

# A methodological framework for archotyping properties and areas for targeted place-based retrofit

Dr. Joseph Day<sup>\*1</sup>, Lawrie Swinfen-Styles<sup>†2</sup> and Dr. Laurie Duncan<sup>‡2</sup>

<sup>1</sup>School of Geography, Earth and Environmental Sciences, University of Birmingham

<sup>2</sup>West Midlands Combined Authority

GISRUK 2026

## Summary

A place-based approach to low-carbon property retrofits has the potential to provide significant advantages over the current ad hoc approach. In this paper, we provide a computationally lightweight framework for archotyping properties and Lower Super Output Areas using publicly available data. This framework can be overlaid with various datasets which can then be used to analyse the potential for place-based interventions, thereby maximising the benefits of this approach.

**KEYWORDS:** Place-based retrofit, decarbonisation, archotyping, energy demand, clustering

## 1. Introduction

Retrofitting existing properties – with insulation and low carbon technologies – is an increasing priority for both government and local/combined authorities. The ramping up of retrofit programs is vital for meeting carbon targets, as well as addressing fuel poverty and equity concerns and to prepare the housing stock for the future impacts of climate change. To date, most retrofits have been performed on an ad hoc basis, confined to siloed schemes for certain tenures. Research in this sector also tends to focus on individual properties, scaling the individual solutions up to national level (Kaveh et al., 2018; Bennadji et al., 2022; Li et al., 2022). Recently, a pattern book for a common UK archetype (the flat-fronted Victorian terraced house) has been published by the National Retrofit Hub (2024). However, this approach to retrofit has several limitations. Auditing and planning retrofits on individual buildings is inefficient due to lead times, travel between sites and differing design constraints between projects. Additionally, the potential benefits of the process are limited to targeted properties, providing little or no advantage to surrounding neighbourhoods.

By comparison, a ‘place-based’ approach – a holistic planning of retrofit across a local area – can improve outcomes. A place-based approach to climate action could deliver “double the socio-economic benefits, at a third of the cost, compared with a national ‘one size fits all’ approach” (Innovate UK, 2022). Nearby properties often have similarities; as such, design constraints can be grouped together for easier planning. Interventions can be designed to make use of the renewable generation potential and demand profile of the entire area, rather than limiting scope to a single site. For example, a warehouse with a large roof area but low demand might still benefit from solar panels by providing energy to a nearby manufacturer or group of households. In this way, grid constraints can both be planned for (by targeting areas with looser constraints) and avoided through intelligently designed co-operation between properties. Areas can be targeted based on a range of desired co-benefits, such as fuel poverty, grid constraints, carbon saving potential, and ease of installation.

In this paper, we describe a methodological framework for archotyping domestic and non-domestic

---

\* J.K.Day@bham.ac.uk

† Lawrie.Swinfen-Styles@WMCA.org.uk

‡ Laurie.Duncan@WMCA.org.uk

properties by several variables, expanding to the Lower Super Output Area (LSOA) scale, which will provide the basis for this data-driven, place-based retrofit planning approach. This framework is scalable, spatially explicit and creates a GIS-informed base layer over which a wide variety of spatially dependent metrics can be laid.

