Benchmarking Domestic Energy Consumption using High-Resolution Neighbourhood Energy Data and City Clustering in the UK

Grace Colverd
gb669@cam.ac.uk
Department of Engineering,
University of Cambridge
Cambridge, UK

Ronita Bardhan
Department of Architecture,
University of Cambridge
Cambridge, UK

Jonathan Cullen
Department of Engineering,
University of Cambridge
Cambridge, UK

Abstract

We present a novel approach to energy benchmarking for cities and towns in England and Wales. By combining high-resolution energy data at the neighbourhood level, using full postcode units of 5-150 households, with comprehensive building stock information, we develop a robust methodology to establish energy performance benchmarks. Our approach employs spectral clustering algorithms to group urban areas with similar building stock characteristics and socio-demographic profiles, ensuring meaningful comparisons. For each cluster, we calculate the Energy Use Intensity (EUI) at the neighbourhood level, which is defined as the annual gas consumption per unit of domestic floor area. This analysis accounts for key factors such as building age, typology, and socio-demographic indicators, enabling us to create context-specific benchmarks. These benchmarks offer valuable insights into the energy performance of different urban typologies within each cluster and assist in the identification of potential areas for targeted improvements. Our findings provide a data-driven framework for city planners, policy-makers, and researchers to assess and compare the energy efficiency of cities and towns at a granular level. The proposed methodology is fast, flexible and nationally scalable.

CCS Concepts

• **Applied computing → Engineering**.

Keywords

Energy Benchmarking, Neighbourhood Energy Consumption, Clustering

ACM Reference Format:
Grace Colverd, Ronita Bardhan, and Jonathan Cullen. 2024. Benchmarking Domestic Energy Consumption using High-Resolution Neighbourhood Energy Data and City Clustering in the UK. In *The 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BUILDSYS '24), November 7–8, 2024, Hangzhou, China.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3671127.3698183

BUILDSYS '24, November 7–8, 2024, Hangzhou, China

ACM ISBN 979-8-4007-0706-3/24/11
https://doi.org/10.1145/3671127.3698183

1 Introduction

The UK's approximately 30 million domestic buildings contribute significantly to the nation's greenhouse gas (GHG) emissions, accounting for an estimated 13% of the total in 2023 [12]. These emissions primarily stem from the combustion of natural gas and other fuels for heating and cooking. Since 1990, residential building emissions have decreased by 35.1%, largely due to changes in housing stock and shifts towards lower-carbon electricity sources. However, recent years have seen a stagnation in this downward trend. The 6.2% reduction observed from 2022 to 2023 is attributed more to warmer-than-average temperatures and demand-side changes driven by rising energy costs rather than improvements in energy efficiency [7]

The UK Government has set an ambitious target of reducing carbon dioxide emissions by 68% by 2030 compared to 1990 levels. However, as of June 2024, the country is not on track to meet this goal, despite a significant reduction in emissions in 2023 [7]. The Climate Change Committee (CCC) emphasizes that the building sector must increase its contribution to emissions reductions, yet only a third of the required reductions to achieve the 2030 target are currently supported by credible plans [7]. While the building sector has seen emissions fall from 105.5 $MtCO_2e$ in 2008 to 77.8 $MtCO_2e$ in 2022, most of this reduction occurred between 2008 and 2014, driven by governmental policy measures supporting energy efficiency investments [7]. Subsequent cuts to these programs have led to a sharp decline in emissions reductions. Energy efficiency measures like the adoption of heat pumps are significantly behind schedule. Currently, only 1% of homes are heated with heat pumps, far short of the 10% needed by 2030 as estimated by the CCC. To meet climate goals, heat pump installation rates in residential buildings must increase tenfold from 2023 levels by 2028 [7]. Clearly, a large upswing in the rate of building emission reductions is required if the UK is to reach its climate change goals, underscoring a need for effective policies and technologies to transform the energy efficiency of the nations building stock.

Reducing building emissions can be achieved in multiple ways: enhancing the energy efficiency of houses through building retrofit [3], transitioning from gas-based to electric-based heating [10], modifying occupier behaviour [5], and decreasing the carbon intensity of the energy supply [17]. Building an effective emissions reduction strategy requires a comprehensive understanding of the building stock and the manner in which energy is consumed. Given the scale of the problem, we are interested in methods of characterising the energy performance of the housing stock which have the potential to scale nationally.

Building Stock Modelling Approaches. Building stock modelling encompasses top-down and bottom-up approaches for residential energy consumption analysis [26]. Top-down methods aggregate the building stock into a single entity to estimate its consumption at scale. Bottom-up methods, considered more accurate, estimate individual building consumption and aggregate results upwards [4, 20]. These include physical and statistical methods, the former simulating energy consumption based on buildings' physical and occupancy parameters [4]. The latter encompasses traditional correlation and regression analysis methods, and machine learning-based prediction methods. Both methods typically use building archetypes for national-level applications. Representative, or 'archetypal' buildings are identified, energy performance predicted or simulated, and results scaled up. The archetype approach is widely used in bottom-up stock modelling for various purposes, including customized energy profiles [23], building stock classification [31], renovation planning [44] and fuel poverty assessment [43].

Issues exist with the archetypal approach in energy modelling. Average consumption and archetypes fail to capture the diversity in actual energy use [32] and oversimplify building characteristics [30]. Household-level studies reveal significant variations in energy consumption, largely due to occupant behaviour. Studies on electricity consumption in Denmark found large standard deviations in electricity use even within the same dwelling type [21], with some homes using 2-3 times more energy than similar ones [32]. These limitations of archetypal approaches highlight the importance of granular, real-world energy consumption data in capturing the true diversity of energy use patterns, which leads us to consider the challenges and opportunities presented by actual energy consumption datasets.

Actual Energy Consumption Data. In response to the limitations of archetypal approaches, researchers and policymakers are increasingly turning to actual energy consumption data to establish more accurate benchmarks and develop targeted emission reduction strategies. Building-level energy consumption datasets fall into either geolocated or anonymised, and their level of accessibility tends to depend on this classification.

More relaxed laws around data disclosure in the USA have led to a cluster of open-access geolocated building-level energy datasets in North America [24]. These datasets have been used to make predictions of energy usage at the building level, and assess drivers of energy performance [15, 47].

In the UK, privacy concerns mean access to geolocated building-level energy consumption data is tightly restricted. The NEED dataset (National Energy Efficiency Data Framework), a governmental building stock dataset which combines building attributes with annualised electricity and gas usage data, publishes a large sample of anonymised data, representative of the UK housing stock, and allows limited access to its full dataset. Government reporting using the NEED data reports on the relationship between building and occupant attributes and energy consumption. The NEED data has been extensively used to investigate the drivers of energy performance [46], assess the energy use intensity (EUI) of London buildings [27], create energy benchmarks for different building typologies [19] and assess the impact of building retrofits [1]. The benefits of such a source of actual consumption data cannot be overstated. Comparisons of the NEED data to a national energy stock model, the Cambridge Housing Model (CHM), revealed that the CHM energy model overestimated gas consumption, particularly for buildings built pre-1930. Dwellings with poor energy efficiency also consumed less gas than expected and hence using such models could overestimate the emission reduction impact of energy retrofits [42].

In addition to the anonymised or access-controlled NEED data, the UK publishes annual energy consumption figures available at sub-national scales. Regional, district and sub-district energy statistics have been published since 2010 [11]. This open-access data has led to a cluster of works investigating drivers of energy performance in the UK, including clustering for benchmarking [33, 48], and predictive modelling [2].

