Article:

Sousa, G., Jones, B.M., Mirzaei, P.A. et al. (1 more author) (2018) An open-source simulation platform to support the formulation of housing stock decarbonisation strategies. Energy and Buildings, 172. pp. 459-477. ISSN 0378-7788

https://doi.org/10.1016/j.enbuild.2018.05.015

An open-source simulation platform to support the formulation of housing stock decarbonisation strategies

Gustavo Sousa[a,b,*], Benjamin M. Jones[b], Parham A. Mirzaei[b], Darren Robinson[a]

[a] *Sheffield School of Architecture, University of Sheffield, S10 2TN*
[b] *Department of Architecture and Built Environment, University of Nottingham, NG7 2RD*

Abstract

Housing Stock Energy Models (HSEMs) play a determinant role in the study of strategies to decarbonise the UK housing stock. Over the past three decades, a range of national HSEMs have been developed and deployed to estimate the energy demand of the 27 million dwellings that comprise the UK housing stock. However, despite ongoing improvements in the fidelity of both modelling strategies and calibration data, their longevity, usability and reliability have been compromised by a lack of modularity and openness in the underlying algorithms and calibration data sets. To address these shortfalls, a new open and modular platform for the dynamic simulation of national (in the first instance, the UK) housing stocks has been developed—the *housing stock Energy Hub (EnHub)*. This paper describes EnHub's architecture, its underlying rationale, the datasets it employs, its current scope, examples of its application, and plans for its further development. In this we pay particular attention to the systematic identification of housing archetypes and their corresponding attributes to represent the stock. The scenarios we analyse in our initial applications of EnHub, based on these archetypes, focus

*Corresponding author

on improvements to housing fabric, the efficiency of lights and appliances and of the related behavioural practices of their users. In this we consider a perfect uptake scenario and a conditional (partial) uptake scenario. Results from the disaggregation of energy use throughout the stock for the baseline case and for our scenarios indicate that improvements to solid wall and loft thermal performance are particularly effective, as are reductions in infiltration. Improvements in lights and appliances and reductions in the intensity of their use are largely counteracted by increases in heating demand. Housing archetypes that offer the greatest potential savings are apartments and detached dwellings, owing to their relatively high surface area to volume ratio; in particular for pre-1919 and inter-war epochs.

Keywords: housing stock, dynamic energy simulation, open-source, modularity, policy support

1. Introduction

The UK's Climate Change Act aims to reduce the 1990 Greenhouse Gases (GHG) emissions level by 80 % by 2050 [1]. To this end, the Committee on Climate Change (CCC) has established a series of incremental targets (or budgets) for the whole energy sector, including the production of 30 % of electricity from renewable sources by 2020, and the reduction of GHG emissions by 50 % by 2025. The UK emitted a total of 564 $MtCO_{2e}$ in 2011, which is 36% below the peak value registered in 1979 and 28 % below that of 1990 [2]. This reduction was mainly caused by a shift from coal to natural gas, by a displacement of industrial activity (primarily to Asia), and by major improvements in the performance of the transport sector [3]. This means that even though the reduction in this period is close to the CCC target, this has largely been achieved in the absence of systematic structural improvements to reduce energy demands. Some of the more significant opportunities for demand reduction are found in the domestic sector, where emissions have been maintained at almost the same level since 1990 [2].

In 2011, the domestic sector contributed 124 $MtCO_{2e}$ to the total emissions; two-fifths of these were caused by the generation of electrical energy in power stations and the remainder by direct combustion of fossil fuels. End-use energy demand in the domestic sector is attributed to four key services: 60% to space heating, 20% to domestic hot water (DHW), 17% to lighting and appliances, and 3% to cooking [4]. This highlights the importance of thermal energy flows in the development of Housing Stock Energy Models (HSEMs).

Efforts have been made to improve the performance of the existing housing stock, but these have mostly been counteracted by the construction of new buildings, serving a larger population, comprised of smaller households. For this reason, a full understanding of the energy flow in dwellings, and the factors influencing them, is required to formulate robust policies and strategies [5, 6] to achieve significant reductions in their carbon emission intensity. This requires further efforts on two fronts. On the one hand, the disaggregated measurement of end-use energy demand to complement existing surveys of housing characteristics for a representative sample of archetypes; and on the other, the formulation and calibration of suitable HSEMs, describing not only the performance of the existing stock, but also how this stock is likely to evolve in response to policy measures designed to reduce carbon intensity [7, 8].

In their recent review, Sousa et al. [9] systematically evaluated, using a detailed matrix characterising their functionalities, usability and accessibility, the attributes of the 29 HSEMs that have hitherto been developed and deployed in the UK. From this they identified the Cambridge Housing Model (CHM) as being the most fully developed. They also concluded that a) the models should be transparent so that their underlying algorithms can be understood and evaluated, and be amenable to improvement; b) future HSEMs should have a modular architecture so that each module can be edited and additional modules can be added; c) their underlying thermal models should be dynamic, to support accurate prediction of indoor temperatures and comfort, and the associated operation of heating systems; and d) databases should track their sources and development, and be continually updated so they can maintain their validity. Furthermore, a successful *dynamic* HSEM would ideally capitalise on available computing power, to support the rigorous and exhaustive testing of alternative decarbonisation hypotheses (concerning both design and energy using practices).

The housing stock Energy Hub (EnHub) platform has been developed in direct response to these observations. It is open and modular in structure, and enhances the virtues of existing HSEMs by dynamically simulating the performance of the building archetypes that comprise the UK housing stock; this latter requiring semantically attributed three-dimensional representations. EnHub also facilitates the straightforward testing of targeted housing stock decarbonisation scenarios, and is readily extensible to support the integration of models predicting household's investments to reduce their carbon intensity, and the associated impacts, in response to policies and strategies designed to stimulate these investments. By dynamically simulating the stock, it also facilitates the (future) study of how households apportion the co-benefits arising from these investments: reducing energy use and emissions on the one hand, and improving indoor thermal comfort and health on the other. Finally, EnHub improves on computational scalability using cloud and high performance computing technologies.

The processing of data to represent the housing stock is achieved using the statistical computing software `R` [10]. This is also used to construct dwelling archetypes, which are then simulated using the dynamic building simulation program `EnergyPlus` [11]. The platform creates geometrically simplified models constructed of contiguous cuboids, following the Domestic Ventilation Model (DOMVENT) [12] and Steadman's model [13]; assigning semantic attributes to these cuboids based on survey data (i.e. of envelope properties and household variables). Thus, the platform is able to derive more informative metrics than has hitherto been possible, including incidences of discomfort, the proportion of the stock that over- or under-heats, the heat gains per square metre of floor area, the disaggregation of energy demands, and the estimation of peak thermal and power demands.

The paper begins by describing, in Section 2, the basis of EnHub: its algorithms and data structures, and the workflow employed in its application. Then, in Section 3, a number of scenarios to decarbonise the domestic stock in the UK are tested and discussed in terms of their effectiveness. The paper closes by critically evaluating the utility of EnHub, and by identifying how its utility can be further enhanced to support the formulation, and the more rigorous testing, of alternative decarbonisation policies and strategies.

2. Methods: Statistical Analysis and Engineering Models

The structure of EnHub takes its inspiration from the CHM, which is at the core of the *Energy Consumption in the UK* study [4], and has been identified as the most flexible and powerful of prior HSEMs [9]. The principle data set underpinning both EnHub and CHM is derived from the English Housing Survey (EHS), which comprised 14,951 dwellings in its 2011 version, and is weighted to represent the 21 million houses in England. This data set is augmented by the Census and the Home Energy Efficiency Database.

The principle differences between EnHub and CHM, besides the more granular and transparent architecture of EnHub, are that:

i. EnHub utilises the dynamic simulation program `EnergyPlus` for energy performance predictions, while CHM uses the Building Research Establishment Domestic Energy Model (BREDEM), a simplified energy balance model.

ii. EnHub represents dwellings volumetrically, thus explicitly representing built form and adjacency (e.g. exposed or shared walls), and facilitating the more explicit modelling of thermal losses and solar gains, while CHM only scales the dwelling archetypes, limiting the analysis of envelope transfers.

iii. EnHub's archetypes represent the housing stock in a structured hierarchical way, which eases communication with its underlying data sets and facilitates convenient testing of modifications to their attributes, while CHM requires direct manipulation of models corresponding to individual EHS entries: testing modifications is far from convenient.

iv. EnHub's architecture is readily extensible to model households' responses to socio-economic drivers influencing investments and changes to behavioural practice that impact on net building energy demand.

v. EnHub employs a process of statistical data reduction to reduce the number of archetypes to simulate, while CHM evaluates every instance of the EHS data sets, with corresponding redundancy.

[Figure 1 about here.]

The steps involved in the application of EnHub are conceptually summarised in Figure 1 and are described in detail in the following sub-sections. Once the main data set is integrated into the platform, the open-source statistical computing platform `R` is used to mine these data and to reduce the sample size by determining the most relevant archetypes contained in the original data set. Then, this reduced data set is re-weighted to match the original totals. The next step uses the archetypes to create a set of semantically enriched volumetric models that are used by `EnergyPlus` to simulate dynamic energy flows.

It is worth noting that both `R` and `EnergyPlus` can be paired with a Graphical User Interface (GUI)—a feature that is under development—or run in Command Language Interface (CLI) mode. In this way, both can be controlled from within a shell[1]. The purpose of using a shell is that 1) the platform can be detached from the operating system and 2) a set of low-level scripts can systematically control the modelling process. This enables parallelisation, so that a High Performance Computing (HPC) facility can be called upon to accelerate the computation tasks, by around two orders of magnitude. In this way, a number of computer nodes can be simultaneously requested, each being allocated an instance of `EnergyPlus` and a corresponding EnergyPlus Input Data File (idf). Hence, the parallelisation of the simulation process depends on available HPC resources. The generated data, on any chosen hardware, can then be re-integrated into the process described in Figure 1. The results are then extrapolated to the subsets of the UK stock represented by these archetypes, and the results are analysed.

2.1. Step 1: Data Mining

To provide a level of confidence in the EHS data set, a data mining process is performed. Data mining involves the application of diverse methods to predict and/or classify typically complex databases [14, 15]. Predictive methods usually apply regressions, although a highly developed sub-type includes hierarchical models (e.g. trees, additive models, neural networks) [5]. Classification methods usually apply cluster and ordination analyses, and are mainly used to study large databases. Both predictive and classification methods identify the proximity[2] between variables, and can be adapted and

[1] A shell is a program which provides an interface between the user and the operating system. Such an interface may be called via a CLI or via scripting files.

[2] The term *proximity* is a general term used in statistics that covers the numerous methods that estimate distance, dissimilarity coefficients and metric inequalities.

complemented to obtain further information about their emerging relationships. For example, the strength of correlation between a range of input variables and an output may determine influential input parameters; some input variables may have a similar influence on the output, so they can be combined or classified to reduce repetition. In studies employing survey data, it is common to interpret a variable as an independent predictor (also known as model misspecification), whereas in reality this should be linked to functional associations. Looking at both the origin of the data and the units of observation helps to identify such associations.

For housing stock studies, the chosen units of observation may be directly derived from real archetypes (representative samples) or deployed as average archetypes (synthetic samples) [16, 17, 18, 19]. Their selection depends on both the purpose of the study and the developer's expertise. The formation of archetypes requires predictive methods to define associations, and to derive fit-for-purpose units of the stock. It is possible to use both techniques to incorporate multiple sources of information, and to enhance the robustness of the archetypes.

When considering a data set with inputs and outputs, the first step is to test its linearity (summarised in the upper part of Algorithm 1). This is useful for revealing similarity in the inputs, statistical independence of the output variables, and—for the case of multiple variables—clusters of data. To this end, two parameters are essential to define the level of linearity among variables: proximity and correlation coefficients. Depending on the structure of the data set (i.e. types of variables, number of cases, weights, survey design), both parameters can be employed to ensure that reduced versions of the dataset are appropriately structured and scaled. Previous studies of nationwide surveys have investigated the usability of the linear association [20, 21, 22], and there appears to be agreement in the value of employing robust methods to account for the influence of categories, and more importantly, to identify redundant variables in terms of their level of significance [5, 23]. Robust methods apply criteria to reduce the impact of multicollinearity, partial effects, and unclear distributions [24]; for instance, by employing median values to identify influential variables, by removing outliers, or by assuming partial distributions to avoid misleading tendencies.

Algorithm 1 Data mining (on top) and reduction (on bottom) processes. Notation: $\leftarrow$ means *assigned from*; $\rightarrow$ means *assigned to*; $\propto$ means *is proportional to*; $:=$ means *creates and assigns*

1. test for linearity
   $L \leftarrow glm(EHS, Out_{EnHubEnergySimulation})$

2. test for correlation
   $C \leftarrow fpca(EHS, Out_{EnHubEnergySimulation})$

3. search for redundancy applying backwards elimination
   $glm(L, C, Out_{EnHubEnergySimulation})_{backwards}$

\---

4. re-sampling process based on mining outcomes
   $EHS_{reduced} \leftarrow (EHS \propto LHS(EHS))$

5. kruskal-wallis test and comparison for consistency
   $KWt(EHS, EHS_{reduced})$

6. obtain population subtotals, i.e. the accumulated sum for each group (n)
   $T_P \rightarrow S_1 + S_2 + ... + S_n$

7. obtain sample subtotals
   $T_s \rightarrow s_1 + s_2 + ... + s_n$

8. compensate with scaling factor for each variable
   $c_n \rightarrow S_n / s_n$

9. adjust parameters accordingly (medians or modes)
   if $n : numeric$ then $s_i := \widetilde{S}_i$
   if $n : categorical$ then $s_i := \widehat{S}_i$

10. re-construct sample using the compensated factors
   $T_s{}' \rightarrow s_1 \cdot c_1 + s_2 \cdot c_2 + ... + s_n \cdot c_n$

In our case, a linearity test is performed via Focused Principal Component Analysis (FPCA) and Generalised Linear Models (GLMs), with a view to identifying potential overlaps among the variables. FPCA extends the classic approach of Principal Component Analysis (PCA) that applies scores to the variance of the data and, by extracting the proximity of those variables that are ranked most highly, identifies redundant variables. It is also useful in testing for correlations between variables, the presence of clusters, and the measurement of variability [25] and collinearity [26]. FPCA is used because it provides a virtual rank of the variables, thus identifying those that are most influential. Furthermore, because non-linear relations and mixed effects are expected in the EHS, FPCA is useful in ranking variables that include categorical variables; acknowledging that, due to the dimensions of the EHS, the outcomes are merely indicative and should be complemented with alternative methods.

