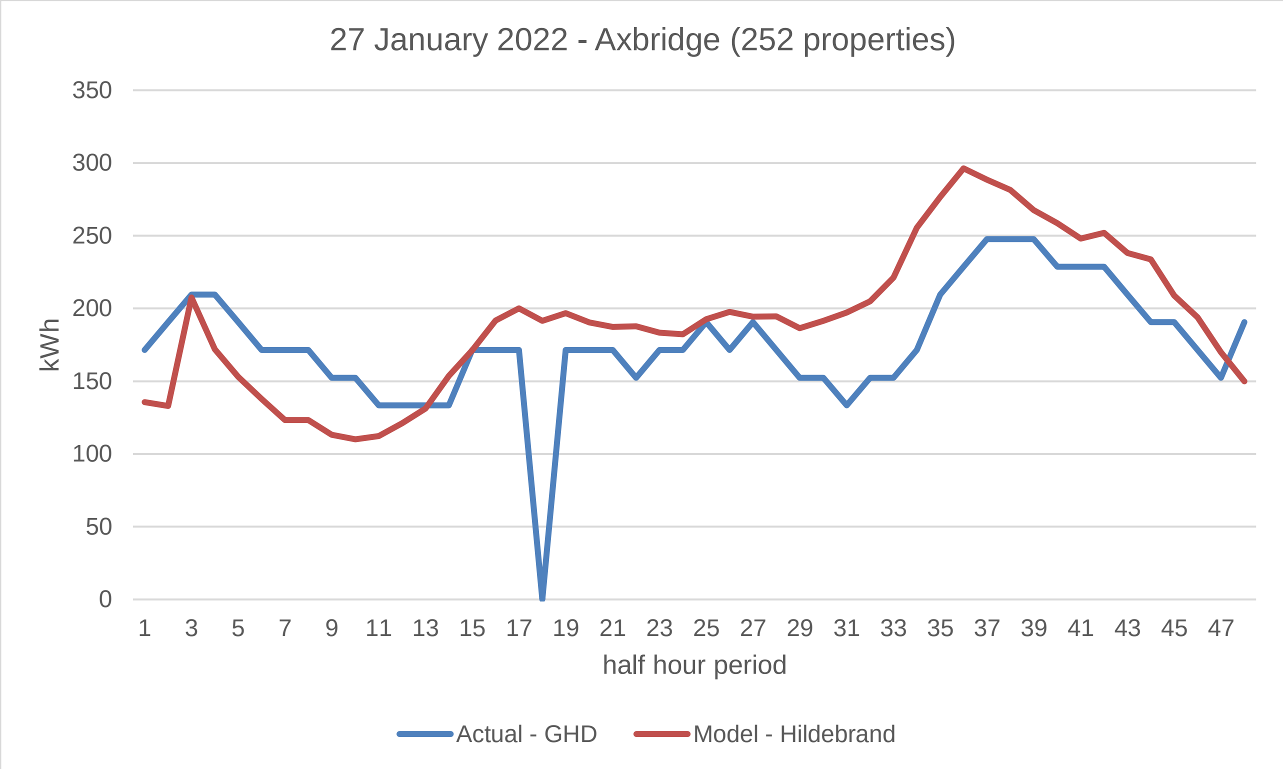


Figure 48 Comparison of model to actual electricity demand in kilowatts for all of Axbridge

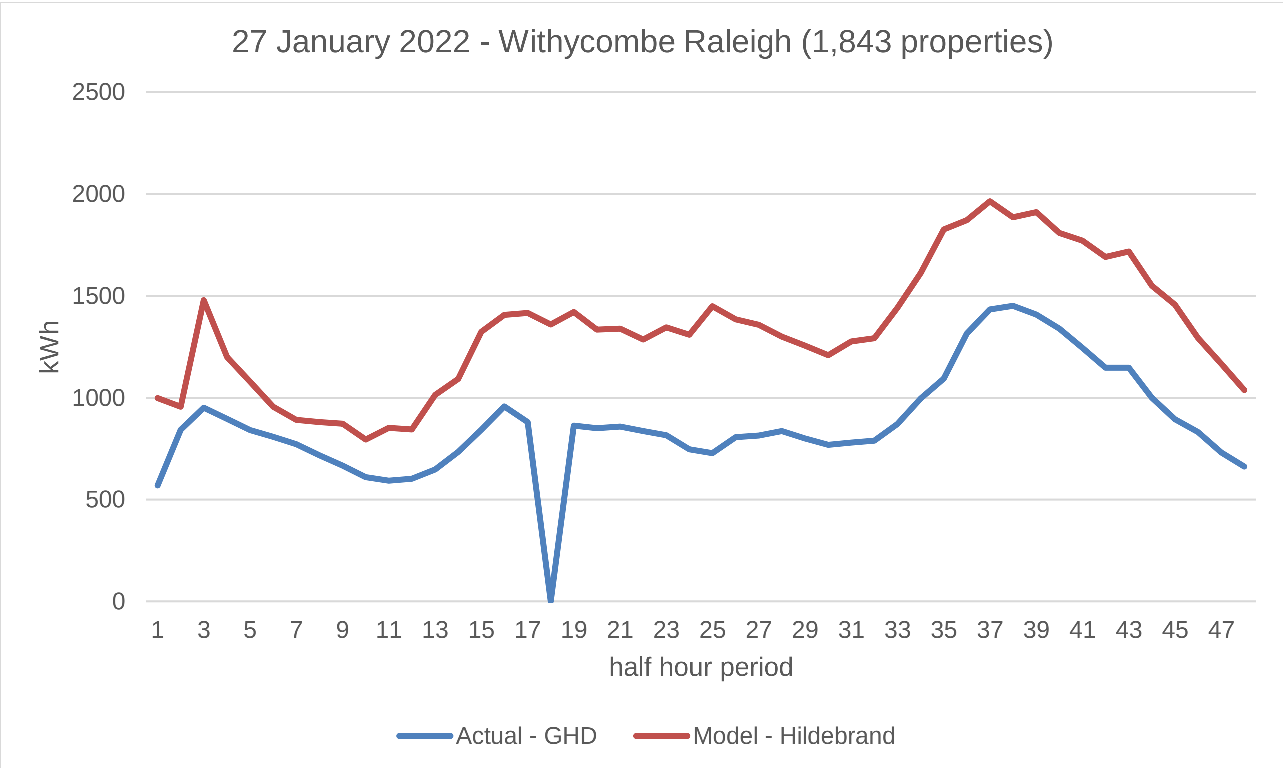


Figure 49 Comparison of model to actual electricity demand in kilowatts for all of Withycombe Raleigh

