Energy Proceedings
Vol 64, 2026
ISSN 2004-2965

Clustering England's and Wales's Housing Stock: From EPC Data to Energy Poverty Insights [#]

Aaron Tesfa Tsion[1], Wei He [1*]
1 King's College London, United Kingdom
(Corresponding Author: wei.4.he@kcl.ac.uk)

ABSTRACT

This paper presents a novel data-driven framework that integrates the analysis of household energy poverty vulnerability with the assessment of technological pathways for the net-zero transition, creating an integrated basis for equitable transition planning. Using the Energy Performance Certificates (EPCs) for England and Wales, we apply K-Prototypes clustering at the LSOA scale to jointly analyse numerical and categorical housing features. Clusters are ranked by a composite "energy poverty transition" score, capturing efficiency, demand, emissions, heating costs, and fuel type. The results reveal pronounced spatial variation in energy poverty transition risk across regions and settlement types, highlighting localised concentrations of vulnerability embedded within the housing stock. The method provides local authorities with a scalable and transparent tool for identifying priority areas for targeted intervention, supporting more equitable and context-sensitive net-zero strategies, while underscoring the need for complementary datasets to inform delivery strategies.

Keywords: energy transition, energy poverty, renewable energy, household energy consumption, sustainable technology, energy data clustering

NONMENCLATURE

Abbreviations	
EPC	Energy Performance Certificate
LSOA	Lower Super Output Area
EPT	Energy Poverty Transition

1. INTRODUCTION

Energy poverty remains a critical socio-technical challenge within the global energy transition, directly affecting household wellbeing and the pace of decarbonisation. The techno-economic framework of energy poverty highlights the interlinked dimensions of affordability, availability, and cleanliness [1]. Households struggling to meet their basic energy needs for heating, cooling, or cooking experience constrained wellbeing and heightened vulnerability [2]. Indicators of affordability often assess energy expenditure as a share of household income [3], while availability captures infrastructure deficits and dependence on insecure or traditional fuels [4]. The cleanliness dimension underscores the health and sustainability impacts of fuels, with reliance on biomass or coal linked to indoor air pollution and severe health outcomes [3]. Prolonged exposure to such conditions forces households into trade-offs with essentials like food and healthcare [5]. Addressing energy poverty does not only alleviate vulnerability but also support climate objectives, as decarbonised systems enhance household resilience and lower greenhouse gas emissions [1] [4].

Despite decades of alleviation policies, energy poverty measurement and mitigation remain deeply contested. The UK's shifting definitions of fuel poverty, from the 10% rule, to the Low Income High Cost (LIHC) approach, and most recently to the Low-Income Low Energy Efficiency (LILEE) indicator, reflect ongoing methodological challenges. The latter incorporates efficiency ratings from Energy Performance Certificates (EPCs) alongside income thresholds, recognising the role of building efficiency in driving energy poverty risks [6]. However, it has been criticised for overlooking households in efficient dwellings who nevertheless face high energy burdens, such as those in council housing or with elevated health-related needs [7]. Such efficiency-centred indicators risk underestimating vulnerability and obscuring socio-demographic and spatial disparities.

This difficulty underscores two persistent barriers: (i) the lack of a standardised, multi-dimensional definition of energy poverty, (ii) the limitations of top-down metrics that fail to capture heterogeneity at the household and local scale. As a result, current policy indicators may misidentify those at risk and underestimate the distributional challenges of the net-zero transition. The complexity of structural transformation further compounds these gaps: while renewable electrification reduces long-term emissions, it

---

[#] This is a paper for the 17th International Conference on Applied Energy (ICAE2025), December 8-12, 2025, Bangkok, Thailand.

may temporarily exacerbate affordability and accessibility issues for vulnerable groups [8] [9].

To address this, a bottom-up, data-driven approach is required, one that leverages large-scale household-level EPC data to uncover hidden patterns of vulnerability and transition readiness. By applying clustering methods to EPC datasets, this study moves beyond aggregated indicators and reveals fine-grained, spatially explicit insights that connect household energy characteristics with sustainability. This study therefore proposes a clustering-based framework applied at the Lower Super Output Area (LSOA) level across England and Wales, that is novel in its dual integration of vulnerability and net-zero transition readiness, enabling localised assessments of energy poverty risk and alleviation strategies.