Energy Benchmarking and Clustering Applications. Energy benchmarks have been widely studied and applied in research. They serve to normalize consumption or account for the effects of specific explanatory variables. Three primary methods for creating energy benchmarks are statistical prediction models, simulations, and approaches using real consumption data. Simulations, which use physical models to simulate data, requiring detailed building information, are typically limited to small samples or archetypal reference buildings due to the data intensity requirements [6, 40]. However, these methods are not suitable for national-scale applications given the aforementioned issue with the archetype approach. Statistical models employ input variables to predict energy consumption. In the UK, regression models have been used to explain 65% of gas usage variance at the sub-district level [33]. However, regression methods have been criticized for oversimplifying input variables and inadequately addressing confounding effects among cross-correlated variables. More recent works have used neural networks to predict electricity and gas consumption at the urban scale in London, at a granularity of 400–1200 households. Results using physical, socio-demographic, and economic input variables with a multilayer neural network achieved an R^2 of 0.998 [2]. Other non-linear methods that have been used include non-linear clustering methods. For instance, Gaussian mixture models and hierarchical clustering have been used to analyze domestic gas consumption and inform policy recommendations at the district level in the UK [48]. Clustering has also been applied to estimate post-retrofit energy savings [16, 22] and identify key variables influencing energy use [9]. In the UK, real-energy consumption benchmarking methods have utilized restricted-access NEED data at the building level in London to create benchmarks for various residential typologies [19].

1.1 Proposed Work

We propose a novel methodology for creating customized energy benchmarks using open-access data to support targeted emissions reduction strategies at national scales. Our approach clusters cities and urban areas in England and Wales based on aggregate statistics known to drive energy consumption. We then normalize small-neighbourhood annual energy consumption by total building floor area to create in-cluster EUI benchmarks. These customized benchmarks and visualization capabilities enable policymakers to assess city performance relative to peers at both city-wide and small-neighborhood levels. Our research offers novelty in three key dimensions:

- Granularity: we generate EUI at the 5–150 household level, level, comparing to prior energy studies at the 400–1200 household level.
- Customization: We create custom benchmarks for urban areas using local energy consumption data rather than relying on data from a single city.
- Accessibility: We leverage open-access energy data and education-access building stock data, in contrast to prior works using restricted datasets

We apply this method to define energy benchmarks for domestic neighbourhoods across 111 cities and urban areas in England and Wales. A case study demonstrates how urban planners and local authorities could use this methodology to target energy retrofits and occupier behaviour interventions. We also compare our EUI results to benchmarks given in the literature.

2 Methods

To create meaningful energy benchmarks, we cluster cities and urban areas in England and Wales based on variables known to be related to building energy consumption. Analysis of NEED data [33], and prior work clustering regions within London for gas performance [2, 48] have identified building age, typology, and socio-demographic factors as key drivers of energy consumption. Our work follows several stages. Firstly, we generate variables relating to EUI, building typology and building age at the postcode level. Next, we aggregate these up to urban levels using mappings from postcode to Output Areas (OA)[1] level and up to cities. We also aggregate socio-demographic variables from OA to the city level. Next, we cluster cities based on building age, typology and demographics. Finally, we create energy benchmarks based on the EUI of postcodes within these cities for intra-cluster comparison. A graphical summary of our methodology is given in Figure 1. The following section details the data generation, the clustering methodology and subsequent comparison to existing benchmarks. Given London's distinctive built environment and the extensive research on neighbourhood energy performance [18, 48], and energy benchmarks specific to the capital [19], we have excluded London from our clustering analysis.

2.1 Data Generation

Our data generation process involves two key stages. First, we derive postcode-level attributes for building age, typology, and gas EUI. Second, we aggregate these derived variables to create city-level metrics. We also aggregate census data at the OA level to further enrich our city-level variables. To aggregate data for urban areas, we sum the counts of different building typologies and ages within postcodes matched to each urban area and then calculate overall percentages. This aggregation follows a hierarchical structure: postcode → OA → LSOA → City/Urban Area. Table 2 outlines

[1] OAs are the lowest level of geographical area for UK census statistics, which aggregate into Lower Super Output Areas (LSOAs) [35].

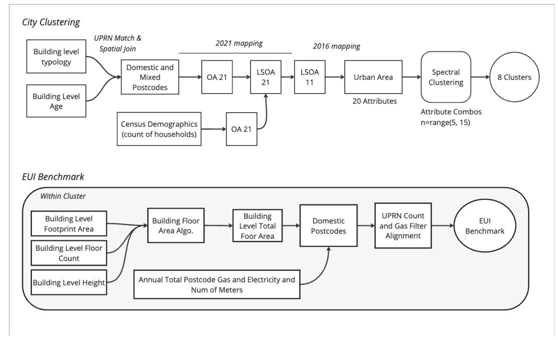

Figure 1: Process diagram

the specific mappings used in this process, all of which are sourced from the Office of National Statistics (ONS) [2].

2.1.1 Energy and Building Stock Variables. We derive Energy Use Intensity (EUI) from annual gas consumption, normalised by total domestic floor area per postcode. Annual postcode domestic energy consumption data is published by the UK Department for Energy Security and Net Zero (DESNZ) [13]. This dataset contains the count of meters per postcode and total, median and mean consumption per postcode, for both gas and electricity, for postcodes with at least five meters. In this analysis, we use annual total gas consumption G_Γ for a postcode Γ.

Total domestic floor area per postcode is calculated using the Verisk UKBuildings dataset[3]. This data provides building footprints, age, typology, floor count and a range of other variables. Buildings are linked to postcodes via the Unique Property Reference Number (UPRN) and a spatial join of building footprint polygons and postcode polygons (see an example of the postcode match in Figure 2). For a postcode Γ, the total floor area A is calculated $A_\Gamma = \sum_{i=0}^{k} (a \cdot f)_i$, where k is the number of domestic buildings in the postcode (excluding domestic outbuildings), a is the area of the building footprint and f is the number of floors. We apply various checks based on reasonable building parameters, e.g. height to base ratio and average floor height to clean the data. Further detail on this process is given in Appendix B. The building typology and age variables are calculated using the groupings given in Table 1, applied to the same building identified as for building floor area. We note that our choice of normalisation factor for EUI is impactful. Prior works have identified that alternate choice of normalisation (e.g. by the number of rooms) affected the distributions of energy performance [27]. Normalisation by floor area tended to give larger buildings smaller EUIs while normalising by rooms shifted in favour of terraces.

We calculate building stock attributes for 1.4M postcodes in England and Wales. Of these, 1.08M match on postcode with the aforementioned energy data. This represents 98% of the postcodes for which domestic gas data is available[4]. The unmatched postcodes likely represent areas with less than 5 meters, which are excluded from the energy dataset for privacy reasons, or areas with

[2] Sourced from 'https://geoportal.statistics.gov.uk/'
[3] Available through the EDINA Digimap Education Licence
[4] DESNZ file contains 1.1M postcodes

Figure 2: Postcode E10 7NP, with domestic buildings footprints. Visualisation code adapted from [28].