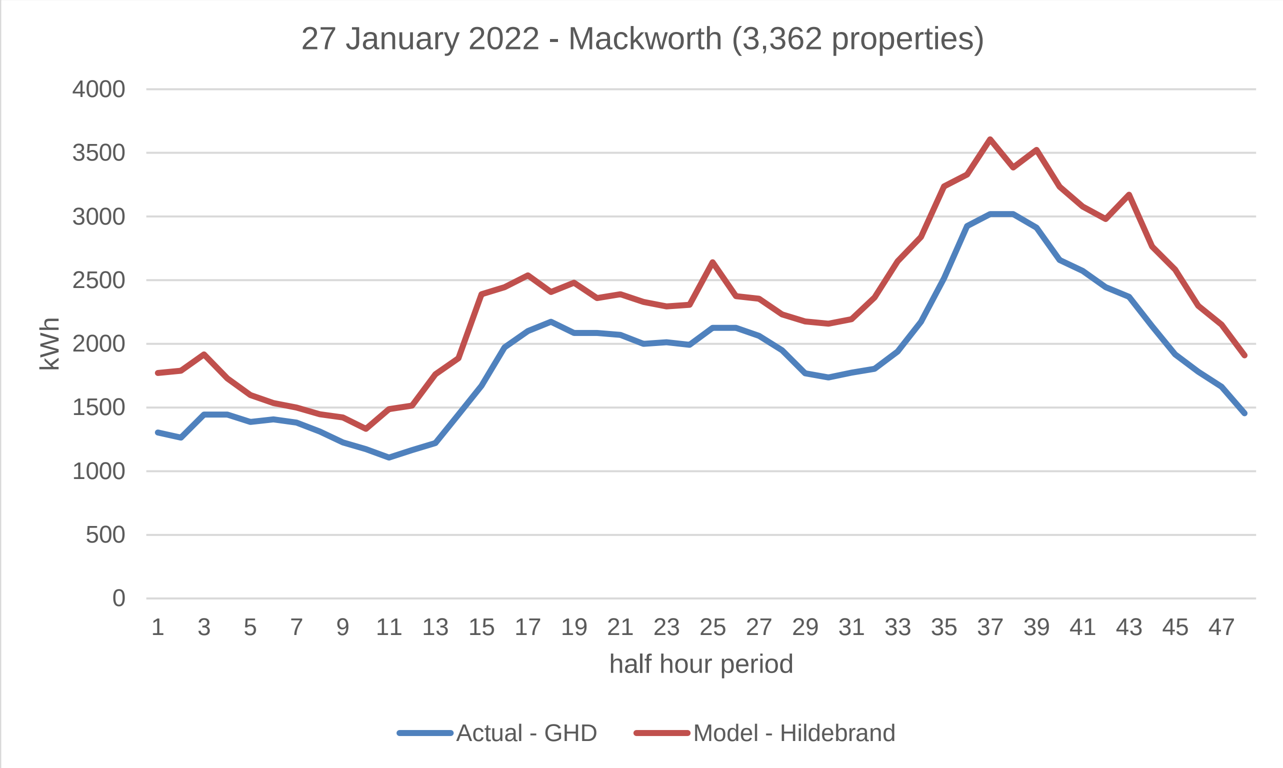
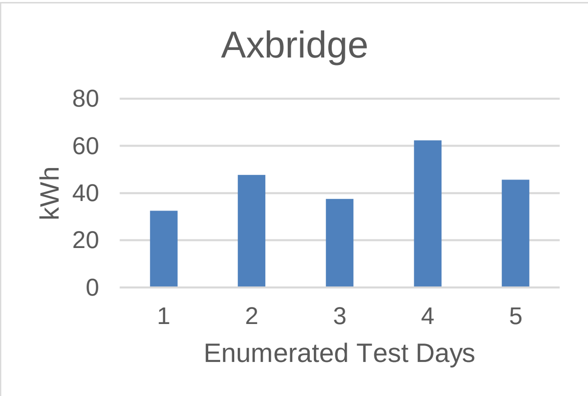
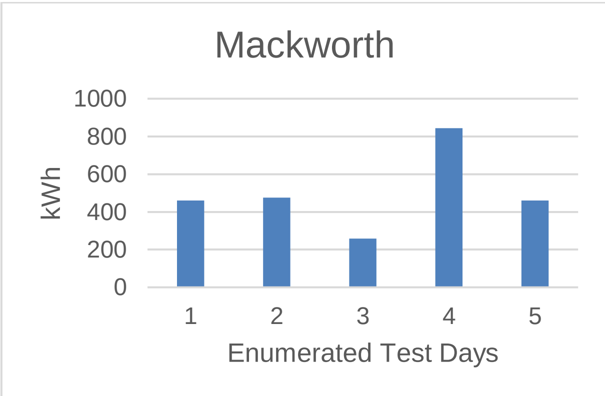


Figure 50 Comparison of model to actual electricity demand in kilowatts for all of Mackworth

Mean Squared Error

To characterise the fit between the actual and model data, a root mean squared error (RMSE) measurement was made between GHD provided data and model runs by Hildebrand. The spreadsheet in Appendix B contains all of the worked data.