1.1 Objectives and Methodological Overview

Grounded in the EPC dataset for England and Wales, the work responds to the on-the-ground realities of the UK housing stock, where heterogeneity in age, fuel type, and property form creates challenges for one-size-fits-all solutions. Traditional approaches to energy poverty tend to treat households as isolated units and overly aggregate outcomes to a regional scale.

K-prototypes clustering is applied to EPC data, allowing simultaneous use of mixed data types that ensures that both technical and socio-structural housing characteristics are integrated into the clustering. By assigning each cluster a locally ranked energy poverty transition (EPT) score, the study combines metrics of household vulnerability within the net-zero transition needs. This dual metric moves beyond conventional energy poverty measures, offering a way to capture who is at risk and in need for mitigation or technology adoption.

Importantly, the analysis operates at the fine-grained LSOA scale, enabling spatially explicit comparisons across England and Wales. This ensures that recommendations remain grounded in local context and policy relevance, while also allowing for regional aggregation and cross-comparison. The combined EPT offers a scalable approach for identifying where household-level interventions can most effectively support a just and inclusive net-zero transition.

2. METHODS

2.1 Data Source

The analysis is based on the EPC dataset provided by the UK Department for Levelling Up, Housing and Communities. The EPC register is legally mandated for all domestic properties prior to sale or rental and currently contains more than 25 million records across England and Wales. This dataset, updated monthly, provides a rich description of the housing stock including efficiency ratings, fuel use, heating costs, and construction age. For this study, all available domestic EPCs from England and Wales were included.

2.2 Indicator Selection

The indicators selected for the analysis encompass both numerical and categorical descriptors of household energy performance. The energy efficiency rating is a composite measure that captures insulation quality, renewable technology uptake, and building service efficiency [6]. Energy consumption (kWh/m^2) functions as a proxy for household demand intensity, while CO_2 emissions (tCO_2/year) highlight the climate impact, transition and health burden associated with energy use [10]. Heating cost (£/year) directly reflects affordability constraints, a central dimension of energy poverty. The total floor area (m^2) provides insight into dwelling size. Complementing these numerical measures are categorical descriptors of property type, main fuel, and construction age band, which introduce structural variation crucial for identifying distinct archetypes of energy poverty risk and transition potential.

2.3 Pre-processing

Several data-cleaning and preparation steps were undertaken prior to clustering:

- Cleaning: Duplicate certificates were removed, construction age bands were harmonised to decade categories, and main fuel descriptors were standardised. Records with implausible floor areas (<6 m^2 or >1000 m^2) were dropped [11].

- Imputation: Missing values were filled through a hierarchical approach: 1. Median/mode within (property type, construction age band). 2. Median/mode within property type. 3. Median/mode within construction age band. 4. Global median/mode fallback.

- Outlier removal: For continuous variables, outliers were filtered using the interquartile range (IQR) method [11] [12] [13].

- Scaling: Numerical variables were min-max scaled to ensure proportional contributions across features with differing ranges.

2.4 Clustering Approach

The K-Prototypes algorithm was applied at the LSOA level, enabling clustering on mixed data types (numerical and categorical). For each LSOA, candidate solutions with *k* between 2 and 6 were tested, with the silhouette score used to select the optimal partition. This approach ensured local adaptability, as the housing mix varies considerably across regions.

2.5 Energy Poverty Transition Score and Ranking

Each cluster centroid was evaluated against an EPT scoring system, producing a relative rank (1: lowest EPT score = best performing cluster) within each LSOA. The scoring system integrated contributions from both numerical and categorical features [14] [15] [16]:
Energy efficiency: higher scores reduce the EPT score (better).
Energy consumption, CO_2 emissions, heating cost: higher values increase the EPT score (worse) [17].
Main fuel: penalties assigned by fuel type:

- Electricity: moderate penalty, except when emissions are low (possible heat pumps).
- Mains gas: small penalty, reflecting decarbonisation needs.
- Oil/LPG: strong penalty, high cost and carbon intensity.
- Solid fuels: strong penalty, reflecting pollution and inefficiency.
- Heat networks: no penalty, assuming central decarbonisation.
- Unknown/invalid fuels: maximum penalty.
- Construction age: stepwise penalty, with pre-1919 dwellings scoring worst and modern dwellings (post-2020) scoring best.
- Property type: flats/maisonettes assigned no penalty (lower demand envelope), houses small penalty, bungalows/park homes higher penalty.