To tackle the problem of multicollinearity, GLMs may also be applied to eliminate redundant variables. GLMs extend ordinary linear models by including a distribution of the expected response and by assuming that more than one variable is dependent [27]. In this way, as with FPCA, GLMs rank the variance in the variables in addition of the level of significance[3].

[Figure 2 about here.]

Figure 2 illustrates significant associations between input variables and three corresponding foci: gas demand, carbon emission, and internal average temperature. Here, the shorter the distance between a variable and the focus, the stronger their correlation [26]. The colour of the variables indicates whether their correlation with the focus is positive or negative, where red is positive and blue is negative. Also, variables in the same quadrant are

[3]The level of significance (or $p - value$) is a trigger at which the null hypothesis is rejected. Although not mandatory, the trigger at 0.05 is commonly employed.

positively correlated, whereas those in opposite quadrants are negatively correlated. For example, Figure 2c describes the degree of correlation of variables with the average internal air temperature *IntrnlT*. Here, both the total floor area (*tfa*) and wall type (*wllcnst*) are at a similar distance from the focus and are located in opposite quadrants. Therefore, they are negatively correlated to each other, and have a similar magnitude of correlation with their focus. However, *tfa* is blue and so is negatively correlated with the focus, whereas *wllcnst* is red and is positively correlated with the focus. These relationships can be explained by the characteristics of different housing archetypes, where older dwellings, constructed pre-1920, incorporate less energy conserving materials in their envelope, and are generally bigger [28]. Finally, the surrounding coloured radial line indicates the limit of acceptable associations, although the high number of variables considered in this case makes this wide.

The linearity test is complemented by performing a process of backwards elimination to remove redundant variables. This process consists of ranking the variance of the variables, and systematically removing those having a minor influence. The level of significance of the whole model becomes the criterion used to determine when the model sufficiently represents the stock (see 3 in Algorithm 1). Because the GLM splits categorical values in its classes to independently evaluate each of them, some uncommon subcategories may be eliminated in the process, but then be re-included in the outcome regarding a main variable. For instance, this is the case when considering the main heating system, where 95% are central gas and electric systems, and so the less common systems (e.g. oil, wood, and coal) are initially removed from the GLM outcome.

Three results of the linearity test are presented in Figure 2 to identify the most influential parameters for each outcome, based on evaluating the full EHS data set using the dynamic simulation program `EnergyPlus`. In this way, it is possible to obtain different ranks according to the chosen outcome variable. Likewise, Appendix A shows the resulting variables for three reference outcomes of the model; here, the analysis reveals that fabric components, particularly the geometric parameters, are essential for the model, so that a relatively explicit representation of them is appropriate. Table A.6 in the Appendix also highlights four variables that are shown as significant in Figure 2: Total Floor Area (TFA), number of floors, DHW and eHS (household size); each of these variables appear within the inner circles in the Figure, and so are considered either for the reduction process or in the construction of our volumetric archetypes.

[Table 1 about here.]

2.2. Step 2: Statistical Reduction

By generating volumetric archetypes, model complexity (in terms of data inputs) and computational cost increase; for each archetype needs to be explicitly simulated. Therefore, it is useful to explore strategies to reduce the number of them. An initial approach to achieve this, is to consider the most influential geometrical parameters of the relatively homogeneous housing stock, i.e. an attempt to encapsulate the different shapes that are present in it. For instance, a cuboid configuration—as illustrated in Figure 7—is different for mid-terrace and end-of-terrace houses, which is relevant when evaluating the effects of shared boundaries. It is also similar for end-of-terrace and semi-detached houses, but with different proportions and semantic attributes.

Variables $C_1 - C_4$ in Table 1 represent the 64 combinations of geometric archetypes that represent the range of UK housing archetypes. These geometric variables are complemented with semantic variables that represent heating systems, period of construction, tenure type, and region. As expected, some of these variables are correlated with each other [29, 30, 9]. For example, older constructions employing solid masonry, have significantly been larger than modern ones, have evolved from solid or oil fuels to gas, and have mainly involved private owners. Gradually, due to the rise of local authority housing schemes [31, 32], in addition to the simultaneous improvements in energy conservation measures and policies [33, 28], the modernisation of the housing stock improved façades to reduce heat losses, introduced electric heating systems, intensified the installation of efficient water heating technologies, and adapted to smaller households [34, 2]. The correlation of such variables constitutes thus a reliable indicator among households, dwellings and their energy performance. This can be seen, for example, in the Government's *Standard Assessment Procedure* [35], which is deployed to catalogue building properties in the UK, and tabulates most of these variables.

By including these semantic variables into the modelling of archetypes, the relevant parameters identified in both the FPCA and the GLM are preserved, so that the stock is effectively characterised. These variables, summarised in the bottom of Table 1, are also employed to identify redundancy in the data set, and hence to reduce its size by applying Latin Hypercube Sampling (LHS), a method that improves the randomisation process of a typical sample, by assigning a suitable distribution that describes the range of possible values, or by specifying parameters for each variable (or dimension) of the data set [36, 37]—this is summarised in the lower part of Algorithm 1. It is worth noting that these properties are considered in a reduced form, otherwise the combination of each category, as specified in the EHS, would increase in size. Therefore, LHS is applied, varying the sample size, the number of variables, and the iteration size. This sampling strategy is convenient because it can handle both numerical and categorical data, whilst ensuring that the statistical structure of the sampled synthetic stock is a good approximation of the original survey dataset (e.g. in terms of dwelling shape, size, envelope properties and technologies)—see Figure 5, of which more later.

The results of the LHS indicate that by using the variables $C_1 - C_8$ (in Table 1), the objective function stabilises with samples over 600 elements and converges at around 1000 (see Figures 3a-3c). Both reference variables and weighting values (used to compensate common properties) are used to optimise the objective matrix for each EHS sampling unit. This reduction process shows that when only a few variables are considered, the resulting data set is highly affected by the sample size, whereas the data set with more variables stabilises much faster. In our case, 1016 unique units are sufficient to represent the structure of the stock. These are defined by the combination of variables minus the non-existent configurations in the survey data, that are able to represent the main composition of the housing stock; a 15-fold reduction of the EHS data set, whilst adequately representing its structure. An alternative method of creating a reduced stock is to associate median parameter values to the 64 units formed from 7 of the 8 categorical variables ($C_1 - C_6$ and C_8) given in Table 1 (C_7 is excluded here because it is too broad), but this risks an inadequate description of the structure of the stock for a modest computational saving.

[Figure 3 about here.]

Figures 5a-5b suggest that the loss of information through statistical reduction does not have a significant impact on the representation of the structure of the stock. For instance, Figure 5a shows that the median and variance of GF area is well represented in the reduced data set; likewise the GF storey height, albeit with slightly reduced variance. The characterisation of categorical variables is similarly well represented, as seen in Figure 5b. Similar results are found for all the input variables forming the data set. In addition to this visual inspection, the resulting variability in the data set is evaluated using a Kruskal-Wallis H Test (KWt) across all the variables. This non-parametric test can be used to identify whether the variables in the reduced data set maintain the original information. The KWt compares the correlation of variables in the data set, so they can be numerical or categorical. It is also used when the examined groups are of unequal size, and so can be used to compare the EHS and reduced data sets of 14,951 and 1016 cases, respectively.

H is a measure of the relative proximity in the correlation of both groups. Large values of H indicate that the data sets are significantly independent. Because the KWt assumes a distribution, then H can be used to measure the level of significance defined by the distribution under the null hypothesis; this hypothesis establishes that the difference between the medians—of each parameter value—in both groups is null (or quasi-null). Therefore, a value below the level of significance suggests that the groups, in a given category, are different.

Table 2 shows that six of the input variables have a p-value < 0.05, showing that they are significant and so should be included in the reduced data set. However, none of them directly affects the composition of the volumetric cuboids. Because the EnHub platform uses the reduced data set, a significant variation is expected in the level of detail of non-primary components, such as secondary heating systems and boiler parameters. To determine whether EnHub's predictions of energy performance are sensitive to these parameters, a One-at-a-time (OAT) sensitivity analysis is applied, in which it is found that dwelling archetype significantly impacts energy performance; but this variability, expressed by H, is more sensitive to the contiguous cuboids whose geometry is configured and scaled using the reduced data set, i.e. modifying width, height, and depth, as well as adjusting the openings ratio, in each of the volumetric archetypes.

2.2.1. Sensitivity Analysis

Sensitivity Analysis (SA) techniques are primarily used to quantify input uncertainties [38, 39, 40, 41] and to measure the variation of model outputs relative to their inputs [42]. SAs are important for tracking errors and assumptions in model inputs and for determining input influence. When applied to housing stocks, they can reveal when the resolution of the data set introduces insufficiently granular representation of the parameters, and can also quantify the extent to which modelling algorithms are unable to effectively project the variability that is found in practice.

As a result, the implementation of SAs is useful to test the effectiveness of strategic changes to housing stocks, identifying potentially impactful parameters to study when evaluating decarbonisation scenarios, and to improve the development of stock models [43, 44, 45].

[Figure 4 about here.]

Global and local SA are two common methods employed in housing stock evaluations. The former considers a set of samples, with simultaneous changes in their variables, so that interactions between them are also represented, to evaluate the effect on the outcome or the response; the latter performs derivative changes at a single point, so that the response is affected by single and sequential changes only [42]. A technique commonly employed for global and local sensitivity analysis is the regression method. Tian [46] claims that it is the most widely used method for SA in building energy analysis, particularly when the inputs are independent. However, this requirement can be problematic when two or more model inputs are correlated. Therefore, it is common to combine both global and local SAs to determine the level of significance of the inputs.

A common local SAs method is the One-at-a-time (OAT) approach, where each input parameter is varied at a defined interval while the others are held constant [47, 43]. The OAT approach has been extended to screening methods [48, 49], which compare the uncertainties in the inputs to those in the outputs. The Morris method (also called the *elementary effects* method) is the technique most often used in this domain. Here, each input variable takes a discrete number of values that are chosen within limits defined by their statistical properties. The method provides a measure of the mean value, μ, which estimates the overall effect of the input on the output, and a measure of the standard deviation, σ, that estimates second- and higher-order effects of the input variable [50, 48]. This process is outlined in Figure 8 where the set of variations is assigned in the input data set, enabling the possibility to compare against a baseline.

The enriched cuboids, global, and OAT analyses are performed to identify the most influential parameters in the data set, many of which have been identified in previous studies [40, 43, 51, 45]; albeit using different simulation workflows. In EnHub, they are directly included in the simulation platform. For example, Figure 4 presents results from the application of SAs by controlling changes to the properties of elements contained in the catalogue of properties used to generate idfs, and by modifying the reduced data set (or indeed by modifying properties related to the external environment, such as weather files). Figure 4a shows that changes in loft (sometimes known as attic) insulation and heating system efficiency significantly affect the energy performance; more so than increasing the wall thickness, as indicated by their gradients. Figures 4b and 4c compare two output variables by applying a screening method, which makes it possible to combine variables of different class. Each input variable defines a dimension for the analysis in which a trajectory of perturbation is defined. The trajectory is chosen by advancing within the same dimension or to one other. In other words, any input variable can be selected and changed from an initial stage, but this is limited to the next possible value, either forwards or backwards in its own dimension. The input change constitutes the next initial stage, and so the trajectory is developed. This allows comparisons between the discrete values of each input variable with those of the output. There is widespread agreement that this method is versatile and useful [48, 49], but the computing time increases significantly and so its application to large data sets is limited. EnHub accomplishes this analysis by selecting variables based on expertise, and their use in national programmes and policies.

The results of the SA module are consistent with those presented in previous studies [43, 51], where a rank of parameters establishes that the differential of temperature between the outside and the inside of a dwelling is the most sensitive component. This is followed by the efficiency of heating systems (space and DHW), and infiltration rates. Thermal transmittance (commonly known as U-values) is also prominent, especially for walls, as is the number of occupants, although this value is limited in impact since it is decoupled here from the associated activities (e.g. use of appliances).

Similarly, yet limited in previous studies due to poor representation of external parameters, solar energy technologies and socio-demographic variables are identified by the OAT analysis as being significant, and so they are directly included into the model archetypes. Moreover, the original units of the data set are directly derived from the EHS, which in turn uses the Postcode Address File (PAF) and the Lower-Super Output Area (SOA) to define their weights[4], which are used to represent the whole English housing stock. These units are selected randomly and merge information from nearby PAFs, and the extrapolation is delimited by census subtotals. The inherent bias in this method, because of the generalisation of categorical variables, affects solar-related parameters, and the definition of representative households. Thus, improving the quality of information describing these variables would improve their significance.

[Table 2 about here.]

[Figure 5 about here.]

Once the data set has been reduced, the units are re-weighted, and their tenure and regional information are re-calibrated following a method employed by many other nationwide surveys [52] (see 6-10 in Algorithm 1).

[4] A weight, or weighting variable, assigns a value to a given case in a data set to compensate how much they should be represented in the analysis [52]. In survey data, weights are scaled to sum to the population total.

This method compensates for the assumptions made in the sampling process which, as in the EHS, introduces a loss and bias of information [51]. The impact of this reduction affects the performance of the simulation algorithms, because their inputs may not exist or may be over-generalised. Hence, there is virtue in developing a modular approach to construct archetypes, independently.

[Figure 6 about here.]

2.3. Step 3: Archetype Construction

The set of geometrically simplified models are constructed in a modular fashion, following an Object-Oriented Modelling (OOM) approach; essentially a collection of interacting objects, where each object is semantically defined with variables, classes and dependencies [53]. An OOM approach provides the ability to detach specific sections of the modelling process, so that they can be refined independently. OOM can be represented graphically using the Unified Modelling Language (UML) architecture to provide a clear and unambiguous structure for the model attribution workflow, and to enhance the integrity of the predictions of the algorithms used to describe the energy flow pathways [54]. That is, the architecture shows the logic in which the components are considered, as well as the distribution and interrelationship of inputs and outputs used throughout the platform.

EnHub utilises a UML class diagram type (CD), focussed on the characteristics of the modules and their relationships. Similar implementations have been used elsewhere to improve the development of highly detailed libraries of classes to support more robust management of databases and modelling algorithms [17, 55]. The EnHub approach is based on both the `idf` documentation [56, 11] and the BREDEM structure [57], applied using the CHM protocol [58]. The UML architecture is presented in Figure 6 and its main components (that form the archetype-specific idfs) are summarised in Table 3.