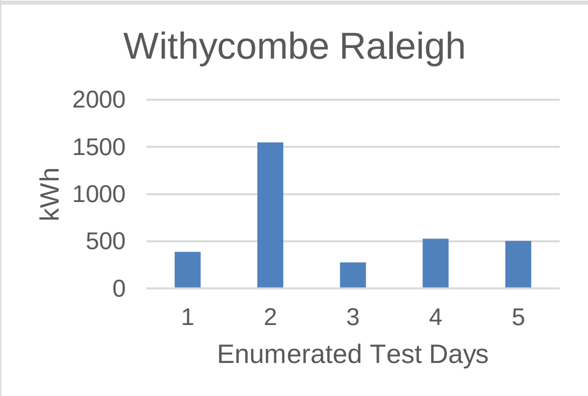


Figure 51 Root mean squared error over the five different test days, error is in kWh for the entire day. Days 1 (March) and 5 (January) are representative of heating days.

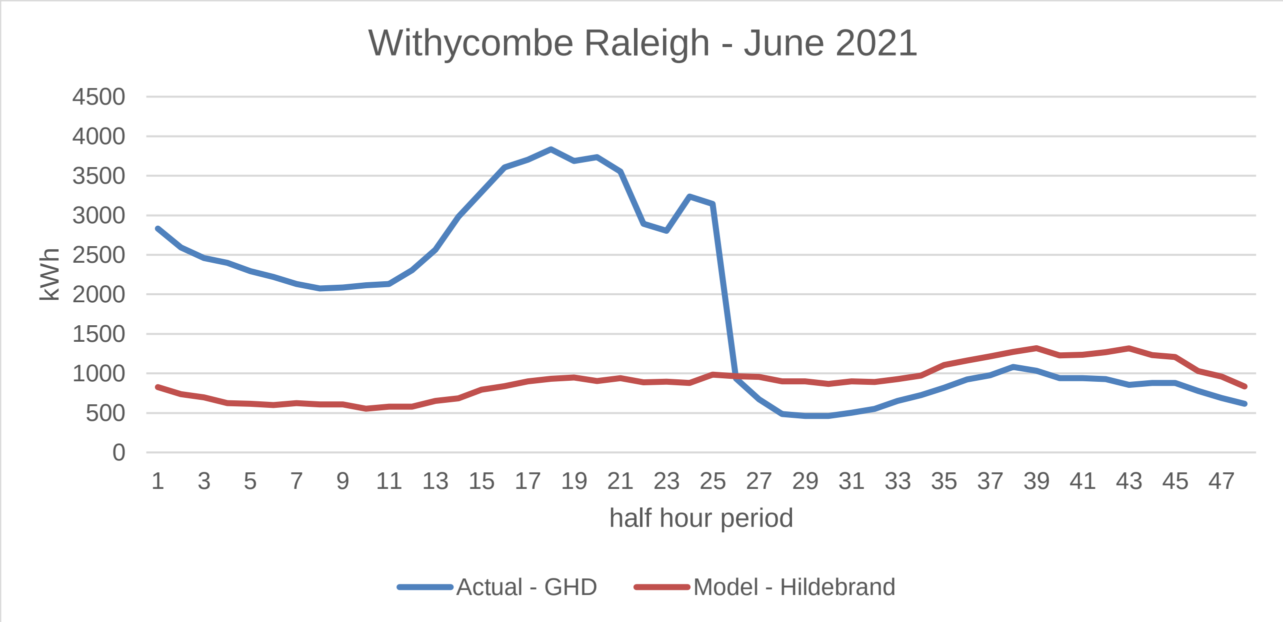


Figure 52 Large overnight load in June 2021 for Withycombe Raleigh which is an anomaly

The large difference in WR June (day 2) is due to some non-domestic load. This highlights the possible difference that could be found when using feeder level data.

Maximum Demand Analysis

To understand how well the model can predict the peak demand, the maximum demand for the 5 test days for both the model and the actual was analysed. There are anonmalies such as in Figure 52 where it appears some industrial load was present overnight.

An interpretation of large positive differences (the observed exceeding the model prediction) it is natural and due to an additional commercial load that happened to occur on that day, whereas negative differences are model errors. Small model errors could be due to natural variation in occupancy whereas larger errors indicate a structural issue with the model. In some cases it can be seen that there are data logger issues, such as Figure 49 where in half hour period 18 the logging drops out.

	Axbridge	Mackworth	Withycombe Raliegh
Mar-21	-8%	-13%	-22%
Jun-21	-19%	-29%	66%
Oct-21	-5%	-10%	-18%
Dec-21	36%	38%	-35%
Jan-22	-20%	-19%	-35%

Table 10. Percentage difference between the observed peak in overall electricity consumption and the maximum value the model predicted for the day. A positive number means the model result was lower than observed, whereas negative means the model was higher.

In the case of natural errors, improvements in accuracy could be made by understanding the cause of the additional load and incorporating it into the model.

In the case of model errors there are three possible sources to be explored, one is the selection of the housing stock itself i.e. better selection of what is connected to the feeder, the second is to look at the peak due to data logger issues and thirdly physical modelling of the network to single homes. Physical modelling would consider power factor and how 3-phase data logging would translate into single phase household modelling.

11. Conclusions

In summary the project has shown that a bottom-up approach is feasible and yields scalable results from household to distribution network and wider. In particular there are major findings that support the approach and minor improvements that would increase the number and type of users that could get benefit from the toolset.

11.1. Major findings

When looking at the samples from the Hildebrand data set and EPC data, there is a reduction of about 50% of the usable data set. This “discount” should be applied in future experimental design if recruitment is required. Once the statistical power is determined, recruiting double the amount of homes will ensure the data set is suitable.

From the Hildebrand data set that was available from April, some house types were under represented. It makes sense that certain house types may not easily be found, such as modern flats with gas. In these cases, additional work may need to be done to disaggregate heating loads from over all electricity consumption.

There appears to be a bias in Hildebrand data towards high users or users with solar, batteries etc. It may be possible to use more of a mass market data set or as the Hildebrand data set gets larger, filter properties to remove the bias.