Variable	Description
% dwellings: Standard Size houses	Typology: Standard size semi-detached, Standard size detached
% dwellings: Estates	Typology: Planned balanced mixed estates and Linked and step linked premises
% dwellings: Small Terraces	Typology: Small low terraces, 2 storeys terraces w/rear extension and Semi type house in multiples
% of dwellings: Large Flats	Typology: Very tall point block flats and Tall flats 6-15 storeys
% of dwellings: Small -Medium Flats	Typology: Medium height flats 5-6 storeys and 3-4 storey and smaller flats
% of dwellings: Large houses	Typology: Very large detached, Large detached, Large semi detached and Tall terraces 3-4 storeys
% of Buildings Aged: Post-1999	Age buckets: Post 1999
% Buildings Aged: 1919-1999	Age buckets: 1919-1944, 1945-1959, 1960-1979, 1980-1989, 1990-1999
% Buildings Aged: Pre-1919	Age buckets: Pre 1919

Table 1: Building stock and typology attributes, used at urban level for clustering

Mapping Name	Date	Source
Postcode -> OA 2021	Aug 2023	ONS
OA 2021 -> LSOA 2021	July 2023	ONS
LSOA 2021 -> LSOA 2011	July 2023	ONS
LSOA 2011 -> City / Town	Dec 2015	ONS

Table 2: Mappings used to create city-level attributes. An updated LSOA to city mapping for the 2021 census was not available at the time of work, hence the conversion from LSOA 2021 -> LSOA 2011.

no domestic gas consumption (i.e. fully commercial). For the EUI benchmarking, the 1.08M of postcodes is filtered to wholly domestic postcodes, defined as postcodes where the building use is 100% residential. EUI E is calculated $E_\Gamma = \frac{G_\Gamma}{A_\Gamma}$ for a postcode Γ with units of kWh/m^2/yr, enabling the energy performance of postcodes with different building sizes to be compared. The total count of wholly domestic postcodes in England and Wales for which we can calculate a residential floor area and a matched gas consumption value is 726k, 67% of the 1.08M. We apply a filter based on the alignment of the count of UPRNs within the postcode (a proxy for the number of households within a postcode) and the count of gas meters within the postcode, given that the majority of households in the UK will have one gas and one electricity meter [19]. Of the 726,776 postcodes with both total floor area and gas data, 75% fall within a 9% difference in number of gas meters and the number of UPRNS, with a long tail up to the maximum difference of 4380%. To exclude postcodes where there is likely an error either on the building stock side or on the energy side, we apply a filter of 40% alignment, leaving 621,415 postcodes. Whilst energy analyses typically exclude based on total energy use, here we follow Godoy-Shimizu et al. [19] and filter based on annual EUI thresholds. We apply an annual gas EUI threshold of $5 < \text{EUI} < 500$ kWh/m^2. We apply some final thresholds to ensure a clean data sample: a count of buildings threshold of 1-200 and total building volume of 50-20,000m^2. These filters reduce the sample to 613,959, representing 55% of the postcodes with domestic gas consumption.

2.1.2 Socio-Demographic Variables. Socio-demographic attributes are sourced from the 2021 Census data for England and Wales at the OA level. City-level attributes are calculated by summing up the count of households with different census attribute values and calculated as overall percentages. Three census themes were chosen: economic activity, household composition and ethnicity.

2.2 Clustering Methodology

After attribute generation at the postcode level and subsequent aggregation to the city level, our dataset comprises 111 cities and urban areas in England and Wales. To create our benchmarks, we test two unsupervised clustering methodologies commonly utilized in the literature: Gaussian Mixture Models (GMM), and Spectral Clustering (SC). GMMs are probabilistic models that represent data as a combination of multiple Gaussian distributions. Each data point is assumed to originate from one of these distributions, with the probability determined by a weight assigned to each distribution. GMMs provide probabilistic cluster assignments, allowing for uncertainty in classification, particularly for data points near cluster boundaries [39]. Previously employed in energy benchmark creation, GMM also offers the advantage of automatically selecting the model parameters (covariance type and number of clusters). Spectral clustering is a technique that identifies clusters in data by leveraging the eigenvectors of a similarity matrix. It is useful for datasets with non-linearly separable or non-convex cluster shapes. While spectral clustering can be computationally expensive for large datasets, given the size of our dataset spectral clustering is feasible. Spectral clustering requires pre-specification of the number

City Clustering Variables
Percentage Standard Houses (%)
Percentage Small Terraces (%)
Percentage Estates (%)
Percentage Age Pre-1919 (%)
Percentage Age 1919-1999 (%)
Percentage One person Household (%)
Percentage White (%)
Percentage Asian (%)
Percentage Economically Active: Employed (%)
Percentage Economically Inactive (%)

Table 3: Final clustering attributes used in the analysis

of clusters and gives a non-probabilistic membership assignment [45],[41]. More detail on the steps of spectral clustering is given in Appendix A.0.1.

To determine the optimal clustering method, we implement both unsupervised clustering algorithms, using different metrics to help select optimal clustering settings within each method. We evaluate two metrics: the Silhouette Score (SS) and the Davies-Bouldin Score (DBS) supplemented by several heuristics relating to the intended use of the clusters. The SS characterises how compact and separated clusters are from each other, calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b), $S_i = \frac{(b_i - a_i)}{max(b_i, a_i)}$ and ranges from $-1 \leq SS \leq +1$, with higher values indicating better clustering [38]. SS is calculated at the member level and is averaged to create macro values for clusters. Negative values for a member indicate potential misclassification as a different cluster is more similar. The DBS is an average similarity measure of each cluster with its most similar cluster, where similarity is the ratio of within-cluster distances to between-cluster distances. DBS ranges from $0 < DBS < \infty$ with lower values indicating better clustering [8]. To account for initialisation sensitivity and get a sense of cluster stability, when calculating SS and DBS, we calculate the metric mean and standard deviation across 20 runs, with 100 initialisations, using Scikit-Learn. We test the numbers of clusters ranging from 5 to 20. This range balances granularity with practicality: more than 20 clusters would be unwieldy for policy recommendations and likely yield single-entry clusters, while less than 5 would not yield granular enough benchmarks. We use the metrics and heuristics for each clustering method to determine the best clustering parameters and then compare methods for robustness, stability and utility.

2.3 Benchmark Comparison

To validate the dataset construction and place our benchmarks in the context of the literature, we compare them to recent works on energy benchmarking in the UK. University College London (UCL) in partnership with the Chartered Institute of Building Survey Engineers (CIBSE), has published domestic energy benchmarks for different typologies of buildings, based on a large sample of actual energy consumption data for London houses and flats [19]. To compare our postcode-level dataset of EUI to the CIBSE EUI benchmarks for different types of buildings, we create categories of postcodes with uniform typologies. Three categories of postcodes

Category	Building Typologies Label
Flats	Very tall point block flats
	Tall flats 6-15 storeys
	Medium height flats 5-6 storeys
	3-4 storey and smaller flats
Terraces	2 storeys terraces with t rear extension
	Linked and step linked premises
	Planned balanced mixed estates
	Semi type house in multiples
	Small low terraces
	Tall terraces 3-4 storeys
Semi + Detached	Large semi-detached
	Standard size semi-detached
	Large detached
	Standard size detached
	Very large detached

Table 4: Categories of Postcode Typology. A postcode is classified as one of the categories if all residential buildings within (excluding outbuildings) fall within one category of typology.

are defined, which only contain residential buildings of certain typologies, i.e. a postcode can be categorised as 'Flats' if it only contains residential buildings with the corresponding building typology labels. The category description and mapping are given in Table 4,

There are several noteworthy differences between the CIBSE methodology and our methodology.

- CIBSE benchmarks use energy data matched to individual meters for households in London from 2016. We use postcode energy data from 2022.
- CIBSE have a sample of 808,559 dwellings, representing 32% of houses and 16% of flats in London. We use a sample of 160,824 postcodes from England and Wales, representing 15% of postcodes with domestic gas usage[5].
- CIBSE distinguish gas-heated dwellings and electricity-heated dwellings and provide separate benchmarks for each. Our method does not distinguish between gas and electricity-heated dwellings, so we instead include all postcodes which have gas data available.