Floor area was excluded from the scoring system, due to its implicit normalisation within the other energy indicators, and literature finds limited independent correlation with energy poverty. Each cluster was then assigned a rank relative to other clusters within the same LSOA, producing a local hierarchy of household archetypes from most to least transition needed.

3. RESULTS AND DISCUSSION

3.1 Clustering Outcomes Across LSOAs

The k-prototypes clustering was applied independently to each LSOA in England and Wales using energy poverty-relevant EPC dimensions. Table 1 shows the distribution of optimal k values across all LSOAs.

Table 1: Selection of k across all LSOAs.

k (clusters)	n_runs	sil_mean	sil_median	sil_std	sil_sem
2	35,738	0.2597	0.2100	0.1842	0.0010
3	35,707	0.1597	0.1339	0.1317	0.0007
4	35,698	0.1252	0.1074	0.1160	0.0006
5	35,692	0.1135	0.0984	0.1119	0.0006
6	35,683	0.1055	0.0904	0.1104	0.0006

Most LSOAs produced between two and four clusters, while larger and more diverse housing stocks required up to six clusters. The silhouette scores indicated that partitions had meaningful internal separation, supporting the robustness of the clustering approach.

3.2 Spatial Distribution of Energy Poverty Transition Score

Figure 1 illustrates which rank dominates in each LSOA, highlighting regional patterns in dominant cluster performance across England and Wales.

Regional differences reveal a modest but consistent spatial gradient that mirrors long-established disparities in housing stock quality and retrofit readiness. Wales, London, and the South of England (South East and South West) record the lowest ranks, indicating a higher prevalence of favourable cluster dominance, while the Midlands and Northern regions show slightly higher averages, reflecting greater concentrations of LSOAs with less optimal energy performance characteristics. These differences likely stem from structural factors: southern and metropolitan areas typically feature newer, higher-efficiency housing, whereas regions with older industrial legacies retain larger shares of pre-1970, poorly insulated dwellings. However, the similarity across all regions suggests that most LSOAs exhibit mid-range performance, implying that spatial energy inequality is driven by localised concentrations of vulnerability.

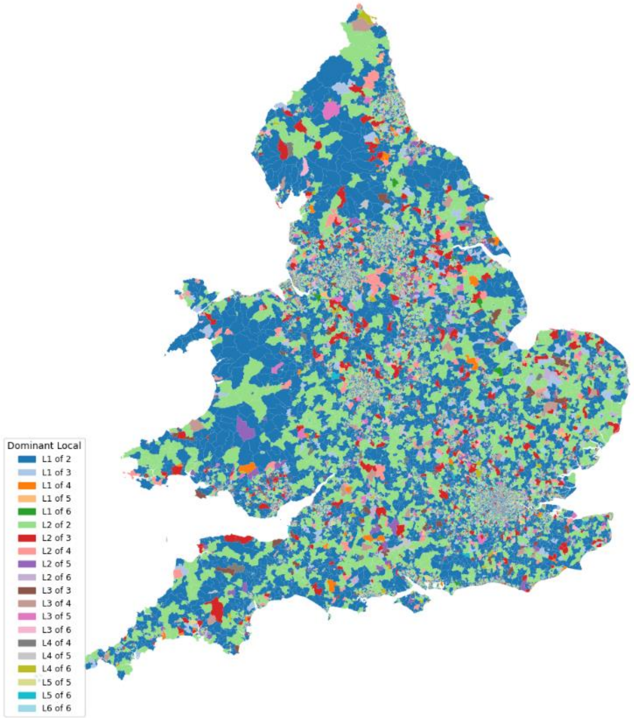

Figure 1: The most dominant energy poverty transition cluster label by cluster amount for each LSOA.

Table 2 complements this pattern by showing the weighted average EPT scores mapping each LSOA to a region according to governmental classification, providing a region-wide measure of energy performance vulnerability.

Table 2: Weighted average EPT scores across English regions and Wales.

Region	count	mean	std
Wales	1917	2.40	0.57
Yorkshire and The Humber	3355	2.33	0.43
West Midlands	3574	2.28	0.43
North West	4567	2.26	0.41
East Midlands	2847	2.25	0.44
North East	1682	2.24	0.42
South West	3407	2.23	0.54
East of England	3758	2.20	0.49
South East	5571	2.14	0.44
London	4994	2.05	0.40

Areas with the highest EPT scores, indicating poorer efficiency and greater retrofit need, are concentrated in Wales, the North of England, and parts of the West Midlands, aligning with regions of older, solid-wall or off-gas housing. In contrast, southern and eastern regions, particularly around London, the East of England, and the South East, show lower averages, reflecting a greater share of upgraded and better-performing dwellings.