Synthetic data is critical as no real data exists for future scenarios. Having parameter driven data also removes the need to expose individual’s data in the case of single dwelling analysis.

It is clear that new loads and time of use pricing signals significantly shift electricity demand and could easily result in new peak hours. Electric vehicles and night storage heaters placed very large loads on the network.

Gas data is critical to understand housing stock and forecast of heat demand. National Grid DSO normally would only have electricity data, the gas data is now available through the smart metering system.

House archetypes are critical to the analytic method as they reduce the number of dimensions and provide selection criteria for scenarios. However the EPC data must be transformed or cleaned so that it can be used to lookup house archetypes from a geographic boundary box.

Using the front-end tool would benefit from geographic selection built in. This would be a set of postcodes or interaction with a map user interface.

Scenarios are driven from the front-end tool. Adding another software layer and parameterising the scenario would enable more scenarios to be generated with the ability to study the effects of small changes in energy efficiency and heating technology.

11.2. Minor improvements

- The user interface for models could be easier to change scenarios and input.
- Move data to a database to share scenarios. This would improve the performance of the tool for large scenarios and keep a base model for each of the National Grid network segments so that they can be used for other analysis.
- Mathematical libraries should be more robustly tested. In some isolated cases, generating sample data showed oscillation which is probably due to the random number generation or issues with the sampling routines that are being used.

- Heating technology simulation was basic for this development work. Energy Plus should be integrated to make available many more heating technology choices and more realistic patterns of consumption.
- Noticed that base load electricity samples that are generated are not independently updating, this is likely a complex software bug where the simulation object is not being regenerated if no parameters are changed.

11.3. Next steps

Next steps for the models should be to make improvements and create a housing stock database for the whole of the service areas.

Improvements would rely on training the models with a larger data set. The data set is now at least 5 times larger than the one that was used. No major issues are foreseen in running the models with this new data, however there is considerable time that would go into extracting and preparing data for a new model run.

The execution of the model to generate data is done on the web browser. Moving the model execution to the server should give more control over sampling functions and allow for more detailed Bayesian calibration with subarchetype information.

The model only considered a single day in the profile generation. A study of the effects of longer periods such as prolonged cold weather should be done to understand the effects of preceding days weather.

Electricity baseload models were done using annualised averages. This could be improved. Some models to explore would be models that link half hours as random variables, such as a Hidden Markov Model; or try and establish Bayesian model like was done for gas, based on occupancy or appliances as the hidden variable.

Maximum demand for home should be considered as a part of the model. This would allow for better fit to network analysis over summed daily energy consumption. Using a machine learning technique the time of day and maximum demand of an appliance for each household would add variation to generated data, consistent with EV loads, whereas they were removed from this analysis.

Detailed research into physical modelling of the network to see if there are better assumptions for how single phase smart meter data corresponds to feeder level logged data, for instance is power factor a material consideration.

12. Appendix A

Frequency of property characteristics for the NGED/NationalGrid service area from the EPC database. A total of 861,872 properties has current EPCs that could be used. The nomenclature used for Property type, Built form, etc was recoded to archetypes specficied by Carbon Trust's Options Tool.

Archetype	Smart meter group	Property type	Built form	Construction age band	Wall type	Floor type	Roof type	COUNT in sample	%
1	A	House	Semi-Detached	before 1930	Solid wall	Suspended	Pitched	33457	3.9%
2	A	House	Semi-Detached	before 1930	Solid wall	Solid	Pitched	21961	2.5%
3	A	House	Semi-Detached	before 1930	Solid wall	Suspended	Flat	176	0.0%
4	A	House	Semi-Detached	before 1930	Solid wall	Solid	Flat	209	0.0%
5	A	House	Semi-Detached	before 1930	Cavity wall	Suspended	Pitched	2294	0.3%
6	A	House	Semi-Detached	before 1930	Cavity wall	Solid	Pitched	1526	0.2%
7	A	House	Semi-Detached	before 1930	Cavity wall	Suspended	Flat	91	0.0%
8	A	House	Semi-Detached	before 1930	Cavity wall	Solid	Flat	103	0.0%
9	B	House	Semi-Detached	after 1930	Solid wall	Suspended	Pitched	43218	5.0%
10	B	House	Semi-Detached	after 1930	Solid wall	Solid	Pitched	32831	3.8%
11	B	House	Semi-Detached	after 1930	Solid wall	Suspended	Flat	173	0.0%
12	B	House	Semi-Detached	after 1930	Solid wall	Solid	Flat	430	0.0%
13	B	House	Semi-Detached	after 1930	Cavity wall	Suspended	Pitched	51598	6.0%
14	B	House	Semi-Detached	after 1930	Cavity wall	Solid	Pitched	123236	14.3%
15	B	House	Semi-Detached	after 1930	Cavity wall	Suspended	Flat	384	0.0%
16	B	House	Semi-Detached	after 1930	Cavity wall	Solid	Flat	1119	0.1%
17	C	House	Mid-Terrace	before 1930	Solid wall	Suspended	Pitched	63746	7.4%
18	C	House	Mid-Terrace	before 1930	Solid wall	Solid	Pitched	27400	3.2%
19	C	House	Mid-Terrace	before 1930	Solid wall	Suspended	Flat	296	0.0%
20	C	House	Mid-Terrace	before 1930	Solid wall	Solid	Flat	255	0.0%
21	C	House	Mid-Terrace	before 1930	Cavity wall	Suspended	Pitched	1659	0.2%
22	C	House	Mid-Terrace	before 1930	Cavity wall	Solid	Pitched	722	0.1%
23	C	House	Mid-Terrace	before 1930	Cavity wall	Suspended	Flat	155	0.0%
24	C	House	Mid-Terrace	before 1930	Cavity wall	Solid	Flat	106	0.0%
25	D	House	Mid-Terrace	after 1930	Solid wall	Suspended	Pitched	11702	1.4%
26	D	House	Mid-Terrace	after 1930	Solid wall	Solid	Pitched	16134	1.9%
27	D	House	Mid-Terrace	after 1930	Solid wall	Suspended	Flat	51	0.0%
28	D	House	Mid-Terrace	after 1930	Solid wall	Solid	Flat	418	0.0%
29	D	House	Mid-Terrace	after 1930	Cavity wall	Suspended	Pitched	15565	1.8%
30	D	House	Mid-Terrace	after 1930	Cavity wall	Solid	Pitched	46886	5.4%
31	D	House	Mid-Terrace	after 1930	Cavity wall	Suspended	Flat	114	0.0%
32	D	House	Mid-Terrace	after 1930	Cavity wall	Solid	Flat	1077	0.1%
33	E	House	Detached	before 1930	Solid wall	Suspended	Pitched	6694	0.8%
34	E	House	Detached	before 1930	Solid wall	Solid	Pitched	13463	1.6%