To link the EHS with the archetype-specific `idf` modules, a base case `idf` is developed where variable flags[5] are used to represent instances of the UML classes. In this way, each EHS sample unit can be associated with its corresponding property fields and used to structure, scale and attribute the volumetric archetype appropriately. These archetypes are explicitly represented as sets of contiguous cuboids (see Figure 7a) that account for the number of floors and the presence of attics, basements, and attachments such as extensions and conservatories; Table 1 summarises the variables used to built such a base case. Specifically, an array of cuboids is built by processing the archetype information; cuboids attached to the main zone, either on top or at the sides, can then be scaled or removed as required. This helps to maintain the connection of shared surfaces. In this way, it is possible to represent the key residential types of the stock, such as detached houses, semi-detached, end of terrace, mid terraced, flats (also known as *apartments*) and bungalows (see Figure 7b). These forms, complemented with their semantic attributes, comprehensively and robustly describe the derived stock of archetypes, but are now based on an explicit volumetric representation (see Figure 7c).

[Figure 7 about here.]

The volumetric representation used by an `idf` includes the assignment of internal zones. EnHub follows the BREDEM representation, using two

[5] In programming, a flag is a predefined string or sequence that is used as a reference when a script is read or executed.

zones. These are initially defined in each cuboid archetype using a main zone, and a distinct secondary zone that represents the rest of the building. For example, in Figure 7a, the ground floor (GF) is initially assigned to the former, whereas the basement, the first floor (1F) and a room in the roof (in some cases, interchangeably referred to as attic), are assigned to the latter. The attached constructions are employed to incorporate differences between detached dwellings and those forming terraced rows. Moreover, relationships between openings (i.e. windows, doors, accesses) and total floor areas are defined for each zone. In Figure 7b, the semi-detached house, as shown in the middle, increases the number of surfaces at which `EnergyPlus` simulates solar radiation. Hence, the relevance of a volumetric representation of archetypes, and the need to appropriately characterise openings.

Employing a base case `idf` to construct the volumetric archetypes helps to respect `EnergyPlus`'s precise coordinate structure. Furthermore, whilst `idf` property fields can be combined in any order, it is convenient to follow the sequential structure of the `EnergyPlus` format [56, 11], because this helps with the readability in the representation of our archetypes, and the convenience with which their attributes may later be manipulated. The steps involved in adapting the base case `idf` are described in the pseudo code shown in Algorithm 2.

Algorithm 2 Construction, simulation and analysis of archetypes;

Notation: $\langle,\rangle$ means *is a subset of*

```
1. construct dwelling archetypes for simulation;
   s represents the size of the data set employed
  loop:
    for i:=1 to s do
      s_{A-general} ← f(SAP, ⟨EHS⟩)
      s_{B-location} ← f(SAP, ONS, ⟨EHS⟩)
      s_{C-schedules} ← f(ONS, ⟨EHS⟩)
      s_{D-envelope} ← f(ONS, ⟨EHS⟩)
      s_{E-geometry} ← f(ONS, ⟨EHS⟩)
      s_{F-internal} ← f(SAP, ONS, ⟨EHS⟩)
      s_{G-heating} ← f(SAP, ONS, ⟨EHS⟩)
```

```
      s_{H-water} ← f(SAP, ONS, ⟨EHS⟩)
      s_{I-outputs} ← f(EPlib, ⟨EHS⟩)
    done

2. simulate dwelling archetypes
  loop:
    for i:=1 to s do
      simulation ← f(i, weather_i, schedules_i)
    done

3. analyse simulated dwelling archetypes
  loop:
    for i:=1 to s do
      summary ← f(i, EHS, Out_i)
      derived ← f(i, EHS, Out_i)
```

A catalogue of appliances and heating systems is included in the base case `idf` to describe internal gains. Their specifications (i.e. power, efficiency and heat gains) and fuel sources are defined using the EHS and the *Energy Consumption in the UK* tables [59]. Similarly, household water services are included in the `idfs`, although they are reduced to a total annual demand. Occupants and their associated (estimated) metabolic gains are also included, and can be directly linked to specific intensity values, such as heat gain per person and floor area per capita.

The dynamic actions of devices and occupants are emulated using schedules. These include time-related events, such as occupant presence, or the use of appliances, window openings, heating systems, and water fixtures. The schedule profiles are prepared for each model by applying a probabilistic function that is dependent on the properties of each archetype, such as the household composition or dwelling type (see the pseudo code shown in Algorithm 3). This means that even though profiles follow similar patterns, each element is unique. This approach enables the behaviours of a specific group of archetypes to be analysed. The resolution of a schedule is independent of the model, but they are conveniently paired with the computational time-step at an hourly resolution [56]. Currently, the generated schedules for each day type are built for an entire year, but they can be seasonal.

Algorithm 3 Probabilistic profile generation

```
1. extract profile from national average daily profiles
  M_n ← extract(data: average usage); n elements

2. extract representative household parameters
  Hh_p ← extract(data: household); p population

3. assign a probabilistic function to each profile
   based on average daily usage
  d_n ← P(M_n)

4. create random profile based on p-function for
   a given period
  loop:
    for i:=1 to 1016 do
      for j:=1 to n do
        for k:=1 to 365 do
          randomise(Hh_i, d_jk)
```

In the same vein, EnHub uses a library of standard `EnergyPlus` Weather (EPW) files to describe local environmental conditions for the defined period. The appropriate weather files, used by `EnergyPlus`, are associated according to the regions that the model archetype has explicitly been allocated to. The regions are aligned to the resolution provided by the EHS: NUTS-1[6] level, which encloses subdivisions of the four UK countries, and is regulated by the European Union. These EPW files are synthesised Test Reference Years (TRYs), where each month represents an average over a period of around 20 years. Each month may come from a different year, and so they are combined using a cubic spline method [60]. This method interpolates the data when gaps and inconsistencies are present.

[Table 3 about here.]

[6] `www.ons.gov.uk/methodology/geography/`

2.4. Step 4: Dynamic Simulation

As noted earlier, the platform architecture has been designed to provide for both flexibility and scalability. For example, it can simulate a single archetype and systematically modify the parameters of *A1* and *C2* (see Table 3) to increase temporal resolution; it can employ a OAT sensitivity analysis module to analyse key independent variables (fabric properties, use intensity, devices efficiency) for one or more archetypes; or it can simulate the entire UK housing stock to evaluate different scenarios (see Sections 1 and 3). Analyses can also be geographically constrained, to simulate the housing stock of a city. This scalability results from the ability to detach processes, as well as from the aforementioned modularity.

The whole stock is simulated for an entire calendar year, plus an additional preheating period, at an hourly resolution. This resolution is reasonable, given that we are not explicitly modelling ventilation, nor accounting for the associated stochastic behaviours of occupants (although this will be amended in the future); we simply utilise a fixed infiltration rate, and ventilation heat gain schedules. Each simulation takes around 30 seconds to complete using a single core high-end computer[7]. However, multiple simulations of the 1016 archetypes are required and so the workflow is coupled with a High Performance Computing (HPC) cluster[8], to reduce the simulation time by over 98%. Figure 1 presents the conceptual structure of EnHub and shows the communication paths between the components of the platform and the tools required to support it.

[7]Software specifications: `EnergyPlus` version 8.4, `R` version 3.2; Hardware specifications: Processor Intel Core i5 2.9 GHz CPU, 16 Gb RAM

[8]166 compute nodes (Dell C6220), with 2x 8-core processors (Intel Sandybridge E5-2670 2.6GHz) and 32GB RAM

2.5. Step 5: Iterative Analysis

The derived energy performance indicators are processed in `R` so that they can be conveniently stored and compared. Thermal comfort indicators are computed in `EnergyPlus` based on the ASHRAE-55 standard [56], but in the future they could be readily complemented with other indicators of instantaneous comfort, or of over- or under-heating risk, by post-processing the `EnergyPlus` results within EnHub.

Each batch of simulations of the 1016 archetypes is stored in a database, in which multiple (input, output and auxiliary) data sets are linked by common identifiers (see Figure 8); enriching analysis options. SA is employed both for reference modelling (choice of archetypes and parameters to sample from) and scenario modelling (choice of parameters to act on). Figure 8 shows that there are two stages to the simulation process: the first stage (A) focuses on data resolution, whereas the second (B) focuses on modelling parsimony and the study of different scenarios.

[Figure 8 about here.]

3. Application to Evaluate De-carbonisation Scenarios

Here we deploy the EnHub platform to evaluate a baseline and a series of measures that are designed to reduce UK housing stock carbon intensity.

3.1. Definition of Scenarios

The workflow summarised in Figure 1 is used to test two CO_2 reduction scenarios that are adapted from national policies [61, 62, 4, 63]. The first emulates a *perfect uptake* scenario, where the fabric properties and appliances of dwellings are upgraded to be as efficient as is plausibly possible, regardless of their current values, associated costs, or the level of disruption in adopting them. The second is a *conditional uptake* scenario (see the pseudo code shown in Algorithm 4), where the upgrade is limited to a fraction of the stock represented by a particular archetype, and considers real-life constraints, such as costs, disruption, and the initial values of the parameters of interest; so diminishing the likelihood of uptake of measures that entail marginal gains or heavy disruption.

Algorithm 4 Implementation of conditioned uptake scenarios

`1. obtain technical potential of implementation`
$technical_i \leftarrow f(savings, trigger, uncertainty)$

`2. obtain income / fuel expenditure ratio`
$income_i \leftarrow f(income, trigger, uncertainty)$

`3. obtain typology feasibility parameter`
$contraints_i \leftarrow f(current, trigger, uncertainty)$

`4. obtain return of investment ratio`
$roi_i \leftarrow f(cost, trigger, uncertainty)$

`5. include random uptake process`
$random_i \leftarrow f(individual, trigger, uncertainty)$

`6. evaluate with thier corresponding uptake triggers`
i`:households`
$U_i \leftarrow f(technical_i, income_i, contraints_i, roi_i, random_i)$

[Table 4 about here.]

The selected scenarios and associated measures are summarised in Table 4, which describes the current status of the dwellings, and the variations that are then parsed to the model, with each scenario considered in isolation. For each scenario, an additional case is tested to understand the implications of different occupancy profiles and patterns of usage; these are based on average profiles defined in the *Energy Consumption in the UK* study [4]. Here, multiple profiles are defined based on their existing probabilistic distributions [4], so that the values relate to each archetype, and to each element associated with an activity (as previously shown in Figure 3).

3.2. Results

[Table 5 about here.]

[Figure 9 about here.]

Table 5 and Figure 9 summarise the outcomes from the perfect and conditional uptake scenarios, whose measures are given in Table 4. They show that when perfect uptake is assumed, A02 and A01 have the greatest impacts, arising from improved fabric performance. Other impactful scenarios are A05, A08, and A06; these latter two being partially counteracted by increased heating loads, affecting indoor temperature. When household constraints are considered in respect of scenario 02 (A01 to B02), the potential emissions reduction is reduced by around 40%. The constraints considered in measures B01-B07 are related to the disposable income of a household or to the increased costs associated with more technically challenging cases, such as housing in conservation areas, solid masonry walls in old buildings, or the insulation of loft spaces that are used for storage purposes with limited accessibility.

[Figure 10 about here.]

The combined scenario (A01+A02) has a relatively high impact on energy demand reduction (319 TWh, 25% less than the baseline). Table 5 shows that their combination is roughly equivalent to all the scenarios combined (All). Likewise, Table 5 outlines the associated potential to reduce energy demand by combining A01 with A03. However, it must be noted that some of the benefits of the individual measures overlap, so that the outcome is not the simple arithmetic sum of each respective saving, and so combining multiple measures does not guarantee a substantial improvement. This can be seen, for example, in Figures 9 and 10 that reveal the extent to which energy demand is reduced by combining all the measures.

[Figure 11 about here.]

Improvements to the fabric may be more likely to be viable for old and large dwellings because of the relatively poor performance of their fabric elements, and their relatively large surface areas. However, sometimes the spaces within these type of buildings have been repurposed and cannot be modified. For example, it is difficult to install energy conservation measures in a loft that has been designated for storage. Figure 11 outlines the proportion of dwelling typologies (by epoch and type) in terms of accumulated floor area in the housing stock. This helps to identify the archetypes that have particular decarbonisation potential, or lack of. For example, pre-1919 terraced dwellings accumulate the greatest floor area in the stock, but their uptake potential is limited by the aforementioned constraints.

[Figure 12 about here.]

Figures 12a and 12b outline differences in the energy intensity of dwellings by archetype. The area formed by the axis variables represents the total energy demand (428 TWh for the baseline case). This helps to better understand the effectiveness of the different measures, particularly when there are real-world constraints that affect their implementation. For instance, Figure 12a shows that, even though the accumulated energy demand reduction in older dwellings is significant, potential carbon reductions appear to be seriously constrained. Figure 12b shows that apartments are particularly sensitive to our modelled perfect uptake measures; albeit constrained by the relatively small cumulative floor area.

[Figure 13 about here.]

Figure 13 suggests that renovation measures may impact on occupants' thermal satisfaction, as the mean indoor temperature is elevated (relative to the base conditions) by around $+0.8\,\mathrm{K}$. This is due to indoor air temperatures that exceed the heating system set-point when it is on (due to excess solar gains), or are higher when the heating system is off (due to better conservation of energy). These improvements are expectedly more noticeable in the cold season, reducing energy bills among the population. Due to a lack of supporting data, any co-benefits to carbon emissions and comfort levels that arise from renovation measures are ignored here (meaning that potential energy savings may not be fully realised due to improved comfort from higher indoor temperatures).

Finally, while scenarios A06-07 and B06-07 show that substitutions of lights and appliances are low-cost upgrades with a quick return on investment, they are not very effective. Increasing the efficiency of appliances is clearly a quick and easy way to reduce electrical energy demand, but this can be counteracted by increased heating demands.

The study of scenarios is useful because it enables the quick and convenient exploration of the limits of potential impacts arising from related policies or strategies. The testing of scenarios that consider the conditional uptake of measures gives an understanding of potential to uptake constraints; but they do not shed light on probable outcomes. This would require a thorough understanding of the underlying drivers affecting the uptake of energy-related improvements by households, and the associated trade-offs between energy demand, costs, comfort, and convenience; likewise, changes to energy-using practices. EnHub will be continually improved to facilitate more granular analyses, and to integrate a platform simulating occupants' energy-using behaviours and associated adaptive comfort, as well as a social simulation platform to predict changes in households' investment and energy-using behaviours.

This combined platform has the potential to considerably improve upon the ability to reliably diagnose where the greatest potential for socio-technical innovation lies on the one hand—with a view to decarbonise the housing stock; and to design, test and implement policies and strategies to realise this potential on the other.

4. Conclusion

This paper presents a new open-source modular platform for the dynamic simulation of the UK housing stock—EnHub, with a view to efficiently evaluating the effectiveness of national decarbonisation strategies. At present, this is restricted to the analysis of the possible implications from perfect or assumed uptake of these strategies, whether physical (building envelope, systems and appliances), or behavioural.