3 Results

3.1 City Clusters

After testing both clustering methodologies, both GMM and SC suggested 8 clusters for our dataset based on metric evaluation. Both produced clusters of full utility, with no clustered with a small number of members. The proposed cluster membership in Principal Component Analysis (PCA) space can be seen in Figure 3. Computation time was not an issue given the small size of the clustering dataset: running both GMM and SC with 100 initialisations took less than a minute on an 8-core CPU with 16GB RAM.

To evaluate the effectiveness of the different clustering methodologies for energy benchmarking, we assessed each method's cluster discernment power relative to gas EUI. Cluster discernment

[5]Number of postcodes reduced from above due to limitations of singular typology

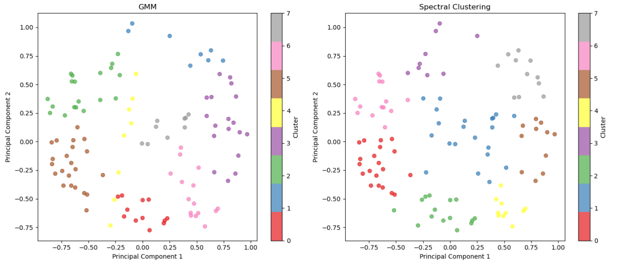

Figure 3: Cluster membership visualised in PCA space, for GMM and SC methods.

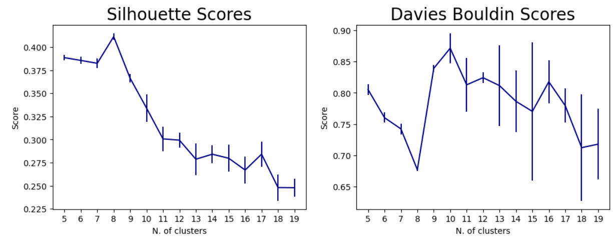

Figure 4: Metric results for SC clustering. Mean across 20 runs plotted with error bars showing STD, for SS and DBS.

power refers to the ability of a clustering method to create distinct, meaningful groups based on the variable of interest. Due to the non-normal distribution of EUI (observed in histogram visualization - skewed right), we used the Mann-Whitney U Test. For each clustering methodology, we conducted pairwise comparisons of median gas EUI between clusters. The null hypothesis for each test was that the gas EUI distributions of the compared clusters were identical. To supplement the significance testing and account for the large sample sizes, we calculated effect sizes using Cliff's Delta, a non-parametric measure. We chose Cliff's Delta due to its robustness with non-normal distributions [29, 34].

All clusters showed statistically significant pairwise differences across each clustering method. However, some of these differences corresponded to negligible effect sizes. When comparing methods, SC produced clusters with more pronounced differences in gas EUI, as evidenced by larger effect sizes [6]. Additionally, SC yielded clusters with smaller cluster-level standard deviations in gas EUI. Based on these findings, we selected SC as the optimal clustering methodology. Table 5 presents the mean values of city clustering variables for each cluster for the SC algorithm. Empirical cumulative distribution functions (ECDFs) are included in the Supplementary Information to provide visualisation of the attribute distributions.

3.1.1 *Energy Benchmarks.*

Returning to the gas EUI, statistical testing revealed three tiers of gas EUI for our 8 clusters: Very high (7), high (3), medium (1,5,6) and low (0,2,4). All clusters tested were significantly different to each other ($p<0.0001$). When evaluating the effect size, effects are negligible internally between the lowest two bands. Medium-large effects were seen between the low band and very high band, and small effects between lower-medium, medium-high and high-v.high bands. The energy benchmarks for gas EUI within the Clusters is given in Figure 5. Box plots are used to show the distribution of EUI, ranked from lowest to highest by median EUI. We cap the y-axis at 99th quartile for visualisation purposes. We now briefly summarise the clusters, employing relative descriptors (low, medium, high) to characterize the ranges of cluster variables.

3.1.2 *Cluster Descriptions.*

Very High EUI Cluster. Cluster 7 has the highest median gas EUI, at 94 kWh/m^2. It emerges as the cluster with very low economic activity, with some of the lowest employment levels. It is characterized by a high percentage of pre-1919 housing stock, higher levels of small terraces, and a low range of standard houses. Cluster 7 is the most ethnically diverse cluster. The older housing stock, low employment levels and high EUI suggests a low level of energy efficiency in the housing stock.

High EUI Cluster. Cluster 3 has a high cluster median gas EUI at 87 kWh/m^2. It is characterised by low percentages of old housing, very low rates of single-person households, low levels of economic activity and mid-range of employment. It is the second most ethnically diverse cluster after Cluster 7. The large amounts of new housing and low range of economic inactivity suggest a cluster whose high EUI is driven by occupancy lifestyles rather than housing stock.

Medium Tier EUI Clusters. Cluster 1, the largest group in our analysis, exhibits the highest gas EUI among the mid-tier clusters, with a median of 84 kWh/m^2. This cluster is characterized by a wide distribution of housing typologies and construction ages, but a relatively tight distribution of ethnicity (approximately 70-85% white). Unlike other clusters with more distinct features, Cluster 1 lacks strong defining characteristics beyond being predominantly white areas, with a wide variety of housing stock and economic activity. This heterogeneity suggests that Cluster 1 might benefit from further analysis, such as sub-clustering or individual case reassignment, to reveal more nuanced patterns or subgroups within its population.

Cluster 6 is similar to Cluster 1 with a median gas EUI of 83 kWh/m^2. It emerges as the most economically active group, characterized by the highest employment rates and the lowest levels of economic inactivity. It has lower rates of single-person households and the lowest proportion of pre-1919 houses, suggesting newer or more developed urban areas.

Cluster 5 represents the lower end of the medium-tier, with a median gas EUI of 80 kWh/m^2. It is characterised by some of the highest levels of older houses pre-1919, along with a large amount of small terraces. It is one of the least economically active clusters, with high economic inactivity, and some of the lowest rates of employment.

Lower Tier EUI Clusters. Cluster 2 has the highest EUI of the low-tier clusters, with a median gas EUI of 76 kWh/m^2. It emerges the most ethnically homogeneous group, with the highest proportion

[6]Tables provided in Supplementary Information

Cluster	Members	% Standard houses	% Small Terraces	% Estates	% B.Age Pre-1919	% B.Age 1919-1999	% 1-person household	% White	% Asian	% Employed	% Econ. Inactive
0	20	57.9	21.4	3.4	13.0	67.6	34.2	87.3	6.2	48.8	30.7
1	21	42.7	33.4	6.4	19.1	64.0	34.0	79.7	9.7	45.6	33.1
2	15	53.8	25.8	2.3	19.6	62.4	38.7	90.3	4.2	46.7	34.1
3	9	48.4	27.6	5.1	10.7	70.1	30.2	57.5	25.8	45.9	29.0
4	10	43.8	34.9	2.5	21.8	59.3	41.1	91.5	4.2	42.0	38.1
5	11	38.6	39.1	3.1	27.0	56.9	37.2	74.1	15.3	40.4	40.9
6	14	53.3	26.0	6.5	7.8	75.9	31.0	77.9	12.8	50.6	27.7
7	11	41.5	40.2	3.4	24.7	59.3	36.0	52.7	32.7	39.8	34.9

Table 5: Clusters, members and mean values of attributes. B.Age = building age. For building age, we can infer % of Post 1919 buildings as the three columns are collectively exhaustive. Note that the census variable for economic activity included a 'None applicable' option which averaged around 18% of responses. The distributions of the variables across the clusters are given in the ECDF plots in the supplementary information.

of white households. This cluster is characterized by an above-average number of standard homes and high rates of single-person households. The economic activity in this cluster falls within the mid-range, suggesting a mix of working and non-working populations. The low EUI even with the higher proportion of standard homes suggest areas at risk for fuel poverty.