35	E	House	Detached	before 1930	Solid wall	Suspended	Flat	64	0.0%
36	E	House	Detached	before 1930	Solid wall	Solid	Flat	173	0.0%
37	E	House	Detached	before 1930	Cavity wall	Suspended	Pitched	1032	0.1%
38	E	House	Detached	before 1930	Cavity wall	Solid	Pitched	923	0.1%
39	E	House	Detached	before 1930	Cavity wall	Suspended	Flat	43	0.0%
40	E	House	Detached	before 1930	Cavity wall	Solid	Flat	43	0.0%
41	F	House	Detached	after 1930	Solid wall	Suspended	Pitched	8621	1.0%
42	F	House	Detached	after 1930	Solid wall	Solid	Pitched	6512	0.8%
43	F	House	Detached	after 1930	Solid wall	Suspended	Flat	114	0.0%
44	F	House	Detached	after 1930	Solid wall	Solid	Flat	107	0.0%
45	F	House	Detached	after 1930	Cavity wall	Suspended	Pitched	28137	3.3%
46	F	House	Detached	after 1930	Cavity wall	Solid	Pitched	78715	9.1%
47	F	House	Detached	after 1930	Cavity wall	Suspended	Flat	206	0.0%
48	F	House	Detached	after 1930	Cavity wall	Solid	Flat	526	0.1%
49	G	Flat	Top floor flat	before 1930	Solid wall	Other premises below	Pitched	18489	2.1%
50	G	Flat	Top floor flat	before 1930	Solid wall	Other premises below	Flat	661	0.1%
51	G	Flat	Top floor flat	before 1930	Cavity wall	Other premises below	Pitched	641	0.1%
52	G	Flat	Top floor flat	before 1930	Cavity wall	Other premises below	Flat	41	0.0%
53	H	Flat	Top floor flat	after 1930	Solid wall	Other premises below	Pitched	11506	1.3%
54	H	Flat	Top floor flat	after 1930	Solid wall	Other premises below	Flat	4016	0.5%
55	H	Flat	Top floor flat	after 1930	Cavity wall	Other premises below	Pitched	43005	5.0%
56	H	Flat	Top floor flat	after 1930	Cavity wall	Other premises below	Flat	6083	0.7%
57	I	Flat	Mid floor flat	before 1930	Solid wall	Other premises below	Other premises above	11804	1.4%
58	I	Flat	Mid floor flat	before 1930	Cavity wall	Other premises below	Other premises above	220	0.0%
59	J	Flat	Mid floor flat	after 1930	Solid wall	Other premises below	Other premises above	22853	2.7%
60	J	Flat	Mid floor flat	after 1930	Cavity wall	Other premises below	Other premises above	27677	3.2%
61	K	Flat	Bottom floor flat	before 1930	Solid wall	Suspended	Other premises above	7246	0.8%
62	K	Flat	Bottom floor flat	before 1930	Solid wall	Solid	Other premises above	7501	0.9%
63	K	Flat	Bottom floor flat	before 1930	Cavity wall	Suspended	Other premises above	502	0.1%
64	K	Flat	Bottom floor flat	before 1930	Cavity wall	Solid	Other premises above	287	0.0%
65	L	Flat	Bottom floor flat	after 1930	Solid wall	Suspended	Other premises above	2455	0.3%
66	L	Flat	Bottom floor flat	after 1930	Solid wall	Solid	Other premises above	6856	0.8%
67	L	Flat	Bottom floor flat	after 1930	Cavity wall	Suspended	Other premises above	7759	0.9%
68	L	Flat	Bottom floor flat	after 1930	Cavity wall	Solid	Other premises above	37805	4.4%

13. Appendix B

Illustrations of electricity profiles for various households. In particular they show:

- normal usage (Figure 53) – a variety of energy spikes normally occurring during waking hours – **10%** night / total usage
- large winter season overnight use (Figure 54) – spikes of energy due to storage heaters – **77%** night / total usage
- large all year round overnight use (Figure 55) – spikes of energy due to EV - **65%** night /total usage

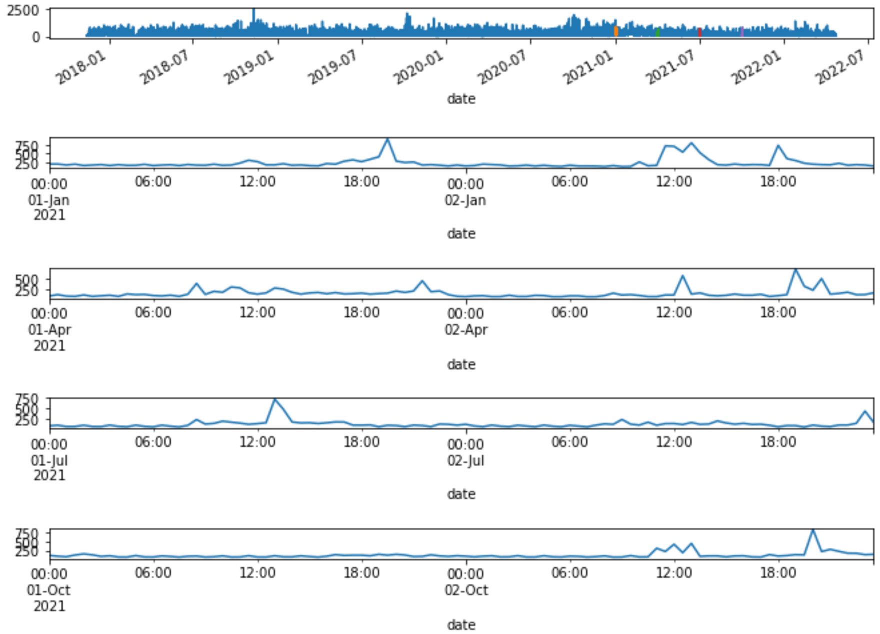


Figure 53 Normal energy patterns for a home without EV, solar or night storage; predominately showing activity during the day time