Housing stock decarbonisation is a long-term endeavour that requires constant updates to the evidence base and recalibration of models. We believe that EnHub can facilitate better inter- and trans-disciplinary collaboration, by providing a standardised hub to combine data models from different perspectives. We also believe that an open-source dynamic simulation platform, such as EnHub, offers greater utility and longevity, compared to its closed and steady-state counterparts. This is reinforced by making each component, of the semantically attributed 3D representations of the archetypes comprising the UK housing stock, editable and verifiable.

To facilitate widespread accessibility, EnHub may be downloaded, upon signing a by-attribution license agreement, from the open source version control repository GitHub. This also helps to facilitate and manage contributions from other potential developers of EnHub—further enriching its future functionality. Furthermore, work is currently underway to improve the usability of EnHub, through a graphical user interface. In principle, this may widen the user base from those interested in national decarbonisation policy, to include those that are interested in regional and city scale decarbonisation policy and planning interventions—by including geographical constraints in its application.

Finally, to address the above restriction of assumed uptake of decarbonisation strategies, we plan to facilitate the combined modelling of household investment behaviours, and their responses to changes to the underlying drivers that influence them (e.g. regulatory, financial, educational, social, technical), and energy-using behavioural practices; this latter through integration with the multi-agent stochastic simulation platform No-MASS [64].

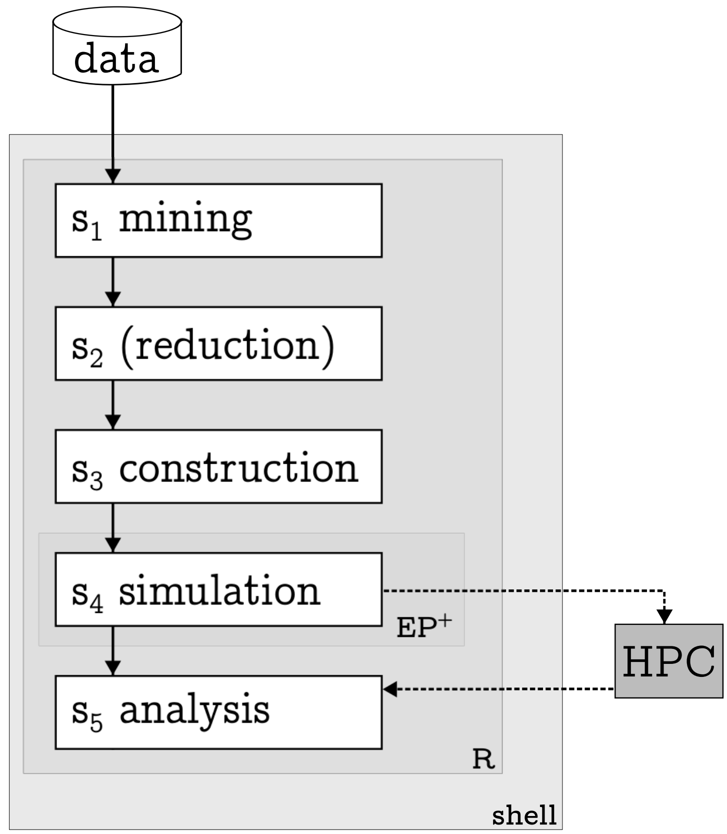
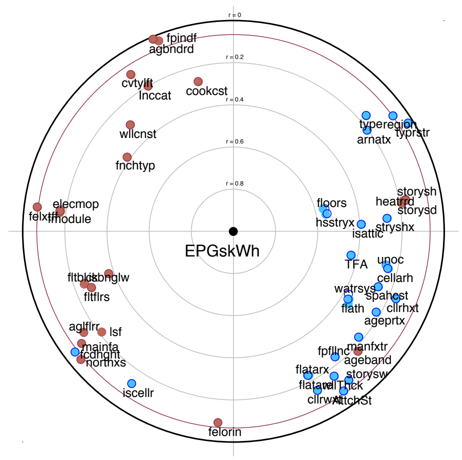
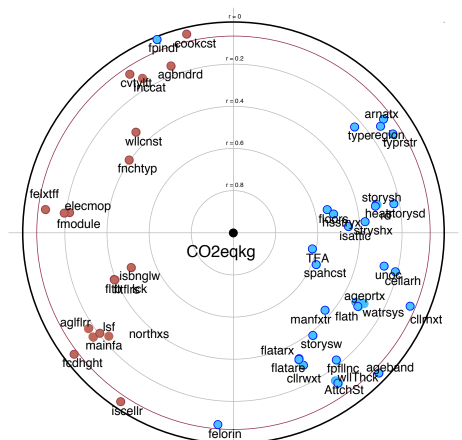


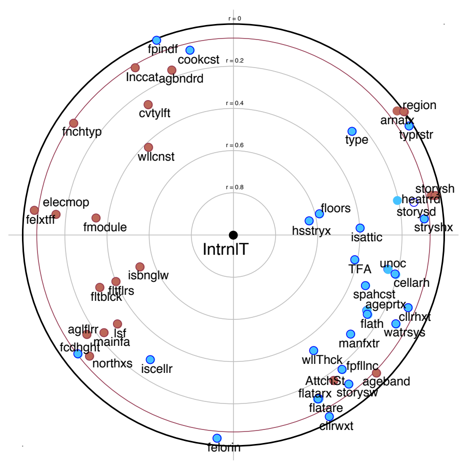
Figure 1: Steps involved in the application of EnHub. The statistical platform `R` is employed to perform all the steps. The computationally-intensive tasks, such as the simulation step, can be performed using a different hardware where `EnergyPlus` is installed.



(a) Gas Demand

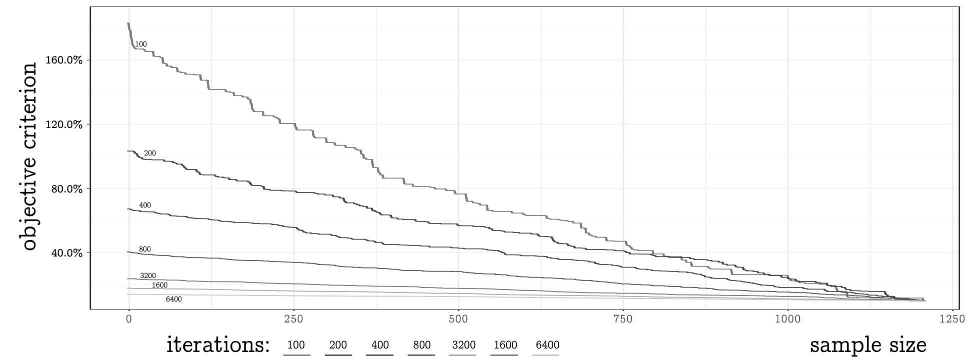


(b) Carbon Emissions

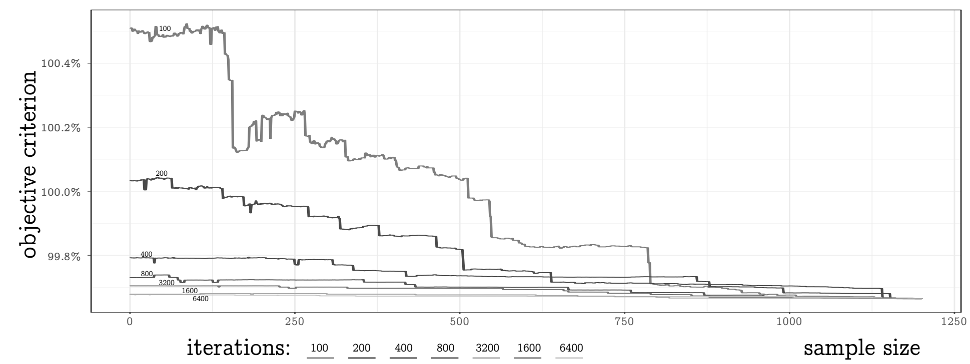


(c) Internal Average Temperature

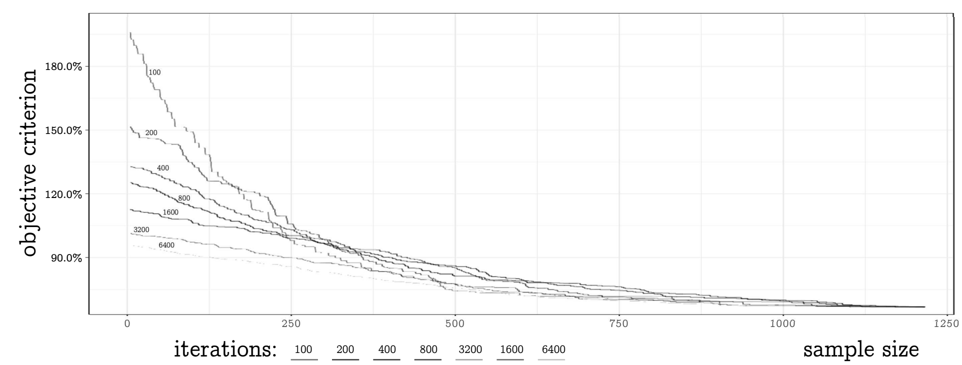
Figure 2: Focused Principal Component Analysis (FPCA) visualisation. The estimated energy demand in the models is the sum of fuel energy demands, in which (a) considers the main fuel used in the stock. To identify other variables affecting the overall energy demand, (b) considers the associated carbon emissions, which is directly dependent of this. In (c), the internal temperature captures the response of the volumetric archetypes to the influence of variables.



(a) 3 pivotal variables - from Table 1

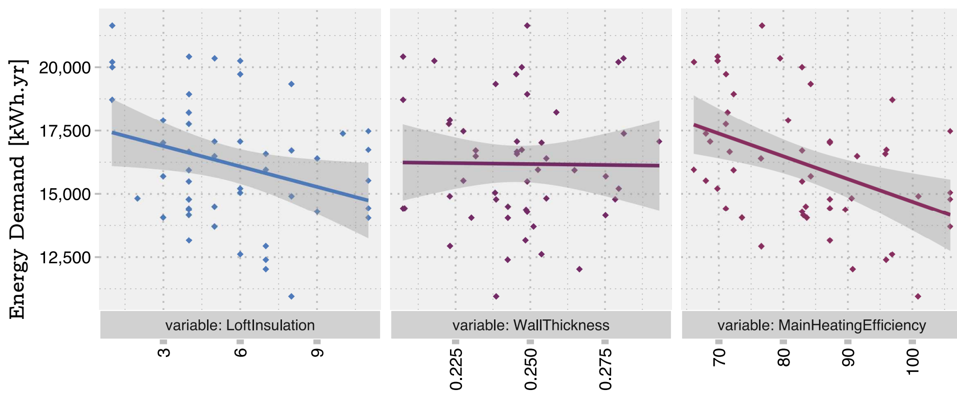


(b) 7 pivotal variables - from Table 1

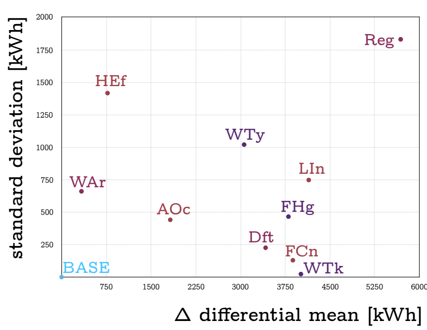


(c) 40 pivotal variables - from EHS

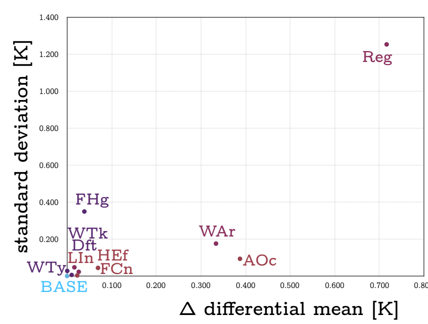
Figure 3: Conditioned LHS iterations, with different pivotal variables. The y-axis indicates the objective criterion which corresponds to the level of perturbation as compared to the initial random dataset. The x-axis represents the resulting sample size. And the lines represent trends according to the required number of iterations to accept a sample case. Here, (a) presents high perturbation and the convergence is diffuse, which demands a higher number of iterations; (b) is restricted to the archetypes, so its perturbation is minor, and stabilises below 1000; and (c) initiates with high variability, but stabilises quickly.



(a) Global SA - insulation, thickness, efficiency. The Confidence Interval (CI) is determined by applying *locally weighted scatterplot smoothing*.

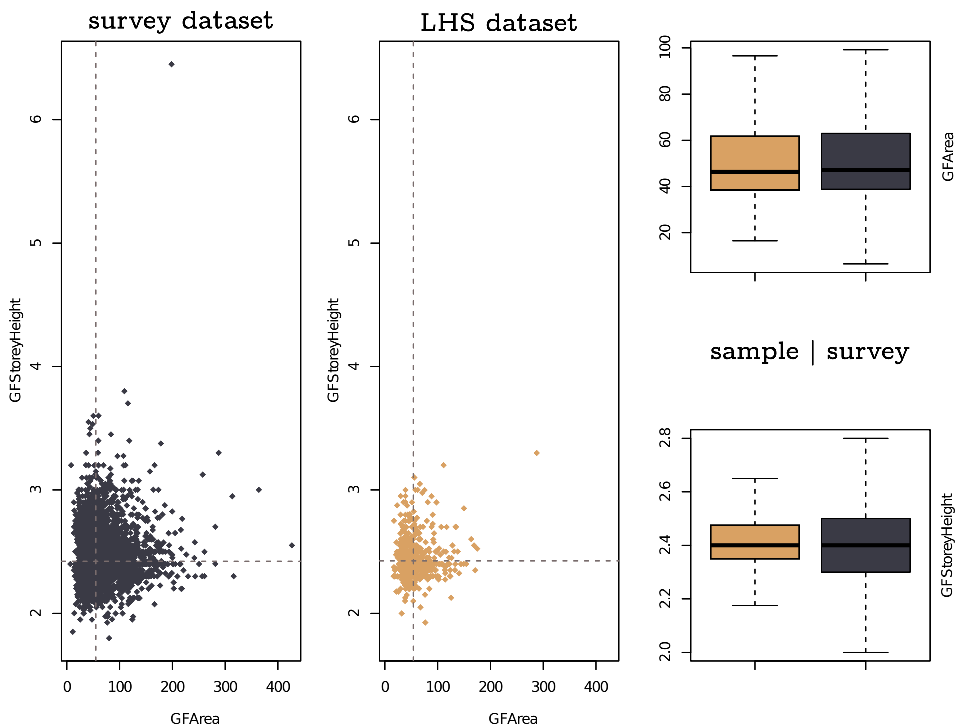


(b) Screening Method - energy demand reference

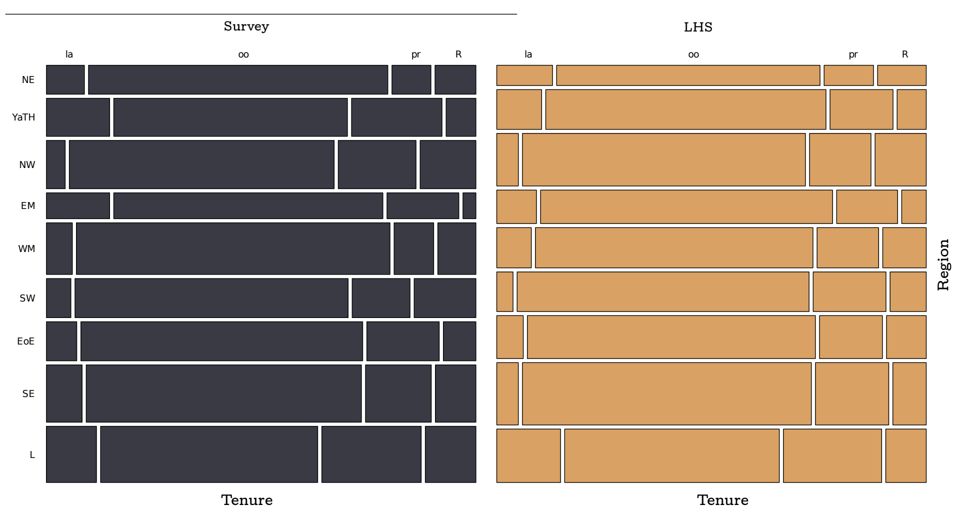


(c) Screening Method - indoor temperature reference

Figure 4: Examples of the SA module: screening method
Notation: AOc: AdultOccupants; Reg: *Region*; LIn: *Loft Insulation*; FCn: *GF Construction*; FHg: *GF Storey Height*; WAr: *Windows Area 1F*; HEf: *Main Heating Efficiency*; WTy: *Glazing Technology*; Dft: *Draught Infiltration*; WTk: *Wall Thickness.*