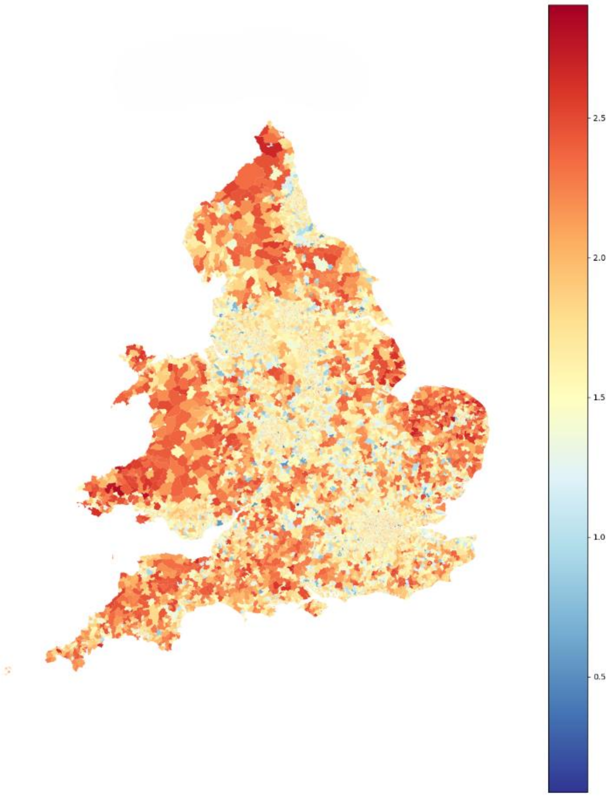

Figure 2: Average weighted EPT scores for each LSOA.

Figure 2 shows that within Wales, much of the South Wales Valleys and rural western areas display elevated scores, while conditions improve around Cardiff, Swansea, and the north-east coastal belt. This spatial alignment between dominant cluster rank and EPT performance underscores the persistence of a north-south divide and highlights the need for geographically targeted retrofit and electrification strategies to prevent deepening regional inequalities.

The weighted average EPT scores in Figure 3Figure 3 exhibit a clear gradient across the rural-urban classification, with the highest values observed in smaller rural settlements located further from major towns and cities, and progressively lower scores in areas with greater urban proximity and scale. Smaller rural areas, particularly those more remote, display markedly elevated EPT scores, indicating higher combined levels of energy poverty vulnerability and transition risk. In contrast, urban areas and larger settlements nearer to major centres show comparatively lower average scores, reflecting greater transition readiness and more favourable housing characteristics. This pattern suggests

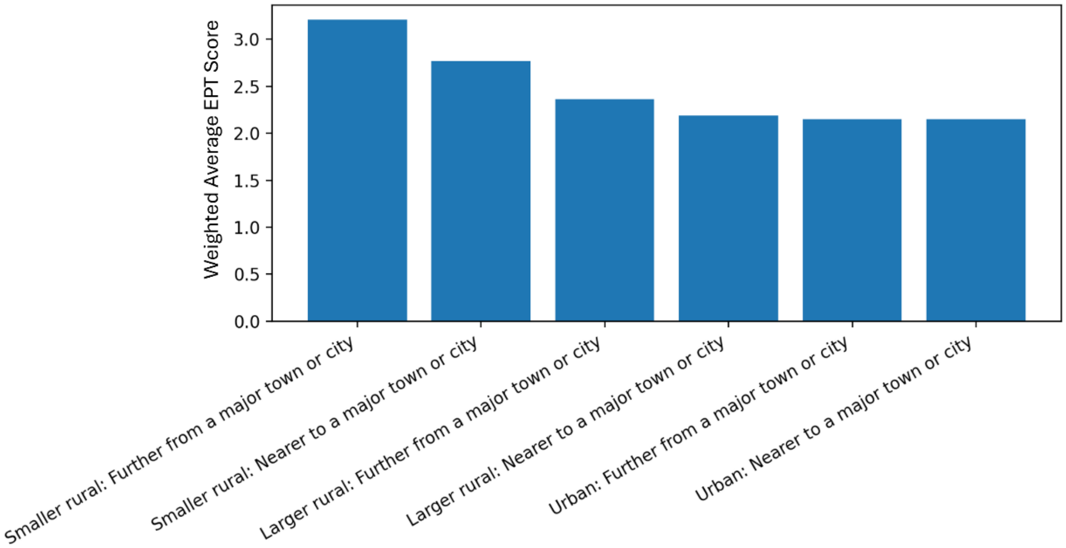

Figure 3: Average weighted EPT scores across UK governmental rural-urban classification.

that remoteness and settlement size play a significant role in shaping energy poverty transition risk, through their association with older housing stock, higher heat demand, and reliance on less flexible heating fuels.