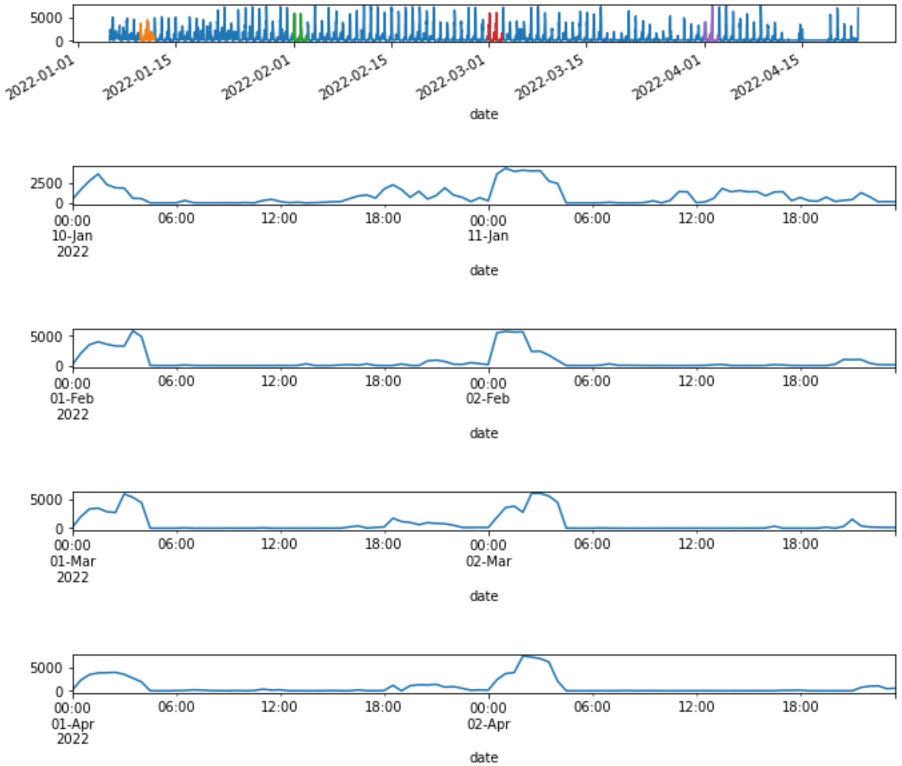


Figure 54 One of the homes where night storage was detected, notice the over night consumption that occur mostly in the colder months of the year

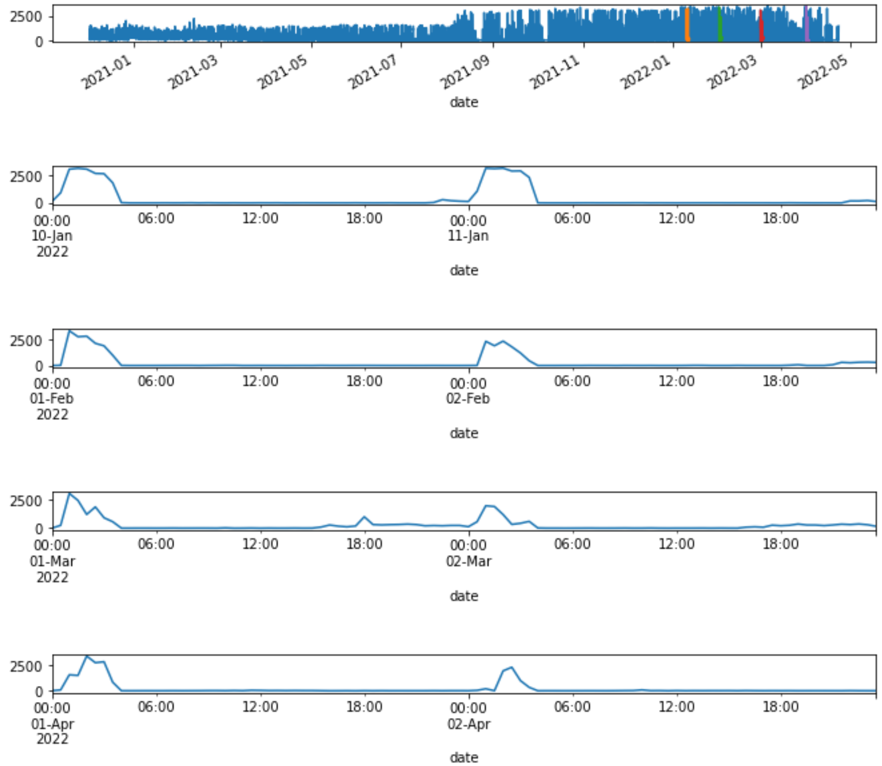
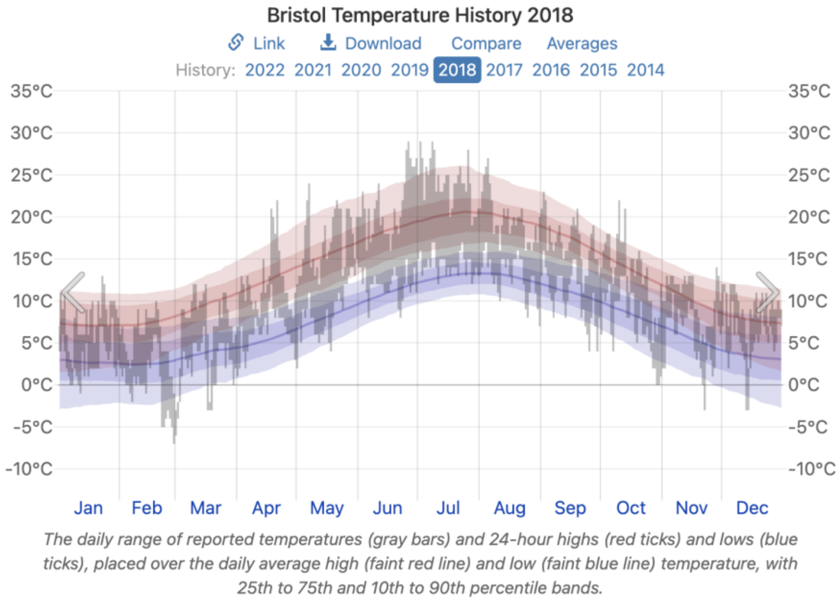