Cluster 0 is the middle of the low-tier cluster, with a median gas EUI of 75 kWh/m^2. This cluster is notable for having the largest proportion of standard homes, with a fairly uniform distribution across the cluster. It contains relatively small amounts of small terraced houses. Importantly, Cluster 0 boasts the second-highest employment levels after Cluster 6. These characteristics suggest urban areas with good energy efficiency measures in place, likely due to newer or well-maintained housing stock and an economically active population.

Cluster 4 has the lowest median gas EUI (72 kWh/m^2). It is distinguished as very ethnically white and exhibits high levels of economic inactivity and low levels of employment. This cluster appears relatively uniform in terms of distribution ranges of ethnicity and employment, suggesting homogeneous, possibly ageing communities. The low EUI combined with the high proportion of single-person households and high economic inactivity suggests high risks of fuel poverty.

To conclude the cluster investigations, we note a good level of discernment in gas EUI and some meaningful actions based on this analysis. Cluster 4 is identified as containing areas potentially at high risk for fuel poverty, warranting particular attention. Cluster 7 shows high potential for housing energy efficiency measures, suggesting opportunities for meaningful emissions reduction. Cluster 3 presents as a suitable candidate for targeting lifestyle-based reduction methods, which could provide emissions reduction without high infrastructure spend. Cluster 1 lacks clear distinguishing characteristics relative to other clusters, potentially indicating a need for additional variables to improve clustering resolution. It is important to note that these cluster assessments provide a broad overview, and intra-urban variability means that areas of socioeconomic hardship may exist even within clusters that generally perform well. Future work should look to investigate these clusters within-cluster distributions in more detail. A short example of how we can investigate within-cluster using the current framework now follows.

3.2 Case Study

The following case study details how the methodology can be used to make targeted and granular assessments of small neighbourhood energy performance. For cluster 6, the EUI performance can be split out by postcode typology, using the categories defined in Section 2.3, to assess what drives the higher and lower ends of the cluster performance. The results of this split are given in Figure 6.

We note that the high gas EUI performance of High Wycombe is driven by postcodes containing semi-detached and detached houses rather than flats and terraces, despite the tendency of our EUI to reduce the values for such buildings. The distribution of such domestic typologies across the urban area of High Wycombe is given in Figure 7. By combining the cluster energy benchmark for these typologies, and mapping them spatially, we can provide local decision-makers with the tools needed to make data-driven energy performance interventions, customised for the local area.

3.3 Comparison to CIBSE Benchmarks

We compare our results to energy benchmarks published in the literature. The box plots for our results of postcode gas and electricity EUI for postcodes with uniform typology are given in Figure 8. The values for the CIBSE benchmarks are extracted from [19].

The EUI of postcodes is consistently lower than the CIBSE benchmarks for all three typology benchmarks, with the largest differences observed for gas EUI in semi+detached postcodes. There are three potential contributing to this disparity. Firstly, total domestic energy consumption was 17% lower in 2022 than in 2016 in absolute terms, 18.5% lower after weather adjustments [14]. Secondly, our method does not include information on occupancy rates, whereas the CIBSE benchmarks only collected energy data from occupied houses. The 2021 census puts the vacancy rate of houses in England at 5.4% and of Wales at 7.0% [36]. Finally, when we compare London's performance to the rest of the regions within our dataset, London's median gas EUI is 13% higher than the average (90.9 vs 80.4 kWh/m^2). We test the statistical significance of this difference, again using a Mann-Whitney U Test necessitated by non-normal gas EUI. Comparing London against the aggregated data from all