(a) GF dimensions



(b) Region and tenure variables

Figure 5: Comparison between the EHS and the reduced data set

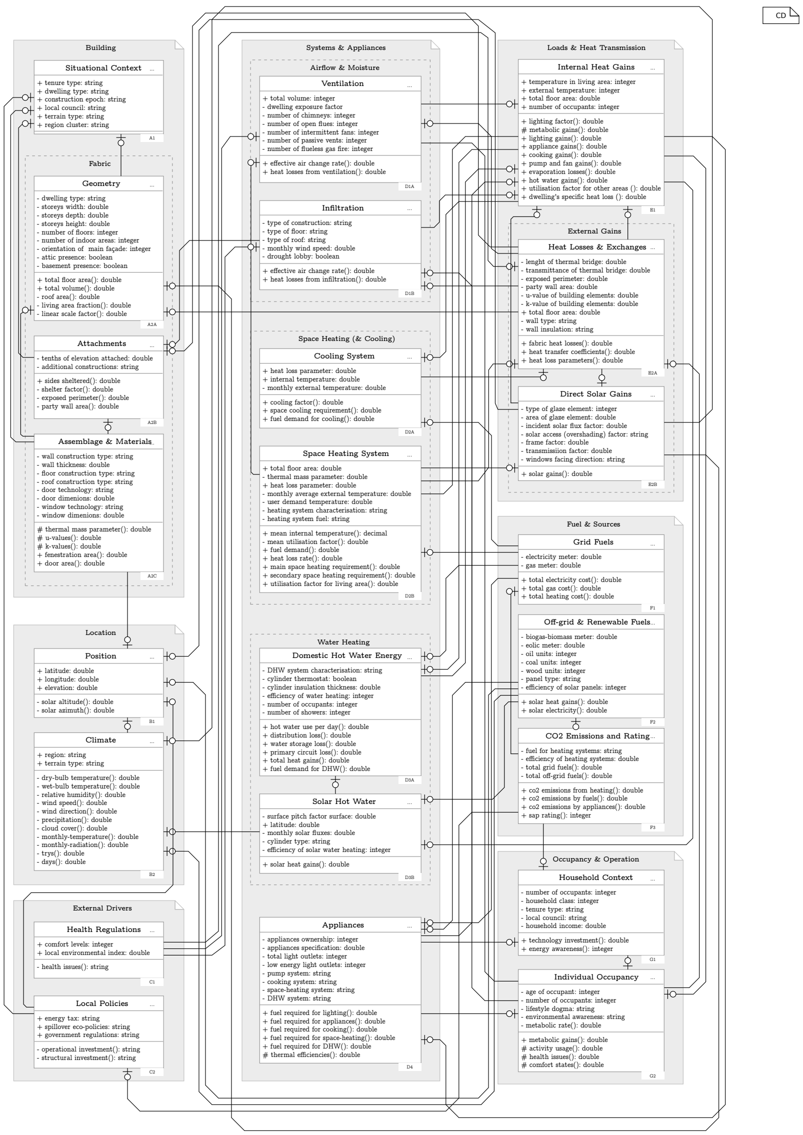
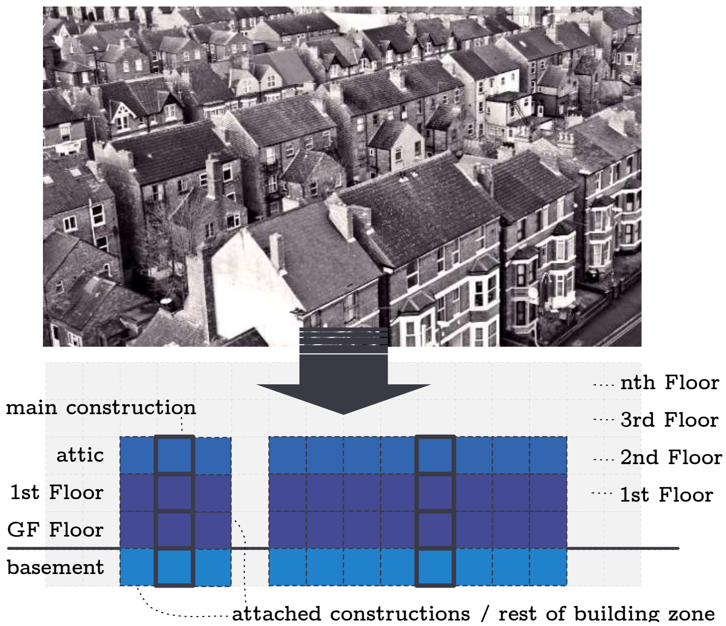
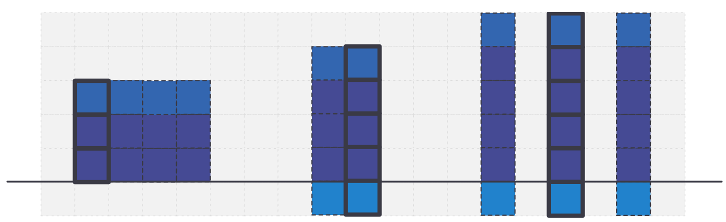


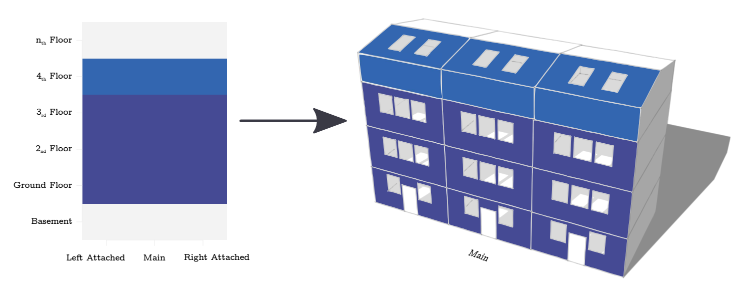
Figure 6: Development of UML architecture aligned with the BREDEM structure (adapted from [55])



(a) Cuboid description: abstraction of a two floors mid-terraced house with attic and basement. Each square represents a zone for the `EnergyPlus` format. – "Victorian Houses" (Author: N. Ramasamy; Flickr link; accessed 15 March 2013; Licensed under CC-BY2.0.; Adapted)



(b) Examples of Cuboid Configurations: left, end-of-terrace house; middle, semi-detached house with attic and basement; right, apartments



(c) Example of Volumetric Representation: terraced houses with attic room

Figure 7: Cuboid abstraction

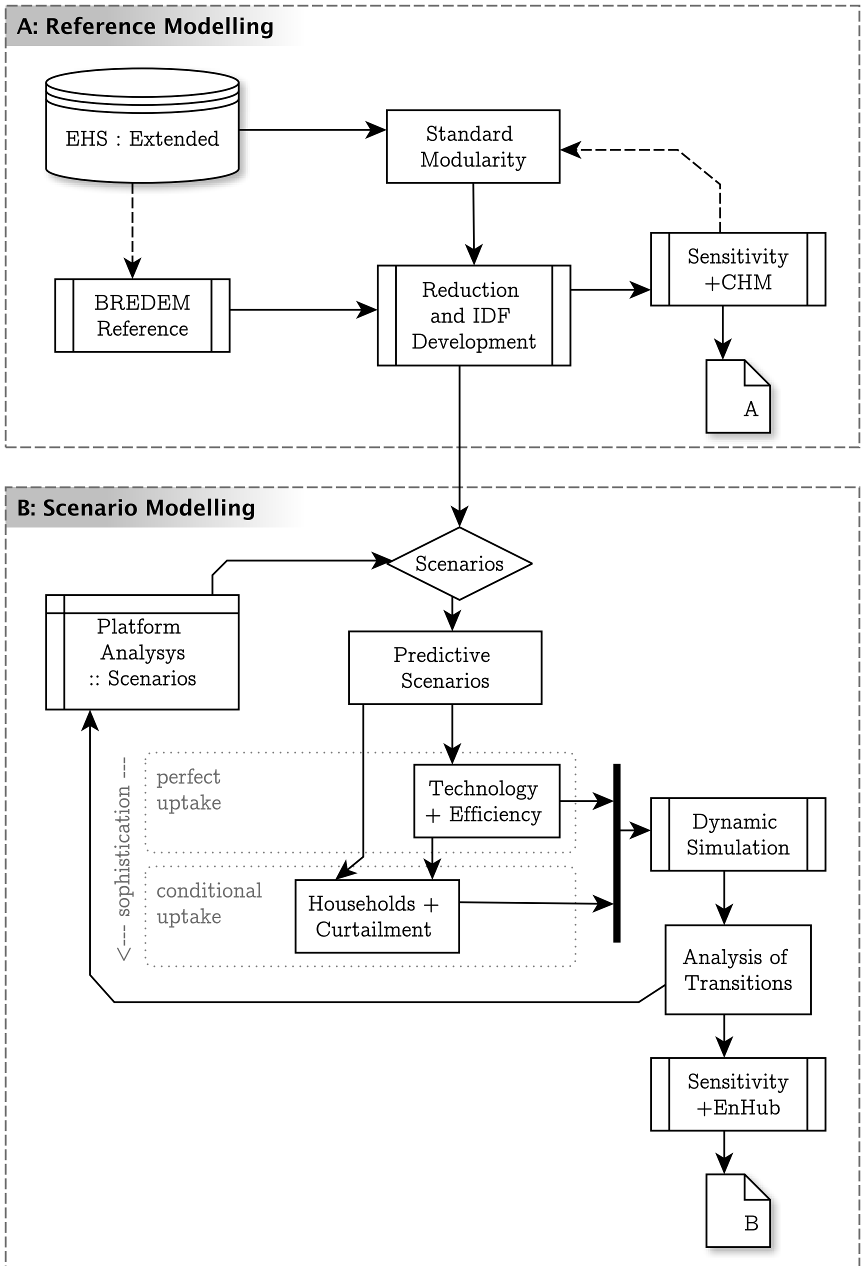


Figure 8: Simulation and analysis flowchart. The (A) *reference modelling* stage is employed to study energy performance of a dwelling archetype, either for further parametrisation, or for measuring specific variables. The (B) *scenario modelling* is employed to evaluate changes to the input variables, to test their potential effectiveness.

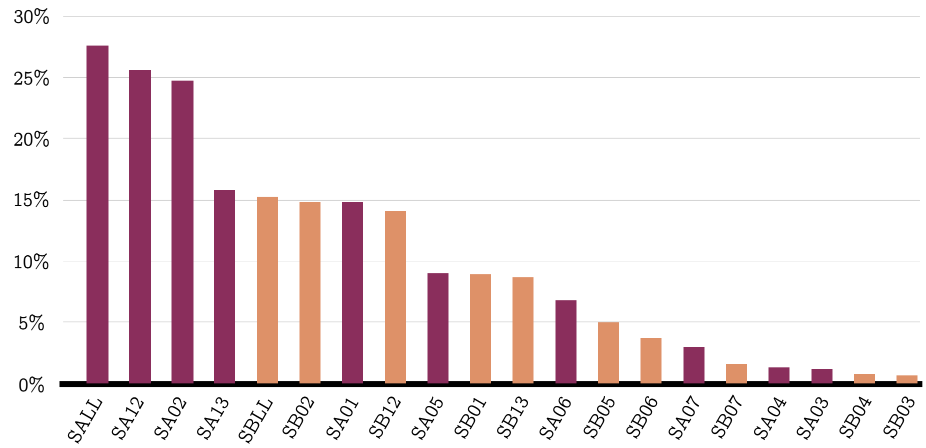


Figure 9: Summarised energy demand reduction, by scenario (Sx) and applied measure (#); **1**: wall insulation, **2**: roof insulation, **3**: glazing, **4**: boiler insulation, **5**: draught proofing, **6**: lighting, **7**: appliances, **8**: occupancy, **ll**: all combined