3.3 Significance and Implications

The analysis delivers a key contribution: EPC-derived clustering demonstrates that adaptive scoring can reveal localised patterns of vulnerability and sustainability. By linking these clusters to energy transition readiness, the study moves beyond descriptive energy poverty research to provide operational value for net-zero prioritisation.

The geographic variation in housing conditions demands locally tailored sequencing and financing: as more remote and smaller rural areas tend to exhibit less favourable housing conditions than larger settlements and urban centres, which generally show comparatively higher transition readiness.

In energy-poor housing contexts, typically the emphasis is on the reduction of space-heating demand through fabric improvements as a foundational step, given its effectiveness in lowering energy costs, improving thermal comfort, and mitigating vulnerability irrespective of fuel type. Where heating systems are addressed, low-carbon alternatives, most prominently heat pumps, are generally considered only once minimum fabric performance thresholds are achieved, reflecting technical efficiency constraints and affordability considerations [18]. For households reliant on high-cost or carbon-intensive fuels, fuel switching is often discussed as a longer-term objective, contingent on infrastructure availability and dwelling suitability. Across these scenarios, the emphasis is placed on addressing structural inefficiencies first, which reduces exposure to volatile energy prices, and improves resilience to future energy system changes. [14] [19] [20]

The clustering and scoring framework developed here provides a systematic means of identifying where households are structurally constrained by poor fabric performance, high energy demand, or unfavourable fuel types. Factors that fundamentally shape the suitability and sequencing of any future intervention. By reframing technology pathways as outcomes contingent on underlying housing characteristics rather than fixed solutions, the analysis supports more nuanced, evidence-led targeting of energy poverty alleviation and transition planning. This highlights the real-world feasibility, underscoring the need for integrated policy action on identifying vulnerable household to enable equitable, scalable implementation of net-zero housing transitions.

4. CONCLUSION

This study demonstrates the feasibility and value of applying k-prototypes clustering to the national EPC dataset to uncover patterns of energy poverty risk. By ranking clusters within each LSOA according to a composite EPT score, the method provides actionable insights for local councils, making it easier to design and target policy interventions.

Beyond its practical utility, the work contributes a conceptual advance by integrating energy poverty vulnerability and transition readiness within a single operational model, offering a template for data-driven just transition planning.

The framework enables a straightforward application across LSOAs, allowing groups of households to be addressed collectively, which creates positive spillover effects for planning and delivery.

At the same time, the analysis highlights limitations: while the approach captures core dimensions of energy poverty and net-zero alignment, more complementary sociodemographic and behavioural data are needed to refine the specificity and give technological recommendations. Nevertheless, the simplicity of the method is also its strength, offering a scalable and transparent tool for councils and policymakers. Future work should validate the approach against real-world outcomes and extend the database to strengthen its robustness and policy relevance.

ACKNOWLEDGEMENT

The author is deeply thankful to Dr Wei He for his essential guidance, encouragement, and support during the course of this research. Appreciation is also extended to King's College London for the resources and institutional assistance that made this work possible.

REFERENCE

[1] F. Rao, Y. M. Tang, K. Y. Chau, W. Iqbal, and M. Abbas, 'Assessment of energy poverty and key influencing factors in N11 countries', *Sustain. Prod. Consum.*, vol. 30, pp. 1–15, Mar. 2022, doi: 10.1016/j.spc.2021.11.002.

[2] M. González-Eguino, 'Energy poverty: An overview', *Renew. Sustain. Energy Rev.*, vol. 47, pp. 377–385, Jul. 2015, doi: 10.1016/j.rser.2015.03.013.

[3] P. K. Adom, F. Amuakwa-Mensah, M. P. Agradi, and A. Nsabimana, 'Energy poverty, development outcomes, and transition to green energy', *Renew. Energy*, vol. 178, pp. 1337–1352, Nov. 2021, doi: 10.1016/j.renene.2021.06.120.