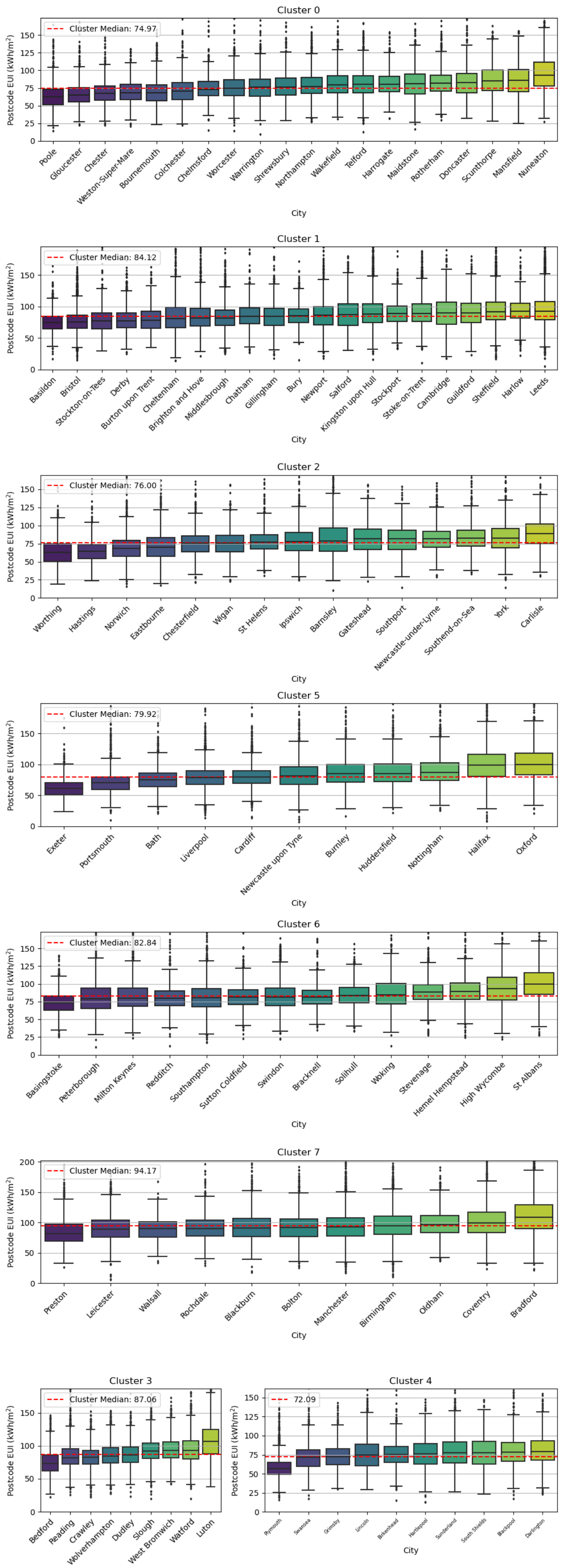

Figure 5: EUI benchmarks for clusters of urban areas

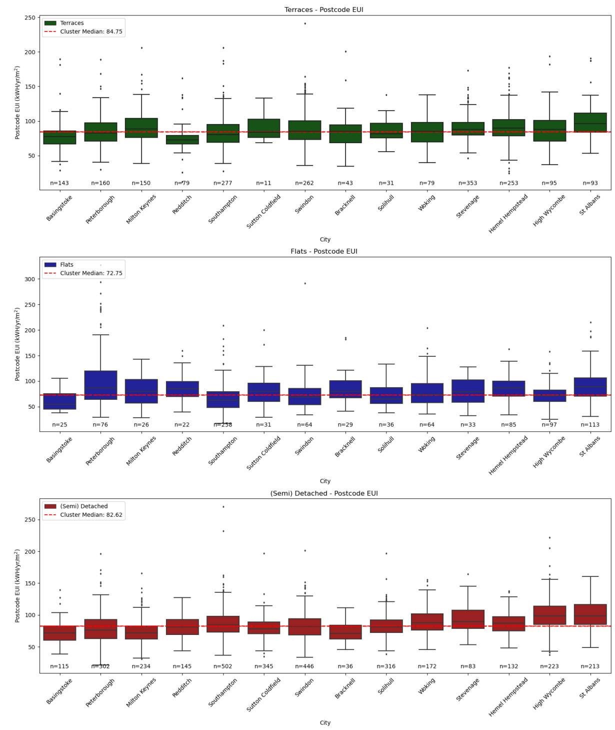

Figure 6: Cluster 6 postcode EUI performance, split by postcode typology

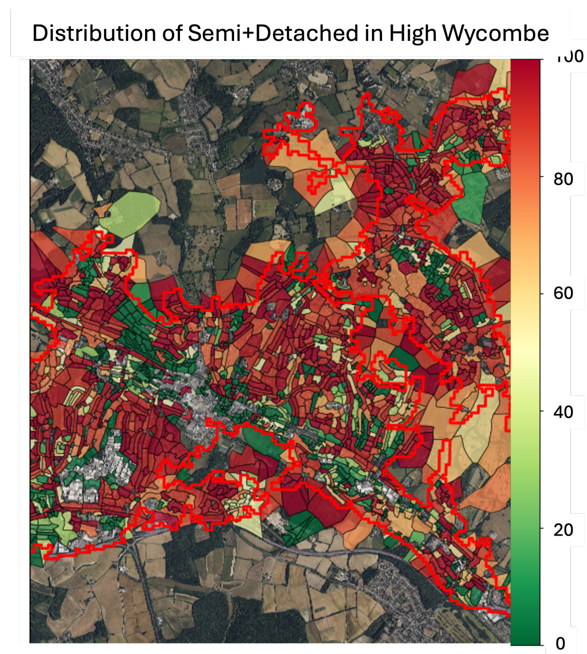

Figure 7: Distribution of domestic neighbourhoods with proportions of semi-detached + detached houses. Contains OS data © Crown copyright and database right 2024 [37]

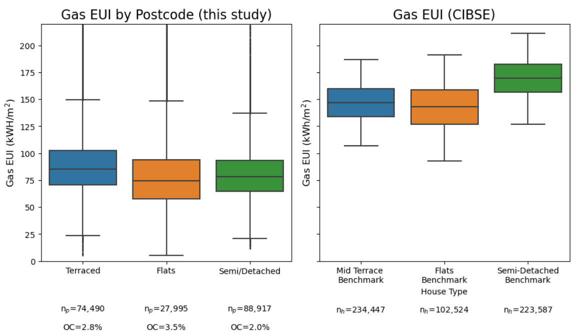

Figure 8: (L) Postcode EUI. n$_p$: number of postcodes in sample. OC: outlier percentage. The Y-axis is capped at the maximum value for 99th quartile across the three typology groups (R) CIBSE benchmarks for mid-terraced, flats and semi-detached houses. n$_h$: number of houses in sample [19].

.

other regions yielded a statistically significant result with a medium effect size ($p < 0.0001$, effect size 0.27). It is important to note that while both our method and CIBSE benchmarks exclude mixed-use buildings, our filter is more restrictive due to higher granularity. This exclusion likely creates an over-index for London due to its urban nature.

After reducing the CIBSE medians to account for the 37% in demand shift, occupancy rates, and London's higher energy performance, we find that our Terraced, Flats, and Semi-detached/Detached medians are 87%, 78%, and 69% of the CIBSE medians, respectively. Testing our distributions against the benchmark median using a one-sided Wilcoxon signed-rank test, we find statistically significant differences, with small, medium and large effect sizes for Terraces, Flats and Semi+Detached houses respectively (p<0.0001, effect sizes: 0.14, 0.33, 0.74). Hence our methods show the closest alignment for terraced houses and diverge more for semi+detached buildings. Note that without the full distribution of CIBSE observations, we can only run a one-sided comparison to the median.

The remaining differences may be caused by a combination of varying occupancy behaviour across typologies and methodological differences. It might be that more second homes and unoccupied buildings fall into the semi+detached categories. There might be overestimation of building volume in our method, potentially more so for semi+detached homes. We note there were instances within our methodology of building floor area being over or underestimated, due to errors in the building stock data and challenges with distinguishing between lived-in areas and ancillary buildings such as basements, garages and garden sheds. The individual assessments of house area/volume in the CIBSE database can more easily correct for these cases. We are also not able to distinguish confidence bounds for our data: the data provider Verisk confirmed varying sources within individual building variables, but the data lacks metadata on either method of attribution (whether hand-labelled or model-derived) and timing of variable updates. While these limitations do not invalidate our findings, they underscore the importance of cautious interpretation and highlight the need for transparency in building stock databases. These limitations should be kept in mind when using our benchmarks, which should be used in the context of their derivation as postcode-level EUI.

To conclude, the differences between our benchmarks and those provided by CIBSE can be attributed to several factors, including temporal shifts in domestic energy usage patterns from 2016 to 2022, the impacts of housing occupancy on the postcode, the unique energy performance of London, occupancy behaviour and a potential overestimation of building volume in our methodology. While our methodology may have some limitations, it still offers significant advantages over existing benchmarks. Our approach utilizes recent energy consumption data from across England and Wales, providing a more representative and current picture of energy use. The method is fast and cheap to implement as it doesn't require extensive matching between meters and buildings. There is no interaction at the homeowner level, preserving privacy and enabling open-source adoption. Leveraging a wide range of observational data reduces misinterpretation of energy data, better captures regional differences, and scales easily to national levels.

4 Limitations and Extension

Our study faces several methodological limitations, both in terms of data generation and high-level methodology, as well as in the specific implementation of clustering techniques. These limitations suggest areas for future extension.

Clustering at the city level produces easily digestible groups and helps target macro policy efforts. However, this approach may fail to capture distinct sub-city clusters and miss valuable connections between sub-city regions. The relatively low number of entries for clustering also means the clustering method is sensitive to new members, such that introducing new regions such as Scotland could change the clusters. Future work could look to implement a two-stage clustering approach at the postcode level, followed by the local authority level, and compare cluster results.

Our implementation of Spectral Clustering demonstrated superior performance compared to Gaussian Mixture Models and generally yielded well-defined clusters. However, one cluster remained relatively poorly defined, suggesting potential limitations in our approach. This anomaly could indicate the presence of missing explanatory variables or the existence of a more complex sub-cluster structure that our current method fails to capture. Further investigation into this cluster may reveal more useful insights into those areas' energy performance and could guide refinements to our clustering methodology.

The quality of the clustering also relies on the quality of the data underpinning it. Our study faces limitations in data representation, particularly regarding EUI at the postcode level. While our city clustering incorporates both domestic and mixed-use postcodes, the EUI benchmark analysis is restricted to domestic postcodes due to the unavailability of non-domestic consumption data at this granularity. This limitation (including 55% of postcodes with domestic gas data available) potentially skews our analysis, under-representing urban centres and overemphasizing suburban areas. The exclusion of mixed-use postcodes presents a significant data gap, especially for metropolitan cities with a higher proportion of such buildings. By focusing on defined towns and cities, we also exclude smaller villages and towns which may have distinct energy performance as well. Future research should prioritize integrating mixed-use postcode data to provide a complete representation of urban EUI patterns and extend to cover all housing regions of the UK. We are also limited in that the energy data excludes any postcodes with less than five meters, so there is some additional skew away from very remote areas.