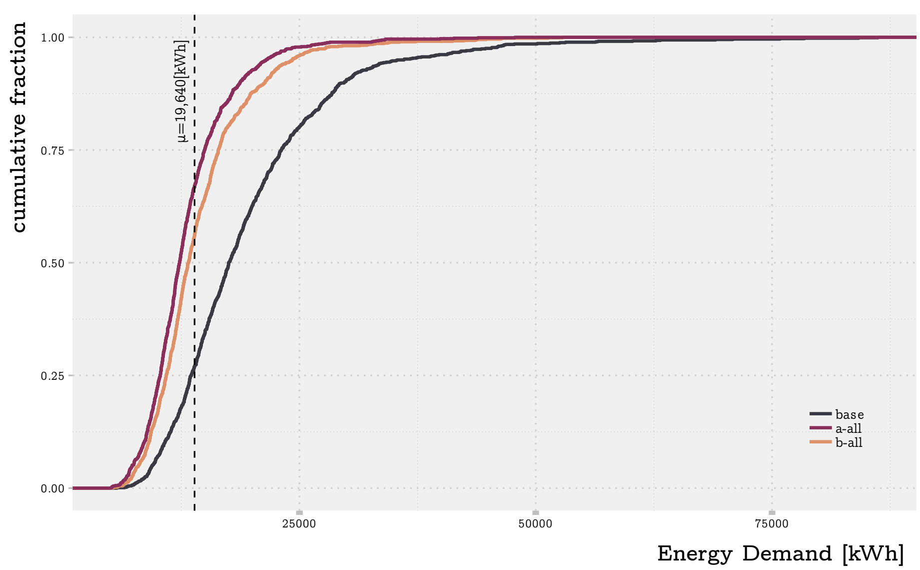


Figure 10: Differential Energy Demand

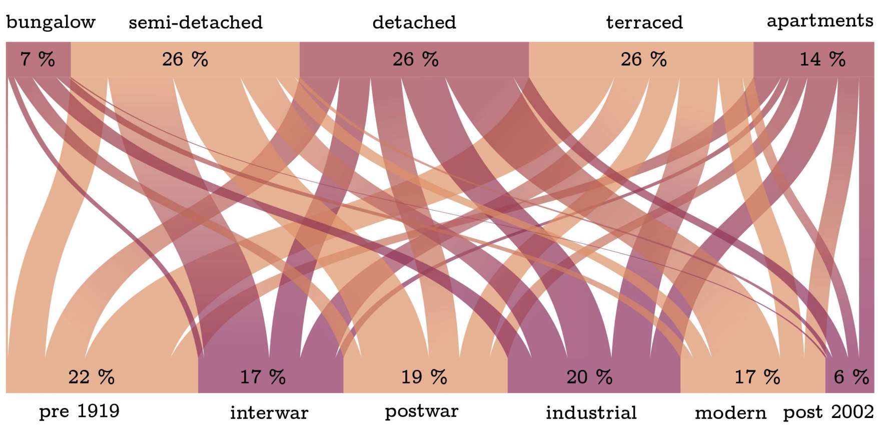
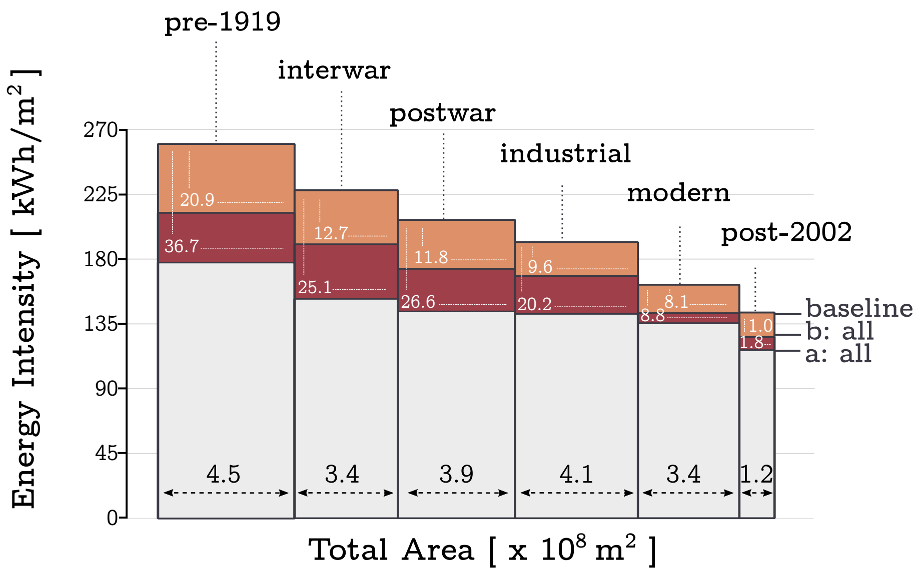
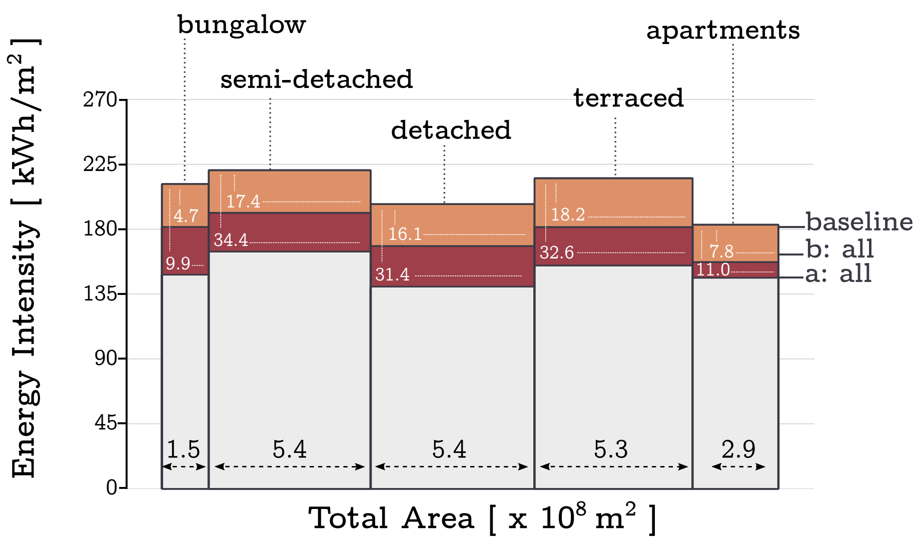


Figure 11: English Housing Stock - Total Floor Area by Archetypes



(a) Intensity by epoch



(b) Intensity by type

Figure 12: Reduction of energy by archetypes (stock indicator)
Total Area (→) x Energy Intensity (↑) = Domestic Sector Energy Demand (□). EnHub estimates this to be 428 TWh for the baseline case. Total energy demand reduction is indicated in white.

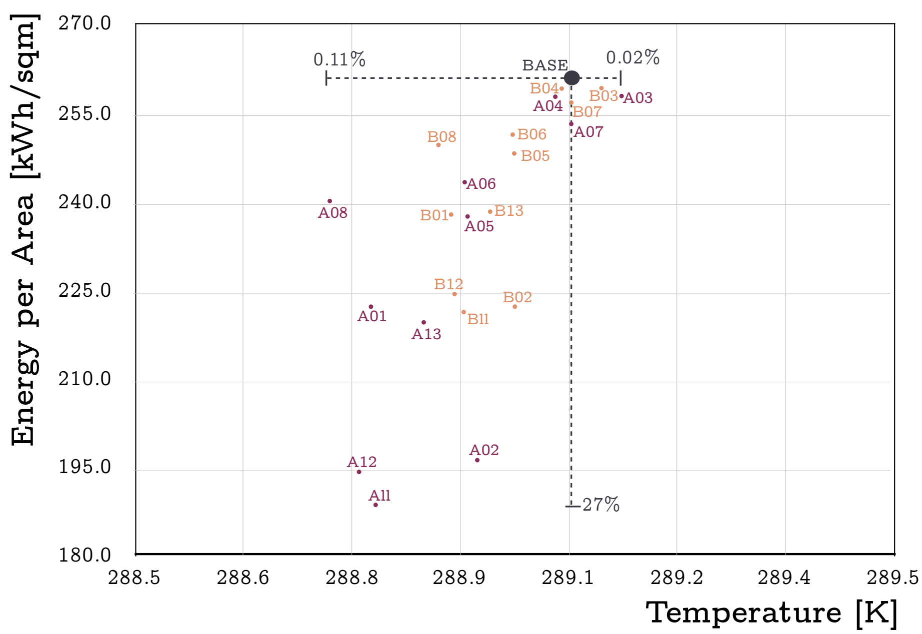


Figure 13: Variations in Energy Intensity and Indoor Temperature

Table 1: Reduction Variables

Id	**Category**	**Values**
C_1	no. floors	[GF, 1F, 2F, 3F]
C_2	basement	[present, absent]
C_3	attic	[present, absent]
C_4	attachments	[left, right, both, no]
C_5	regions	[north, central, south]
C_6	heater type	[gas, electric, other]
C_7	tenure	[private, social]
C_8	built age	[pre 1920, pre 1975, modern]

NOTE: The initial reduction considers the combination of the first group of variables (above dashed line): 4x2x2x4

Appendix A. Statistical Analysis

Table A.6 summarises the process of backwards elimination performed to identify redundancy in the EHS data set.

[Table 6 about here.]

[Table 7 about here.]

Table 2: KWt summary

variable	**name**	**p.value ↓**	**significance ↑**
felextrf	External insulation (right side)	0.027	yes
.attach_right	Attachment (right side)	0.029	yes
ageoldx	Age of oldest person	0.031	yes
unoc	Under occupancy	0.038	yes
pyngx	Age of youngest person	0.043	yes
.main_fa_xtra	Main Extra Construction	0.047	yes
finnofir	Total number of open fireplaces	0.051	not
.attachR	Attachment (right side)	0.056	not
DHWSystem	DWH system	0.112	not
mainfuel	Main Fuel	0.120	not
finchtyp	Main Heating System	0.129	not
⋮			
TFA	Total Floor Area	0.409	not
.insulation_front	Insulation (front wall)	0.909	not
FstFStoreyHeight	1F height	0.944	not
GFStoreyHeight	GF height	0.965	not

NOTE: The Table presents an abridged tabulation of the variables' significance. Most of these variables are computed directly from the raw EHS datasets; however, there are derived (or processed) variables that complement the reference survey data.

Table 3: `EnergyPlus` typology-specific components with UML reference

idf sections / UML blocks	A1	A2a	A2b	A2c	B1	B2	C1	C2	D1a	D2b	D2a	D2b	D3a	D3b	D4	E1	E2a	E2b	F1	F2	F3	G1	G2
A	x	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
B	x	.	.	.	.	.	.	x	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	.	.	.	.	.	.	x	x	.	.	.	.	.	.	.	.	.	.	.	.	.	x	x
D	x	.	x	x	x	.	.	x	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E	x	x	x	x	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F	x	.	.	.	x	x	.	.	x	.	.	.	.	.	x	x	x	x	.	.	.	.	.
G	.	.	.	.	.	x	x	x	x	x	x	x	.	.	.	.	.	.	.	.	.	x	x
H	x	.	.	.	.	.	.	x	.	.	.	.	x	x	x	.	.	.	.	.	.	x	x
I	x	.	.	.	.	.	x	x	.	.	.	.	.	.	.	x	x	.	x	x	x	.	.

A **id: Simulation Parameters**
- *timestep* resolution
- *building name* as global identifier
- *surface convection* algorithm
- *heat balance* algorithm

B **id: Location and Period**
- *location* information
- linked weather file
- simulation *period*
- seasonal attributes

C **id: Schedules associated**
- linked *schedules*
- simulation *timestep*

D **id: Surface and Constructions**
- *materials* catalogue
- *constructions* catalogue
- specific openings properties
- specific values for walls, floors, roofs

E **id: Thermal Zones and Geometry**
- opening areas and walls
- faade dimensions
- internal heights
- zone lists

F **id: Internal Gains**
- occupants properties
- equipment properties

G **id: HVAC**
- main heating system type
- additional heating system
- heat zone assignment

H **id: DHW System**
- dwelling water demand
- heat gains from DHW

I **id: Outputs**
- energy demand by end-use
- fuel demand
- heat gains
- comfort estimates

NOTES: Each idf section summarises the main parameters included in the EnergyPlus structure. A more detailed summary of cross-referenced components is provided in Figure 6

Table 4: Selection of prominent national policies to test scenarios for perfect and conditional uptakes

id	*Scenario*	*Status*
A: perfect uptake		
A01	**Solid wall insulation**	Non-insulated: 47%; semi-upgraded: 21%; efficient 32%
A02	**Loft insulation**	Non-insulated: 45%; semi-upgraded: 35%; efficient 20%
A03	**Double glazing**	Non-efficient windows: 45%; semi-upgraded: 35%; efficient 20%
A04	**Cylinder insulation**	Non-insulated:15%; below 50mm: 50%; efficient:35%
A05	**Draught proofting**	Below permitted: 65%; efficient: 35%
A06	**Low energy lights**	Non-efficient lighting: 35%; semi-upgraded: 50%; efficient 15%
A07	**Efficient household appliances**	Non-efficient appliances: 50%; semi-upgraded: 45%; efficient 5%
A08	**User Behaviour**	Typical profiles based on national studies.
B: conditional uptake		
B01	**Solid wall insulation**	This represents: 60% of the A01 upgrade.
B02	**Loft insulation**	This represents: 45% of the A02 upgrade.
B03	**Double glazing**	This represents: 56% of the A03 upgrade.
B04	**Cylinder insulation**	This represents: 46% of the A04 upgrade.
B05	**Draught proofting**	This represents: 33% of the A05 upgrade.
B06	**Low energy lights**	This represents: 40% of the A06 upgrade.
B07	**Efficient household appliances**	This represents: 40% of the A07 upgrade.
B08	**User Behaviour**	Test different assumptions on occupancy and minimum EUIs

NOTES: Assumptions based on both SAP catalogue and The Building Regulations. **01** emulates U-values below 0.3 (W/m^2K); **02** emulates a thickness of 270mm and U-value below 0.2 (W/m^2K); **03** emulates a U-value of at least 3.1 (W/m^2K) as established in The Building Regulations; **04** emulates a cylinder insulation over 55mm; **05** emulates a maximum infiltration level of 0.25 ach; **06** emulates low-energy lighting at 100%; **07** emulates low-energy for average household appliances; **08** emulates behaviour-related schedules associated to household appliances usage and metabolic gains.