Figure 55 Suspected purchase of an EV in August of 2021

14. Appendix C - Extreme Days

The extreme day of 28 February 2018 was used and as can be considered roughly a 1 in 20 year of extremes as it falls out of the 95 percentile of weather patterns. The days leading up to the extremely cold weather were not considered. It may be a future topic to explore the impact of depleted thermal stores.



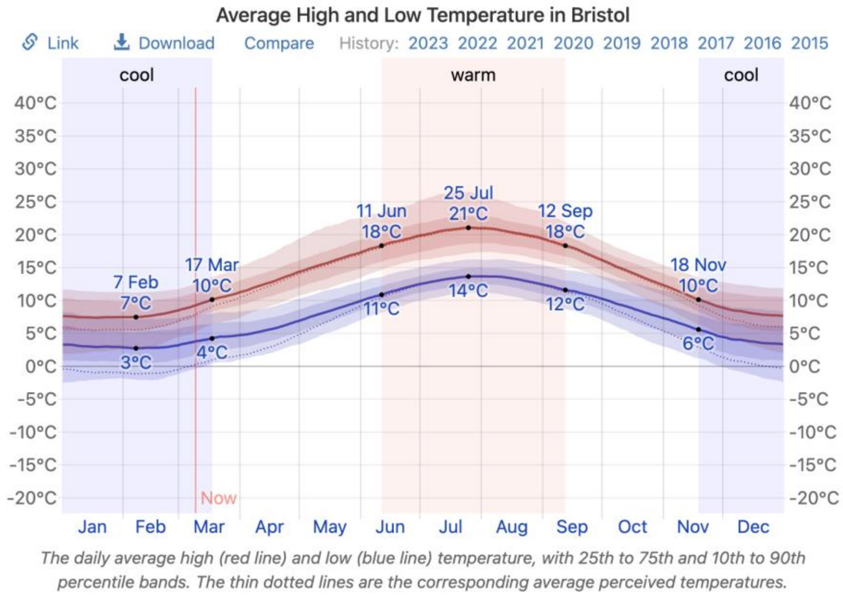


Figure 56. Historical weather for Bristol, left graph shows 2018 with February 28th weather extremely cold; historical norms for all years is shown on the right for comparison.

15. Appendix D – Mean gas consumption

The following table is the mean gas consumption by half hour for each of the archetypes.

	Flat			House			
built_form	Bottom floor flat	Mid floor flat	Top floor flat	Detached	Mid-Terrace	NO DATA!	Semi-Detached
timeofday							
00:00:00	271.771761	133.709760	239.007317	212.774580	179.744825	47.277325	208.416031
00:30:00	263.417050	150.022432	223.585106	213.788067	170.174730	40.290563	190.318649
01:00:00	253.371085	146.413283	203.218065	214.287345	164.879446	43.732179	191.451675
01:30:00	253.214719	165.011628	207.548288	228.012948	170.258743	44.281738	197.608636
02:00:00	234.120364	174.222507	217.659135	258.028653	189.174829	47.807536	210.839410
02:30:00	245.317026	173.880765	226.625979	282.465893	193.956747	48.709437	223.517015
03:00:00	255.466239	227.154513	240.404735	348.879382	231.388173	58.355058	263.588922
03:30:00	272.470935	231.997635	279.189043	451.142020	286.304510	108.828979	317.021377
04:00:00	349.960119	286.755814	425.634764	705.065778	385.847024	130.524262	451.643238
04:30:00	390.710821	318.153695	480.036837	1027.109818	538.382363	149.666893	628.509565
05:00:00	500.724538	411.270095	630.762770	1523.470265	762.907825	423.683407	934.123736
05:30:00	587.616721	553.594563	685.588956	1874.193841	1029.774933	635.823270	1191.171691
06:00:00	681.496648	750.983452	976.535100	2215.026571	1205.387533	835.687352	1450.743237
06:30:00	786.369513	476.619460	1000.833050	2226.723033	1279.902903	1394.101729	1557.275191
07:00:00	862.411274	490.942277	815.745276	2070.344467	1308.792037	1382.754237	1509.654487
07:30:00	776.137636	414.991529	797.007321	1783.615931	1100.358420	1533.367797	1325.586029
08:00:00	773.761921	499.458432	739.918620	1517.012889	968.099342	1213.966102	1147.961076
08:30:00	686.571470	333.212766	663.680590	1254.601565	837.178318	963.844068	978.965473
09:00:00	630.845302	365.294326	597.404994	1099.073529	747.331109	802.149831	862.983275
09:30:00	593.418954	350.379236	583.155675	972.889943	684.300706	598.197694	783.147380