Beyond the issues of data completeness, we also face challenges related to the quality and reliability of the data we have. Our methodology solves some of the identified errors within the building stock data, but we do not have confidence bounds on derived attributes such as building floor area given the data limitations relating to the variable sources. In particular, we note that the variability of values for some subset of the data is high, e.g. for flats. Further work should look to improve validations on the dataset and refine methodology in areas with high variability. We note that EUI at the postcode level is often easy to manually verify input correctness of with satellite data. An interesting extension could be developing automated quality control methods, potentially using vision-language models, or testing other input-building stock data sources.

5 Conclusions

In conclusion, we have developed a novel methodology for creating customized energy benchmarks based on actual consumption data for 111 towns and cities in England and Wales. Our approach yields eight clusters based on city-level attributes known to correlate with energy consumption, along with corresponding gas EUI benchmarks for each cluster. Analysis of the cluster characteristics and median EUI suggested several clusters as candidates for either energy efficiency housing upgrades, behavioural interventions or interventions around fuel poverty. We illustrate the utility of this method through a case study of High Wycombe, demonstrating how neighbourhoods with semi-detached and detached houses drive the cities' high energy performance relative to their cluster group. Our methodology leverages open-access energy data and geo-located neighbourhood units, offering significant potential for incorporating additional variables. The approach is efficient, flexible, and scalable, addressing privacy concerns and closed-access data limitations by utilizing open-access small neighbourhood data and can be easily updated annually with new energy data, maintaining its relevance.

Acknowledgments

This work was supported by the UKRI Centre for Doctoral Training in Application of Artificial Intelligence to the study of Environmental Risks (reference EP/S022961/1)

References

[1] Hassan Adan and Franz Fuerst. 2016. Do energy efficiency measures really reduce household energy consumption? A difference-in-difference analysis. *Energy Efficiency* 9, 5 (Oct. 2016), 1207–1219. https://doi.org/10.1007/s12053-015-9418-3
[2] Abdo Abdullah Ahmed Gassar, Geun Young Yun, and Sumin Kim. 2019. Data-driven approach to prediction of residential energy consumption at urban scales in London. *Energy* 187 (Nov. 2019), 115973. https://doi.org/10.1016/j.energy.2019.115973
[3] Jamal Alabid, Amar Bennadji, and Mohammed Seddiki. 2022. A review on the energy retrofit policies and improvements of the UK existing buildings, challenges and benefits. *Renewable and Sustainable Energy Reviews* 159 (May 2022), 112161. https://doi.org/10.1016/j.rser.2022.112161
[4] Usman Ali, Mohammad Haris Shamsi, Cathal Hoare, Eleni Mangina, and James O'Donnell. 2021. Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis. *Energy and Buildings* 246 (Sept. 2021), 111073. https://doi.org/10.1016/j.enbuild.2021.111073
[5] Hui Ben and Koen Steemers. 2019. Assessing the impact of a differentiated retrofit approach in UK domestic buildings. *Journal of Physics: Conference Series* 1343, 1 (Nov. 2019), 012173. https://doi.org/10.1088/1742-6596/1343/1/012173 Publisher: IOP Publishing.
[6] Esfandiar Burman, Dan Rigamonti, Judit Kimpain, and Dejan Mumovic. 2012. Performance Gap and Thermal Modelling: A Comparison of Simulation Results and Actual Energy Performance for an Academy in North West England. Loughborough, UK.
[7] Climate Change Commitee. 2024. Progress in reducing emissions: 2024 Report to Parliament.
[8] David L. Davies and Donald W. Bouldin. 1979. A Cluster Separation Measure | IEEE Journals & Magazine | IEEE Xplore. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1, 2 (April 1979), 224 – 227. https://doi.org/10.1109/TPAMI.1979.4766909
[9] Chirag Deb and Siew Eang Lee. 2018. Determining key variables influencing energy consumption in office buildings through cluster analysis of pre- and post-retrofit building data. *Energy and Buildings* 159 (Jan. 2018), 228–245. https://doi.org/10.1016/j.enbuild.2017.11.007
[10] C. Deb and A. Schlueter. 2021. Review of data-driven energy modelling techniques for building retrofit. *Renewable and Sustainable Energy Reviews* 144 (July 2021), 110990. https://doi.org/10.1016/j.rser.2021.110990
[11] Department for Energy Security and Net Zero. 2015. Lower and Middle Super Output Areas gas consumption - GOV.UK. https://www.gov.uk/government/statistics/lower-and-middle-super-output-areas-gas-consumption
[12] Department for Energy Security and Net Zero. 2024. 2023 UK greenhouse gas emissions, provisional figures. https://assets.publishing.service.gov.uk/media/6604460f91a320001a82b0fd/uk-greenhouse-gas-emissions-provisional-figures-statistical-release-2023.pdf
[13] Department for Energy Security and Net Zero. 2024. Postcode level domestic gas and electricity consumption: about the data. https://www.gov.uk/government/publications/postcode-level-domestic-gas-and-electricity-consumption-about-the-data
[14] Department for Energy Security and Net Zero. 2024. *UK ENERGY IN BRIEF 2023, UK ENERGY IN BRIEF 2017.* Technical Report. https://www.gov.uk/government/collections/uk-energy-in-brief
[15] Thomas R. Dougherty, Tianyuan Huang, Yirong Chen, Rishee K. Jain, and Ram Rajagopal. 2021. SCHMEAR: scalable construction of holistic models for energy analysis from rooftops. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. ACM, Coimbra Portugal, 111–120. https://doi.org/10.1145/3486611.3486666
[16] Kailun Feng, Weizhuo Lu, Yaowu Wang, and Qingpeng Man. 2022. Energy-Efficient Retrofitting under Incomplete Information: A Data-Driven Approach and Empirical Study of Sweden. *Buildings* 12, 8 (Aug. 2022), 1244. https://doi.org/10.3390/buildings12081244 Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.
[17] Timothy J. Foxon. 2013. Transition pathways for a UK low carbon electricity future. *Energy Policy* 52 (Jan. 2013), 10–24. https://doi.org/10.1016/j.enpol.2012.04.001
[18] Abdo Abdullah Ahmed Gassar and Seung Hyun Cha. 2020. Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review. *Energy and Buildings* 224 (Oct. 2020), 110238. https://doi.org/10.1016/j.enbuild.2020.110238
[19] Daniel Godoy-Shimizu, Rob Liddiard, Stephen Evans, Sung Min Hong, Dominic Humphrey, Paul Ruyssevelt, Dejan Mumovic, and Philip Steadman. 2024. Producing domestic energy benchmarks using a large disaggregate stock model. *Building Services Engineering Research and Technology* 45, 3 (May 2024), 217–239. https://doi.org/10.1177/01436244241229907 Publisher: SAGE Publications Ltd STM.
[20] Solène Goy, Volker Coors, and Donal Finn. 2021. Grouping techniques for building stock analysis: A comparative case study. *Energy and Buildings* 236 (April 2021), 110754. https://doi.org/10.1016/j.enbuild.2021.110754
[21] Kirstem Gram-Hanssen. 2004. Chapter 7 Domestic electricity consumption - consumers and appliances. In *The Ecological Economics of Consumption*. Edward Elgar Publishing. https://china.elgaronline.com/edcollchap/9781843765127.00015.xml Section: The Ecological Economics of Consumption.
[22] Benedetto Grillone, Gerard Mor, Stoyan Danov, Jordi Cipriano, and Andreas Sumper. 2021. A data-driven methodology for enhanced measurement and verification of energy efficiency savings in commercial buildings. *Applied Energy* 301 (Nov. 2021), 117502. https://doi.org/10.1016/j.apenergy.2021.117502
[23] SeyedehRabeeh HosseiniHaghighi. 2021. Characterizing and Structuring Urban Data for Housing Stock Energy Modeling. (2021).
[24] Xiaoyu Jin, Chong Zhang, Fu Xiao, Ao Li, and Clayton Miller. 2023. A review and reflection on open datasets of city-level building energy use and their applications. *Energy and Buildings* 285 (April 2023), 112911. https://doi.org/10.1016/j.enbuild.2023.112911