Table 5: Scenarios results

(a) Perfect Uptake

	BASE	**A01**	**A02**	**A03**	**A04**	**A05**	**A06**	**A07**	**A08**	**A01+02**	**A01+03**	**ALL**
Energy Demand [TWh]	428	365	322	423	423	390	399	415	394	319	361	311
Internal Temperature $[K]$	289.1	0.26	0.12	-0.07	0.02	0.14	0.14	0.00	0.32	0.28	0.19	0.26
Emissions $[\frac{tCO_2}{hh}]$	5.7	-18%	-31%	-2%	-1%	-9%	-7%	-3%	-8%	-32%	-20%	-35%
Energy per area $[\frac{kWh}{sqm}]$	261	222	196	258	258	237	243	253	240	194	219	188
Energy per capita $[\frac{MWh}{p}]$	11.0	9.4	8.3	10.9	10.9	8.4	8.4	11.0	10.1	8.2	9.3	8.0
Uptakers [%]	-	68	79	45	65	65	85	95	-	79	68	100

(b) Conditional Uptake

	BASE	**B01**	**B02**	**B03**	**B04**	**B05**	**B06**	**B07**	**B08**	**B01+02**	**B01+03**	**BLL**
Energy Demand [TWh]	428	390	365	425	425	407	412	421	462	368	391	363
Internal Temperature $[K]$	289.1	0.16	0.07	-0.04	0.01	0.08	0.08	0.00	0.18	0.15	0.11	0.14
Emissions $[\frac{tCO_2}{hh}]$	5.7	-11%	-19%	-1%	-1%	-5%	-4%	-2%	8%	-17%	-11%	-19%
Energy per area $[\frac{kWh}{sqm}]$	261	237	222	259	259	248	251	256	249	224	238	221
Energy per capita $[\frac{MWh}{p}]$	11.0	9.4	8.3	10.9	10.9	9.6	9.6	11.0	10.5	9.5	10.1	9.3
Uptakers [%]	-	36	32	24	29	52	84	94	-	36	36	100

Abbreviations

ASHRAE American Society of Heating, Refrigerating, and Air-Conditioning Engineers

BREDEM Building Research Establishment Domestic Energy Model

CCC Committee on Climate Change

CHM Cambridge Housing Model

CLI Command Language Interface

DHW domestic hot water

DOMVENT Domestic Ventilation Model

EHS English Housing Survey

EnHub housing stock Energy Hub

EPW `EnergyPlus` Weather

FPCA Focused Principal Component Analysis

GHG Greenhouse Gases

GLM Generalised Linear Model

GUI Graphical User Interface

HEED Home Energy Efficiency Database

HPC High Performance Computing

HSEM Housing Stock Energy Model

HVAC Heating, Ventilation and Air Conditioning

idf EnergyPlus Input Data File

KWt Kruskal-Wallis H Test

LHS Latin Hypercube Sampling

No-MASS Nottingham Multi-Agent Stochastic Simulation

NUTS Nomenclature of Territorial Units for Statistics

OAT One-at-a-time

OOM Object-Oriented Modelling

PCA Principal Component Analysis

PAF Postcode Address File

SA Sensitivity Analysis

SOA Super Output Area

TRY Test Reference Year

UK United Kingdom

UML Unified Modelling Language

TFA Total Floor Area

References

[1] Parliament of the United Kingdom, Climate Change Act 2008, Tech. rep. (2008). `doi:10.1136/bmj.39469.569815.47`.

[2] Department of Energy & Climate Change (DECC), Digest of United Kingdom Energy Statistics 2013, Crown Copyright, London, 2013.

[3] P. S. Mallaburn, N. Eyre, Lessons from energy efficiency policy and programmes in the UK from 1973 to 2013, Energy Efficiency 7 (1) (2014) 23–41. `doi:10.1007/s12053-013-9197-7`.

[4] Department of Energy & Climate Change (DECC), Energy Consumption in the United Kingdom (ECUK), Tech. Rep. July, Department of Energy and Climate Change, London (2013).

[5] L. G. Swan, V. I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews 13 (8) (2009) 1819–1835.

[6] M. Kavgic, A. Mavrogianni, D. Mumovic, A. J. Summerfield, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, Building and Environment 45 (7) (2010) 1683–1697.

[7] M. Hinnells, B. Boardman, S. J. Darby, G. Killip, R. Layberry, Transforming UK homes : achieving a 60 % cut in carbon emissions by 2050, in: European Council for an Energy Efficient Economy, no. June, 2007, pp. 1105–1109.

[8] L. Collins, S. Natarajan, G. J. Levermore, Climate change and future

energy consumption in UK housing stock, Building Service Engineering 31 (1) (2010) 75–90. doi:10.1177/0143624409354972.

[9] G. Sousa, B. M. Jones, P. A. Mirzaei, D. Robinson, A review and critique of UK housing stock energy models, modelling approaches and data sources, Energy and Buildings 151 (2017) 66–80. doi:10.1016/j.enbuild.2017.06.043.

[10] R Core Team, R: A Language and Environment for Statistical Computing (2016).

[11] R. K. Strand, D. B. Crawley, C. O. Pedersen, R. J. Liesen, L. K. Lawrie, F. C. Winkelmann, W. F. Buhl, Y. J. Huang, D. E. Fisher, EnergyPlus : A New-Generation Energy Analysis and Load Calculation Engine for Building Design Introduction : What is EnergyPlus ?, in: ACSA Technology Conference, Vol. 2, NREL, 2000.

[12] B. Jones, P. Das, Z. Chalabi, M. Davies, I. Hamilton, R. Lowe, A. Mavrogianni, D. Robinson, J. Taylor, Assessing uncertainty in housing stock infiltration rates and associated heat loss: English and UK case studies, Building and Environment 92 (2015) 644–656. doi:10.1016/j.buildenv.2015.05.033.

[13] P. Steadman, S. Evans, M. Batty, Wall area, volume and plan depth in the building stock, Building Research & Information 37 (5-6) (2009) 455–467. doi:10.1080/09613210903152531.

[14] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer-Verlag New York, 2009. doi:10.1007/978-0-387-84858-7.

[15] B. S. Everitt, S. Landau, M. Leese, D. Stahl, Cluster Analysis, Wiley, 2011.

[16] IEA-ECBCSP (International Energy Agency Energy Conservation in Buildings and Community Systems Programme), Stock aggregation, in Annex 31 Energy related environmental impact of buildings, Tech. rep., IEA (2005).

[17] S. Natarajan, J. Padget, L. Elliott, Modelling UK domestic energy and carbon emissions: an agent-based approach, Energy and Buildings 43 (10) (2011) 2602–2612. doi:http://dx.doi.org/10.1016/j.enbuild.2011.05.013.

[18] I. Ballarini, S. P. Corgnati, V. Corrado, Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project, Energy Policy 68 (2014) 273–284. doi:10.1016/j.enpol.2014.01.027.

[19] T. Loga, B. Stein, N. Diefenbach, TABULA building typologies in 20 European countriesMaking energy-related features of residential building stocks comparable, Energy and Buildings 132 (2016) 4–12. doi:10.1016/j.enbuild.2016.06.094.

[20] K. J. Baker, R. M. Rylatt, Improving the prediction of UK domestic energy-demand using annual consumption-data, Applied Energy 85 (6) (2008) 475–482.

[21] I. G. Hamilton, P. J. Steadman, H. Bruhns, A. J. Summerfield, R. J. Lowe, Energy efficiency in the British housing stock: Energy demand

and the Homes Energy Efficiency Database, Energy Policy 60 (2013) 462–480.

[22] A. Mastrucci, O. Baume, F. Stazi, U. Leopold, Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam, Energy and Buildings 75 (2014) 358–367. doi:10.1016/j.enbuild.2014.02.032.

[23] J. Palmer, N. Terry, S. Firth, T. Kane, D. Godoy-shimizu, P. Pope, Energy use at home : models, labels and unusual appliances, Tech. Rep. October, Cambridge Architectural Research Ltd, Loughborough University (2013).

[24] P. J. Huber, E. M. Ronchetti, Robust Statistics, Vol. 523, Wiley, 2009. doi:10.1002/0470010940.

[25] J. N. R. Jeffers, Two Case Studies in the Application of Principal Component Analysis, Journal of the Royal Statistical Society. Series C: Applied Statistics 16 (3) (1967) 225–236.

[26] B. Falissard, Focused Principal Component Analysis: Looking at a Correlation Matrix With a Particular Interest in a Given Variable, Journal of Computational and Graphical Statistics 8 (4) (1999) 906–912.

[27] P. McCullagh, J. A. Nelder, Generalized Linear Models, 2nd Edition, Chapman and Hall/CRC, 1989. doi:10.1007/978-1-4899-3242-6.

[28] S. Nicol, C. Beer, C. Scott, The Age and Construction of English Housing, IHS BRE Press, 2014.

[29] Department of Energy & Climate Change (DECC), Energy Consumption in the United Kingdom (ECUK) : domestic data tables (2013).

[30] P. Wyatt, A dwelling-level investigation into the physical and socio-economic drivers of domestic energy consumption in England, Energy Policy 60 (2013) 540–549. doi:10.1016/j.enpol.2013.05.037.

[31] E. R. Scoffham, The shape of British housing, London, 1984.

[32] V. Brechling, S. Smith, The pattern of Energy Efficiency Measures amongst Domestic Households in the UK, Tech. rep., Institute for Fiscal Studies, London, UK (1992).

[33] UK Government, Building Act 1984 (1984).

[34] J. Palmer, I. Cooper, United Kingdom Housing Energy Fact File, Tech. rep., Department of Energy and Climate (2013).

[35] Department of Energy & Climate Change (DECC), The Government's Standard Assessment Procedure for Energy Rating of Dwellings, Tech. Rep. October, Building Research Establishment, Watford (2012).

[36] B. Minasny, A. B. McBratney, A conditioned Latin hypercube method for sampling in the presence of ancillary information, Computers and Geosciences 32 (9) (2006) 1378–1388. doi:10.1016/j.cageo.2005.12.009.

[37] F. Zhao, S. H. Lee, G. Augenbroe, Reconstructing building stock to replicate energy consumption data, Energy and Buildings 117 (2016) 301–312. doi:10.1016/j.enbuild.2015.10.001.

[38] K. J. Lomas, H. Eppel, Sensitivity analysis techniques for building thermal simulation programs, Energy and Buildings 19 (1) (1992) 21–44. doi:10.1016/0378-7788(92)90033-D.

[39] M. S. de Wit, Identification of the important parameters in thermal building simulation models, Journal of Statistical Computation and Simulation 57 (1-4) (1997) 305–320. doi:10.1080/00949659708811814.

[40] I. Macdonald, P. Strachan, Practical application of uncertainty analysis, Energy and Buildings 33 (3) (2001) 219–227. doi:10.1016/S0378-7788(00)00085-2.

[41] M. S. de Wit, G. Augenbroe, Analysis of uncertainty in building design evaluations and its implications, Energy and Buildings 34 (9) (2002) 951–958. doi:10.1016/S0378-7788(02)00070-1.

[42] A. Saltelli, Global sensitivity analysis : the primer, John Wiley, Chichester, 2008.

[43] S. K. Firth, K. J. Lomas, A. J. Wright, Targeting household energy efficiency measures using sensitivity analysis, Building Research & Information 38 (1) (2010) 31. doi:10.1080/09613210903236706.

[44] G. Calleja Rodríguez, A. Carrillo Andrés, F. Domínguez Muñoz, J. M. Cejudo López, Y. Zhang, Uncertainties and sensitivity analysis in building energy simulation using macroparameters, Energy and Buildings 67 (2013) 79–87. doi:10.1016/j.enbuild.2013.08.009.

[45] M. Kavgic, D. Mumovic, A. J. Summerfield, Z. Stevanovic, O. Ecim-Djuric, Uncertainty and modeling energy consumption: Sensitivity analysis for a city-scale domestic energy model, Energy and Buildings 60 (2013) 1–11. doi:10.1016/j.enbuild.2013.01.005.

[46] W. Tian, A review of sensitivity analysis methods in building energy analysis, Renewable and Sustainable Energy Reviews 20 (2013) 411–419. doi:10.1016/j.rser.2012.12.014.

[47] D. M. Hamby, A review of techniques for parameter sensitivity analysis of environmental models., Environmental monitoring and assessment 32 (2) (1994) 135–54. doi:10.1007/BF00547132.

[48] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity analysis in practice: a guide to assessing scientific models, Wiley, 2004.

[49] P. Heiselberg, H. Brohus, A. Hesselholt, H. Rasmussen, E. Seinre, S. Thomas, Application of sensitivity analysis in design of sustainable buildings, Renewable Energy 34 (9) (2009) 2030–2036. doi:10.1016/j.renene.2009.02.016.

[50] M. D. Morris, Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics 33 (2) (1991) 161–174. doi:10.1080/00401706.1991.10484804.

[51] M. Hughes, J. Palmer, V. Cheng, D. Shipworth, Sensitivity and uncertainty analysis of England's housing energy model, Building Research & Information 41 (2) (2013) 156–167. doi:10.1080/09613218.2013.769146.

[52] J. Buckley, S. King-Hele, What is weighting?, Tech. rep., UK Data Service, University of Essex and University of Manchester (2014).

[53] M. Page-Jones, L. L. Constantine, Fundamentals of Object-Oriented Design in UML, Addison-Wesley Professional, Boston, US, 2000.

[54] Object Management Group, OMG Unified Modeling Language, Tech. rep. (2015).

[55] A. Mhalas, Assessment and decision support for energy performance improvement of dwellings: framework and prototype development, Phd thesis, Teesside University (2013).

[56] B. D. B. Crawley, L. K. Lawrie, EnergyPlus : Energy Simulation Program, Ashrae 42 (4) (2000) 49–56. doi:10.1.1.122.6852.

[57] J. Henderson, J. Hart, BREDEM 2012 A technical description of the BRE Domestic Energy Model, Tech. rep. (2012).

[58] M. Hughes, P. Armitage, J. Palmer, A. Stone, Converting English Housing Survey Data for Use in Energy Models, Tech. Rep. January, Cambridge Architectural Research Ltd. and University College London (2012).

[59] Department of Energy & Climate Change (DECC), Energy Consumption in the United Kingdom (ECUK) : data tables (2013).

[60] D. Virk, M. Eames, CIBSE Weather Files 2016 release: Technical Briefing and Testing, Tech. rep., Chartered Institution of Building Services Engineers (CIBSE) (2016).

[61] L. D. Shorrock, J. Henderson, J. Utley, Reducing carbon emissions from the UK housing stock, BRE Press, 2005.

[62] E. Foda, D. Allinson, K. J. Lomas, D. Loveday, Using Uk Green Deal Assessment Data Of A Stock Of Dwellings To Run A Batch Of Building Energy Models, in: Building Simulation and Optimization Conference, no. Xml, 2014.

[63] Department of Energy & Climate Change (DECC), Summary of Analysis using the National Energy Efficiency Data-Framework: Part I Domestic Energy Consumption (June 2013).

[64] J. Chapman, P.-O. Siebers, D. Robinson, On the multi-agent stochastic simulation of occupants in buildings, Journal of Building Performance Simulation (2018) 1–18doi:10.1080/19401493.2017.1417483.