	Flat			House			
built_form	Bottom floor flat	Mid floor flat	Top floor flat	Detached	Mid-Terrace	NO DATA!	Semi-Detached
timeofday							
10:00:00	555.546048	347.708259	554.539981	897.987288	636.176559	433.710312	723.807000
10:30:00	516.773037	338.742362	543.975085	863.465635	624.624092	336.212687	683.777309
11:00:00	500.663811	336.840725	538.751561	875.362320	608.842230	309.213365	683.953044
11:30:00	508.313087	331.832446	562.153148	842.834894	613.155558	203.295455	651.303104
12:00:00	504.160585	347.764683	553.975763	870.315575	623.783843	252.898236	672.977504
12:30:00	466.286505	338.295428	501.408786	807.076422	578.723646	180.265105	634.612958
13:00:00	452.271694	330.546315	485.583882	794.821796	564.485025	171.038357	627.809647
13:30:00	453.308253	335.214539	493.788808	804.583564	559.612622	165.262729	636.106798
14:00:00	456.780607	328.482270	486.570042	871.539664	565.074507	159.980991	660.024890
14:30:00	466.207377	321.136104	470.712623	969.857097	601.344097	197.936864	701.866972
15:00:00	496.326614	465.079575	608.074185	1163.703164	676.051657	226.550085	826.470746
15:30:00	504.751719	383.623991	683.047673	1333.189044	785.371114	360.791171	934.119880
16:00:00	554.723873	433.090193	750.167149	1595.345919	938.395305	485.435993	1094.739712
16:30:00	578.173715	420.672115	842.415744	1653.197958	1015.189416	881.078098	1162.752278
17:00:00	701.159958	408.602403	840.585471	1714.507116	1084.748697	1081.432937	1267.401684
17:30:00	705.375779	407.733753	798.366345	1649.487820	1073.653744	808.242784	1223.179268
18:00:00	719.599091	410.645530	867.981555	1592.331875	1063.045275	756.707980	1220.837788
18:30:00	724.885355	413.987003	788.694023	1483.799442	1024.604173	615.847538	1159.660684
19:00:00	720.641093	413.305238	783.422669	1386.332265	950.043579	575.692699	1094.354042
19:30:00	706.932949	382.731246	726.697826	1271.973887	876.190500	514.253311	1006.337184
20:00:00	672.518875	424.923804	678.008911	1163.039975	794.100494	417.138200	922.389361
20:30:00	630.969824	371.308722	643.093014	1037.896954	725.928994	379.107640	830.668094

	Flat			House			
built_form	Bottom floor flat	Mid floor flat	Top floor flat	Detached	Mid-Terrace	NO DATA!	Semi-Detached
timeofday							
21:00:00	606.192987	336.217563	623.335811	871.586090	641.980888	358.301291	702.966387
21:30:00	582.619023	305.189801	574.822922	703.017559	543.965321	265.471647	579.275016
22:00:00	452.337100	237.556387	532.839773	452.324261	398.981782	130.093718	396.352221
22:30:00	439.345641	228.359307	455.322024	367.529049	326.570989	99.269179	334.915741
23:00:00	339.969718	176.012593	347.637808	247.993207	216.855341	55.670401	237.175755
23:30:00	297.337933	173.566903	275.970046	236.292421	206.338875	54.855397	220.458344

16. Appendix E – Bayesian classification results

The following tables are the traces and metrics that were produced for the inference step to estimate HTC based on degree days and energy on a daily basis. Each table summaries the run for the archetype. The HTC gamma distribution is then used within the generative steps of the simulator.

16.1. Detached House, after 1930 construction

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.510	0.001	0.508	0.512	0.000	0.000	976.0	1047.0	1.0
sigma	39.133	0.118	38.899	39.352	0.004	0.003	970.0	1102.0	1.0

16.2. Detached House, before 1930 construction

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.575	0.004	0.567	0.582	0.000	0.000	963.0	1424.0	1.00
sigma	48.465	0.394	47.753	49.212	0.016	0.011	620.0	942.0	1.01

16.3. Semi detached House, after 1930 construction

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.371	0.001	0.369	0.372	0.000	0.000	853.0	1089.0	1.0
sigma	29.356	0.099	29.165	29.534	0.004	0.003	726.0	1126.0	1.0

16.4. Semi detached House, before 1930 construction

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.495	0.003	0.490	0.500	0.000	0.000	1217.0	1368.0	1.00

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
sigma	39.489	0.281	39.004	40.057	0.011	0.008	668.0	626.0	1.01

16.5. Semi detached House, after 1930 construction

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.302	0.001	0.300	0.305	0.000	0.000	863.0	1074.0	1.0
sigma	25.536	0.170	25.246	25.880	0.006	0.004	891.0	1162.0	1.0

16.6. Semi detached House, before 1930 construction

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.346	0.002	0.342	0.349	0.00	0.000	568.0	1084.0	1.0
sigma	31.127	0.232	30.678	31.574	0.01	0.007	577.0	733.0	1.0

16.7. Bottom floor flat, before 1930 construction

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.357	0.005	0.348	0.368	0.00	0.000	1100.0	1007.0	1.0
sigma	20.300	0.382	19.619	20.993	0.01	0.007	1583.0	1560.0	1.0

16.8. Bottom floor flat, after 1930 construction

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.188	0.003	0.183	0.192	0.000	0.000	802.0	1167.0	1.0
sigma	20.426	0.315	19.847	21.026	0.011	0.008	782.0	1030.0	1.0

16.9. Mid floor flats (including before and after 1930 construction)

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.181	0.004	0.174	0.189	0.000	0.000	924.0	1146.0	1.0
sigma	18.179	0.401	17.452	18.967	0.013	0.009	1001.0	1264.0	1.0

16.10. Top floor flat, after 1930 construction

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.235	0.003	0.230	0.240	0.000	0.000	957.0	1119.0	1.0
sigma	18.592	0.281	18.079	19.114	0.009	0.006	1018.0	1182.0	1.0

16.11. Top floor flat, before 1930 construction

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
htc	0.344	0.004	0.336	0.352	0.00	0.000	1161.0	1236.0	1.0
sigma	23.368	0.371	22.644	24.037	0.01	0.007	1305.0	1222.0	1.0