[25] John Hedengren. [n. d.]. Machine Learning for Engineers: Spectral Clustering. https://apmonitor.com/pds/index.php/Main/SpectralClustering

[26] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic. 2010. A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment* 45, 7 (July 2010), 1683–1697. https://doi.org/10.1016/j.buildenv.2010.01.021

[27] Rob Liddiard, Daniel Godoy-Shimizu, Paul Ruyssevelt, Philip Steadman, Stephen Evans, Dominic Humphrey, and Rayan Azhari. 2021. Energy use intensities in London houses. *Buildings and Cities* 2, 1 (April 2021), 336. https://doi.org/10.5334/bc.79

[28] Dr Mathew Lipson. 2021. Plotting OpenStreetMap images with Cartopy. https://www.theurbanist.com.au/2021/03/plotting-openstreetmap-images-with-cartopy/

[29] Guillermo Macbeth, Eugenia Razumiejczyk, and Rubén Daniel Ledesma. 2011. Cliff's Delta Calculator: A non-parametric effect size program for two groups of observations. *Universitas Psychologica* 10, 2 (May 2011), 545–555. http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S1657-92672011000200018&lng=en&nrm=iso&tlng=en Publisher: Pontificia Universidad Javeriana.

[30] Alessio Mastrucci, Emil Popovici, Antonino Marvuglia, Luís De Sousa, Enrico Benetto, and Ulrich Leopold. 2015. GIS-based Life Cycle Assessment of urban building stocks retrofitting- a bottom-up framework applied to Luxembourg. Atlantis Press, 47–56. https://doi.org/10.2991/ict4s-env-15.2015.6 ISSN: 2352-538X.

[31] É. Mata, A. Sasic Kalagasidis, and F. Johnsson. 2014. Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. *Building and Environment* 81 (Nov. 2014), 270–282. https://doi.org/10.1016/j.buildenv.2014.06.013

[32] Janine Morley. 2011. The significance of difference: Understanding variation in household energy consumption. In *Energy efficiency first: The foundation of a low-carbon society.* Belambra Presqu'île de Giens, France.

[33] Jonathan Morris, David Allinson, John Harrison, and Kevin J. Lomas. 2016. Benchmarking and tracking domestic gas and electricity consumption at the local authority level. *Energy Efficiency* 9, 3 (June 2016), 723–743. https://doi.org/10.1007/s12053-015-9393-8

[34] Neil Ernst. 2021. cliffs-delta: Python Package to calculate Cliff's delta. https://github.com/neilernst/cliffsDelta

[35] Office for National Statistics. [n. d.]. Statistical geographies. https://www.ons.gov.uk/methodology/geography/ukgeographies/statisticalgeographies

[36] Office for National Statistics. 2023. *Number of vacant and second homes, England and Wales, statistical bulletin.* Technical Report. https://www.ons.gov.uk/peoplepopulationandcommunity/housing/bulletins/numberofvacantandsecondhomesenglandandwales/census2021

[37] Ordnance Survey. [n. d.]. Code-Point with Polygons | Data Products | © Crown copyright and database rights 2024 Ordnance Survey (AC0000851941). https://www.ordnancesurvey.co.uk/products/code-point-polygons

[38] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 85 (2011), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

[39] Reynolds, Douglas A. 2009. Gaussian mixture models. *Encyclopedia of biometrics* 741 (2009), 659–663. http://leap.ee.iisc.ac.in/sriram/teaching/MLSP_16/refs/GMM_Tutorial_Reynolds.pdf

[40] Mehdi Shahrestani, Runming Yao, and Geoffrey K. Cook. 2014. A review of existing building benchmarks and the development of a set of reference office buildings for England and Wales. *Intelligent Buildings International* 6, 1 (Jan. 2014), 41–64. https://doi.org/10.1080/17508975.2013.828586 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17508975.2013.828586.

[41] Jianbo Shi and J. Malik. 2000. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 8 (Aug. 2000), 888–905. https://doi.org/10.1109/34.868688 Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[42] A. J. Summerfield, T. Oreszczyn, J. Palmer, I. G. Hamilton, F. G. N. Li, J. Crawley, and R. J. Lowe. 2019. What do empirical findings reveal about modelled energy demand and energy ratings? Comparisons of gas consumption across the English residential sector. *Energy Policy* 129 (June 2019), 997–1007. https://doi.org/10.1016/j.enpol.2019.02.033

[43] Paraskevi Vatougiou, Peter McCallum, and David P Jenkins. 2020. Assessing Residential Retrofit Measures to Reduce the Risk of Fuel Poverty: A Modelling Approach at Community Scale. 8.

[44] Evelien Verellen and Karen Allacker. 2022. Developing a Building Stock Model to Enable Clustered Renovation—The City of Leuven as Case Study. *Sustainability* 14, 10 (Jan. 2022), 5769. https://doi.org/10.3390/su14105769 Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

[45] Ulrike von Luxburg. 2007. A tutorial on spectral clustering. *Statistics and Computing* 17, 4 (Dec. 2007), 395–416. https://doi.org/10.1007/s11222-007-9033-z

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

[47] Xinran Yu, Zhengbo Zou, and Semiha Ergan. 2023. Extracting principal building variables from automatically collected urban scale façade images for energy conservation through deep transfer learning. *Applied Energy* 344 (Aug. 2023), 121228. https://doi.org/10.1016/j.apenergy.2023.121228

[48] Mingda Yuan and Ruchi Choudhary. 2020. A two-step clustering framework for locally tailored design of residential heating policies. *Sustainable Cities and Society* 63 (Dec. 2020), 102431. https://doi.org/10.1016/j.scs.2020.102431

A Clustering

A.0.1 Spectral Clustering. A brief summary of spectral clustering is taken from John Hedengren [25]. Spectral clustering involves several steps:

Constructing a similarity matrix using a kernel function (e.g., Gaussian kernel) to measure pairwise data point similarities. Computing a degree matrix and Laplacian matrix from the similarity matrix. Calculating eigenvectors of the Laplacian matrix to create a spectral embedding of the data. Applying K-means clustering to this lower-dimensional spectral representation.

This approach has roots in graph theory and is also known as segmentation-based object categorization. Its main advantage is flexibility in handling complex data structures where traditional clustering methods may fail. However, spectral clustering can be computationally expensive for large datasets, especially if the similarity matrix is dense .

B Building Level Pre-Processing

We implement a custom building-level pre-processing pipeline to derive the floor count at building level.

- Bucket building age variables to ensure non-overlapping categories
- Bucket building heights into discrete subgroups.
- Apply outbuilding update: for buildings with height 3m, floor count 2 and no UPRN, convert into outbuildings.
- Apply average floor count updates:
  - Identify any building with average floor height outside of 2.2-5.3m as having an error with either height or floor count
  - Derive the MWB: minimum width of the building footprint and calculate the ratio of height and MWB: $R_H = \frac{H}{MWB}$.
  - If the average floor height is outside of floor height thresholds, and $H > 3 * MWB$, set the height to null.
  - If average floor height is outside of thresholds and $H \leq 3 * MWB$, set floor count to null.
- Fill missing values for height and floor count (this includes both missing data and the incorrect heights/floor counts that were set to null in the prior step). The missing values are filled using global averages.
  - Global Fill: use the global average tables to update any missing height or floor count with the global average. Global averages are derived from the whole dataset.
- Calculate total floor area per building from height: height → global floor count × premise area.