Table A.6: GLM rating derived from the stock energy evaluation in EnHub

Cat[n]	Variable (disaggregated)	Electricity Demand	Gas Demand	Internal Temperature	Cat	Variable (disaggregated)	Electricity Demand	Gas Demand	Internal Temperature
		(a)	(b)	(c)			(a)	(b)	(c)
ctx	ageband1900-1918	-133.618 (389.421)	8,482.547*** (2,262.874)	0.409* (0.231)		cellarwxtra	0	2,532.359*** (730.670)	0.490*** (0.107)
	ageband1919-1944	-629.732* (328.798)	-11,876.630 (8,643.790)	0.239 (0.236)		facadeheight	273.043*** (51.018)		0.149*** (0.037)
	ageband1945-1964	73.834 (321.574)	-430.885 (845.892)	0.054 (0.236)		felcavlfQuestion Not Applicable	5,684.806** (2,509.387)	-3,015.674 (5,396.284)	
	ageband1965-1974	100.749 (343.880)	-21,779.950*** (4,980.188)	-0.668** (0.275)		felcavlfYes	-95.957 (250.376)	1,245.668** (542.601)	
	ageband1975-1980	-360.280 (494.311)	-2,959.396** (1,240.613)	-1.012*** (0.309)		felextbfQuestion Not Applicable		-2,164.365* (1,228.670)	
	ageband1981-1990	-1,244.199*** (415.128)	-2,817.250*** (1,035.777)	-1.237*** (0.294)		felextbfYes		-5,080.463** (2,066.522)	
	ageband1991-1995	-1,441.712** (660.545)	-1,940.470 (1,541.450)	-1.342*** (0.333)		felextffQuestion Not Applicable		-8,095.438** (3,940.779)	0.480 (0.519)
	ageband1996-2002	-1,099.831 (706.980)	-4,849.826*** (1,605.351)	-1.686*** (0.346)		felextffYes		2,739.233 (2,169.354)	0.522*** (0.167)
	agebandPost 2002	-699.838* (398.058)	-3,972.823*** (995.753)	-1.200*** (0.292)		felextlfQuestion Not Applicable	-5,705.763** (2,510.862)	3,111.386 (5,404.663)	
	agebandPre 1850	652.560** (322.442)	-1,230.242* (743.533)	0.062 (0.104)		felextlfYes	-1,232.903** (555.795)	3,559.246** (1,695.976)	
	arnatxother urban centre	508.858 (441.301)	-1,226.125 (976.276)	-0.152 (0.138)		flatarea	-598.199** (249.916)	-1,552.052*** (564.090)	-0.135* (0.080)
	arnatxrural	665.680 (603.577)	-4,597.320*** (1,376.714)	-0.827*** (0.188)		flatareax	605.104** (255.388)	1,640.602*** (584.958)	0.131 (0.082)
	arnatxrural residential	1,384.238** (624.434)	458.260 (1,401.394)	0.704*** (0.197)		flatblocks	-823.425*** (123.492)	-362.117** (174.361)	-0.500*** (0.095)
	arnatxsuburban residential	759.530* (426.704)	-672.271 (960.061)	-0.123 (0.137)		flatd		-807.233* (424.288)	
	arnatxvillage centre	1,813.287*** (574.279)	-1,721.748 (1,289.697)	-0.315* (0.183)		flatfloors		764.245 (579.051)	0.287*** (0.095)
	felorienNorth		3,274.620*** (799.526)	0.454*** (0.113)		flath	-939.991 (691.929)	6,037.718*** (1,524.396)	-0.826*** (0.231)
	felorienNorth-east		3,141.772*** (955.304)	0.175 (0.133)		flatw		-773.434* (424.761)	
	felorienNorth-west		3,283.262*** (990.398)	0.353** (0.138)		floors		6,312.159*** (971.071)	-0.436*** (0.163)
	felorienSouth		1,627.948* (877.513)	0.238* (0.122)		housestoreyx		-2,084.330** (860.011)	0.235* (0.122)
	felorienSouth-east		2,547.355*** (985.400)	0.287** (0.140)		isattic		2,646.479*** (525.979)	0.330*** (0.080)
	felorienSouth-west		2,130.764** (912.852)	0.439*** (0.130)		isbungalow			-1.979*** (0.163)
	felorienUnknown		19,412.910*** (7,019.450)	-2.370*** (0.538)		iscellar			1.418*** (0.115)
	felorienWest		1,669.975* (915.189)	0.326** (0.130)		lsf		812.543*** (266.981)	0.117** (0.055)
	fmoduleHouse (single unit)		-247.084 (1,065.942)	-0.232 (0.247)		mainfa		-9.688** (4.637)	-0.004*** (0.002)
	fmodulePurpose built flats (multiple units)		1,625.671* (930.768)	0.742** (0.334)		mainfaxtra			0.012*** (0.004)
	northaxis	-2.143** (1.015)		-0.001 (0.0003)		storeysd		343.387*** (109.681)	0.040** (0.016)
	regionEast Midlands	96.216 (533.030)	-59.336 (1,174.576)	-0.674*** (0.162)		storeysdxtra	97.679* (58.020)		
	regionLondon	604.959 (431.783)	176.792 (952.167)	-1.014*** (0.138)		storeysh	-1,445.879** (648.238)	6,945.203 (4,331.311)	2.005*** (0.669)
	regionNorth East	958.309 (587.260)	-829.361 (1,277.758)	-1.572*** (0.180)		storeyshxtra		-8,748.733** (3,931.356)	-2.617*** (0.611)
	regionNorth West	968.633** (424.720)	1,310.224 (944.440)	-1.791*** (0.132)		storeysw	277.440*** (49.060)	571.876*** (133.378)	0.113*** (0.019)
	regionSouth East	279.611 (419.133)	-1,187.339 (927.938)	-0.593*** (0.132)		TFA			-0.007*** (0.001)
	regionSouth West	1,141.305** (446.948)	-1,653.824* (983.404)	-0.150 (0.140)		typeFLAT - Non-residential plus flat			-0.235 (0.443)
	regionWest Midlands	189.217 (513.133)	780.760 (1,145.227)	-1.351*** (0.161)		typeFLAT - Purpose built			-0.602* (0.320)
	regionYorkshire and the Humber	293.968 (464.006)	1,950.620* (1,029.346)	-1.534*** (0.141)		typeHOUSE - Detached			-0.704*** (0.250)
gmt	attachleft	-391.851* (236.908)				typeHOUSE - End terrace			-0.502* (0.256)
	AttachStateLeft		2,572.780*** (579.114)	-0.262*** (0.095)		typeHOUSE - Mid terrace			-0.670*** (0.258)
	AttachStateNone		-2,276.800 (1,963.311)	-1.400*** (0.292)		typeHOUSE - Semi-detached			-0.505** (0.248)
	AttachStateRight		2,672.694*** (640.036)	-0.515*** (0.099)		typeHOUSE - Temporary			-1.482** (0.717)
	cavityfront	1,927.267** (953.862)	-1,171.927 (762.740)			typerstrflat			0.223 (0.197)
	cavityleft	-1,654.055* (854.971)		-0.199 (0.137)		typerstrmansard			0.246 (0.276)
	cavityright	-1,904.343** (871.570)				typerstrmixed types			0.162 (0.251)
	cellard	-245.193** (105.304)				typerstrpitched			-0.047 (0.168)
	cellarh	8,290.003** (3,889.835)		2.424* (1.275)		typewindouble-glazed- UPVC		-1,842.613* (1,000.875)	
	cellarhxtra	-4,052.833* (2,248.383)		-1.214* (0.699)		typewindouble-glazed- wood		-1,633.188 (1,213.154)	
	cellarw		396.782 (283.693)			typewinmixed types		-813.180 (1,208.819)	

NOTES: *p<0.1; **p<0.05; ***p<0.01

[n] gmt: geometrical, hhd: household, ctx: contextual, hts: heating

Table A.6: GLM rating derived from the stock energy evaluation in EnHub - continuation

Cat[n]	Variable (dis)	Electricity Demand	Gas Demand	Internal Temperature	Cat	Variable (dis)	Electricity Demand	Gas Demand	Internal Temperature
		(a)	(b)	(c)			(a)	(b)	(c)
	typewinsingle-glazed- metal		-5,474.360*** (1,637.496)			Inccatlow	-2,272.356** (883.929)		-0.948*** (0.265)
	typewinsingle-glazed- UPVC		-3,061.403 (3,826.228)			Inccatmed	-1,296.777** (564.885)		-0.568*** (0.179)
	typewinsingle-glazed- wood casement		-903.250 (1,094.423)			litecost		12.320*** (2.821)	
	typewinsingle-glazed- wood sash		-1,342.717 (1,100.831)			spahcost	2.931*** (0.721)	3.599** (1.420)	0.0003*** (0.0001)
	wallconstruction		-314.297*** (55.034)	-0.037*** (0.008)		unocUnder occupying			0.104* (0.060)
	wallThickness	3,158.325 (2,137.725)		1.632** (0.679)		wathcost	4.012*** (1.153)		
hhd	agehrpx	127.945*** (44.833)	-26.602 (19.553)		hts	finchbcd	-0.337** (0.166)	0.525 (0.344)	
	ageoldx	-120.718*** (44.934)				finchtypCommunal/CHP	-8,148.802*** (1,757.142)		-0.715*** (0.241)
	agepartx			0.003 (0.002)		finchtypElectric ceiling/underfloor	2,747.625* (1,415.476)		0.430 (0.342)
	allincx		0.167** (0.068)			finchtypQuestion Not Applicable	1,976.628 (1,981.790)		-0.304 (0.394)
	cookcost	14.192 (9.501)		0.008*** (0.002)		finchtypRoom heaters	4,747.898*** (1,247.743)		0.148 (0.183)
	eHIindpoverty		866.333* (479.746)			finchtypStorage heaters	2,087.242** (1,012.540)		0.287* (0.164)
	eHS	5,262.710** (2,624.621)	-994.118* (587.506)			finchtypWarm air	-146.093 (993.967)		0.430* (0.260)
	elecmopPre payment			-0.056 (0.079)		findistyRadiators	-2,159.239** (974.605)		
	elecmopStandard credit			-0.151** (0.069)		findistyUnderfloor	6,250.357*** (1,639.902)		
	finmhfueCommunal - from boiler	-4,143.596** (1,963.889)	901.710 (4,178.862)			finwhcylQuestion Not Applicable		-2,867.514 (1,892.430)	
	finmhfueElectricity - 10hr tariff	-4,048.115 (2,520.113)	5,551.307 (5,280.428)			finwhcylUnknown		479.019 (415.612)	
	finmhfueElectricity - 24hr tariff	-844.776 (3,029.931)	8,629.020 (6,210.958)			finwhmms12.5mm	-1,383.850* (745.333)	-2,191.281 (1,652.363)	
	finmhfueElectricity - 7hr tariff	-1,850.275 (2,206.837)	6,891.481 (4,542.809)			finwhmms25 mm	-141.379 (247.521)	-1,302.315** (545.759)	
	finmhfueElectricity - standard	-2,155.658 (2,229.460)	6,901.943 (4,615.493)			finwhmms38 mm	-2,126.907*** (659.002)	1,565.841 (1,428.598)	
	finmhfueGas - Bulk/LPG	-1,377.994 (2,304.633)	8,704.779* (5,002.130)			finwhmms50 mm	-2,219.884** (891.607)	542.889 (1,781.393)	
	finmhfueGas - Mains	-1,569.778 (2,500.844)	11,998.340** (5,246.026)			heat4xfixed room heating	-3,641.989*** (1,147.232)		
	finmhfueOil	-6,372.868*** (2,302.414)	-2,000.661 (4,332.658)			heat4xstorage heater	-1,318.605 (1,076.429)		
	finmhfueQuestion Not Applicable	-5,810.017** (2,828.069)	11,896.630** (5,792.274)			heaterredGas			0.253* (0.152)
	finmhfueSolid fuel - anthracite	-1,579.105 (3,423.123)	-4,690.303 (7,223.806)			heaterredOil			0.318* (0.175)
	finmhfueSolid fuel - coal	-4,261.434 (3,324.294)	17,046.870** (7,301.591)			heaterredOther			0.198 (0.228)
	finmhfueSolid fuel - smokeless fuel	-5,820.910* (3,195.725)	-673.376 (6,400.846)			heaterredSolid			0.733*** (0.162)
	finmhfueSolid fuel - wood	-2,461.404 (2,340.024)	11,823.600** (4,930.865)			mainfuelbulk LPG	1,945.978 (3,598.372)	-17,664.080** (7,855.263)	
	finnofir		826.143** (328.304)			mainfuelcommunity heating from boilers	11,088.280*** (3,564.820)	-1,658.345 (7,420.785)	
	fpbasinc	0.065 (0.046)				mainfuelcommunity heating from CHP/waste heat	6,962.044* (3,633.674)	-4,040.558 (7,844.030)	
	fpflgbNot in FP - basic income definition	581.527* (334.254)	2,222.139*** (751.629)			mainfuelelectricity (10 hr. off peak)	6,386.972* (3,274.319)	-7,916.335 (7,105.703)	
	fpflgfNot in FP - full income definition	-768.360** (356.503)	-1,698.766** (759.975)			mainfuelelectricity (7 hr. off peak)	6,685.988** (3,211.061)	-10,389.010 (6,887.299)	
	fpfullinc	-0.092* (0.051)	-0.082 (0.056)	-0.00001 (0.00000)		mainfuelelectricity (7 hr. on peak)	4,714.173 (4,626.694)	-1,898.283 (9,227.019)	
	fpindf	-2,135.674* (1,289.204)		1.340*** (0.376)		mainfuelelectricity (standard tariff)	7,450.393** (3,158.378)	-11,035.800 (6,885.515)	
	fplihcgap	1.549** (0.617)				mainfuelgas (mains)	1,653.429 (2,781.317)	-6,828.595 (6,036.646)	
	fplihcqdtHigh income low costs		-2,222.905*** (576.592)			mainfuelheating oil	5,545.196 (3,462.841)	-9,192.091 (7,521.782)	
	fplihcqdtLow income high costs		46.263 (578.766)			mainfuelhouse coal	3,559.605 (3,749.324)	-23,362.920*** (8,232.796)	
	fplihcqdtLow income low costs		-1,821.565** (751.431)			mainfuelsmokeless fuel	4,299.161 (4,278.794)	-8,032.143 (9,274.553)	
	fuelexpn	-2.211*** (0.674)	-2.563* (1.310)			mainfuelwood	2,558.473 (3,068.993)	-17,550.650*** (6,657.916)	
	gasmopn/a = No gas		-2,034.873*** (745.060)			watersysElectric immersion heater	1,875.594*** (444.175)	-1,504.898 (951.205)	0.315** (0.135)
	gasmopPre payment		-2,675.785*** (689.754)			watersysInstantaneous	2,587.915*** (528.159)	-750.204 (1,106.023)	0.551*** (0.160)
	gasmopStandard credit		-1,629.660*** (611.569)			watersysWith central heating	809.960 (504.763)	696.070 (949.491)	0.191 (0.146)
	hhincx		-0.414*** (0.100)			Constant	-1,444.060 (6,124.506)	-20,724.350* (10,611.130)	17.058*** (1.676)
	hyeargrx		0.161*** (0.049)						

NOTES: *p<0.1; **p<0.05; ***p<0.01

[n] gmt: geometrical, hhd: household, ctx: contextual, hts: heating