[4] J. Zhao, Q. Jiang, X. Dong, and K. Dong, 'Assessing energy poverty and its effect on CO2 emissions: The case of China', *Energy Econ.*, vol. 97, p. 105191, May 2021, doi: 10.1016/j.eneco.2021.105191.

[5] J. M. Fry, L. Farrell, and J. B. Temple, 'Energy poverty and food insecurity: Is there an energy or food trade-off among low-income Australians?', *Energy Econ.*, vol. 123, p. 106731, Jul. 2023, doi: 10.1016/j.eneco.2023.106731.

[6] Office for National Statistics, 'How fuel poverty is measured in the UK'. Accessed: Mar. 08, 2024. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/howfuelpovertyismeasuredintheuk/march2023

[7] N. Simcock, H. Thomson, S. Petrova, and S. Bouzarovski, Eds, *Energy poverty and vulnerability: a global perspective*, First issued in paperback. London: Routledge, 2019.

[8] 'Global Energy Crisis', IEA. Accessed: Apr. 02, 2024. [Online]. Available: https://www.iea.org/topics/global-energy-crisis

[9] P. Bolton, 'Gas and electricity prices during the "energy crisis" and beyond', *House Commons Libr.*, vol. Number 9714, Apr. 2024, Accessed: Apr. 08, 2024. [Online]. Available: https://commonslibrary.parliament.uk/research-briefings/cbp-9714/

[10] J. Chen *et al.*, 'Transforming building retrofits: Linking energy, equity, and health insights from The World Avatar', Jun. 2025, doi: 10.17863/CAM.119233.

[11] Matthew Wills, Jonathan Radcliffe, Neha Mehta, 'Use of domestic energy performance certificates and spatial mapping to evaluate the variation in heat demand and associated emissions reduction', doi: https://doi.org/10.1016/j.nexres.2025.100922.

[12] H. Liu, J. Li, Y. Wu, and Y. Fu, 'Clustering With Outlier Removal', *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2369–2379, Jun. 2021, doi: 10.1109/TKDE.2019.2954317.

[13] M. Anastasiadou, V. Santos, and M. S. Dias, 'Evaluating Energy Performance Certificate Data with Data Science', in *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, Dec. 2021, pp. 1–5. doi: 10.1109/ICECET52533.2021.9698806.

[14] Department for Energy Security & Net Zero, 'National Energy Efficiency Data-Framework (NEED): Summary of Analysis, Great Britain, 2024', Jun. 2024. Accessed: Sep. 20, 2025. [Online]. Available: https://assets.publishing.service.gov.uk/media/66e0203ad65d5c23df086710/NEED-report-june-2024.pdf

[15] S. Bouzarovski and S. Petrova, 'A global perspective on domestic energy deprivation: Overcoming the energy poverty–fuel poverty binary', *Energy Res.*

Soc. Sci., vol. 10, pp. 31–40, Nov. 2015, doi: 10.1016/j.erss.2015.06.007.

[16] H. Tutar and D. Štreimikienė, ‘Assessing the effects of low-carbon transition on energy poverty: Empirical evidence from selected EU member states’, *Energy Res. Soc. Sci.*, vol. 127, p. 104315, Sep. 2025, doi: 10.1016/j.erss.2025.104315.

[17] G. M. Huebner, I. Hamilton, Z. Chalabi, D. Shipworth, and T. Oreszczyn, ‘Explaining domestic energy consumption – The comparative contribution of building factors, socio-demographics, behaviours and attitudes’, *Appl. Energy*, vol. 159, pp. 589–600, Dec. 2015, doi: 10.1016/j.apenergy.2015.09.028.

[18] P. Carroll, M. Chesser, and P. Lyons, ‘Air Source Heat Pumps field studies: A systematic literature review’, *Renew. Sustain. Energy Rev.*, vol. 134, p. 110275, Dec. 2020, doi: 10.1016/j.rser.2020.110275.

[19] C. Piccardo, A. Dodoo, L. Gustavsson, and U. Tettey, ‘Retrofitting with different building materials: Life-cycle primary energy implications’, *Energy*, vol. 192, p. 116648, Feb. 2020, doi: 10.1016/j.energy.2019.116648.

[20] L. McCarthy *et al.*, ‘Decarbonising Off Grid Homes: A Rapid Evidence Review’, Jun. 2024, [Online]. Available: https://discovery.ucl.ac.uk/id/eprint/10207034