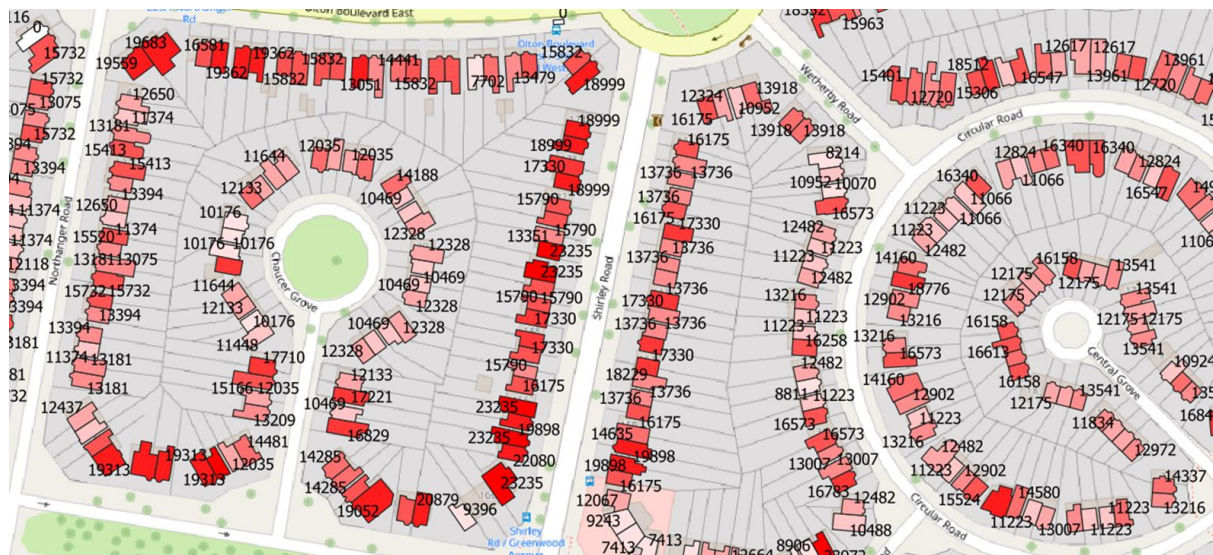
## 2. Study area and key datasets

The West Midlands Combined Authority (WMCA) region was the test site for this work; the methodology is applicable to any geography (accounting for differences in data sources and availability).

The Ordnance Survey datasets AddressBase Premium, Master Map Topography Layer and Building Heights were used to gather property data for the region. These are definitive sources for property addresses (with co-ordinates) and building footprint polygons, assigning each a Unique Property Reference Number (UPRN) and Topographic Identifier (TOID), respectively.

The UK Government's Energy Performance Certificate (EPC) Database provided key property information, including insulation measures and property age. This data was cleaned by removing duplicates and keeping only the most recent EPC for each UPRN. A 'nearest neighbour' pass was then used to assign data for properties that have no or outdated EPCs (not all homes have EPCs, since they are only required when selling or renting property). Non-domestic EPCs could be matched from the same source, although they contained less information that was practically useful for our analysis.

UK Government datasets based on metered usage were used to infer gas and electricity consumption at an individual property level. These were the Postcode Level Annual Domestic Gas Consumption (PLADGC), National Energy Efficiency Database (NEED), non-domestic NEED (ND-NEED) and Building Energy Efficiency Survey (BEES). A representation of the output of these datasets can be seen in Figure 1.



**Figure 1** The domestic building footprints from Master Map Topography Layer labelled with their annual gas consumption in kWh and overlayed on Open Street Map

## 3. Methodology

### 3.1. Domestic properties

The methodology was implicitly spatial as it relied upon combining data of known and estimated

attributes at different levels of geography (e.g. a known postcode gas consumption, to an unknown UPRN or TOID level consumption). NEED data gives expected annual gas consumption for a list of property archetypes, split by property type, age, bedroom count and presence of a gas meter. This is summarised in Table 1.

**Table 1** Summary of NEED domestic archetype variables

Property type	Property age band	Bedroom count	Gas meter present?
End-terrace	Pre-1919	1 bedroom	Yes
Mid-terrace	1919-1944	2 bedrooms	No
Flat	1945-1964	3 bedrooms	
Detached	1965-1982	4 bedrooms	
Semi-detached	1983-1992	5+ bedrooms	
Bungalow	1993-1999		
	Post 1999		

These variables formed the basis for our archetypes but were simplified with the removal of bedroom count, Bungalow property type and the combining of 1983-1992 and 1993-1999 age bands. Bungalow is not a classification within AddressBase Premium (exploratory methods to infer bungalow status from building height proved uncertain) and 1983-1999 was merged as one category to give each age band approximately 20-year time coverage. This reduced the total number of archetypes to 30 (60 when including gas and non-gas heated) for ease of future analysis, and to reduce the number of clusters (with less variation within each cluster) produced by the subsequent K-means clustering.

To group properties within the region into these archetypes, UPRNs from AddressBase Premium were assigned an EPC, (TOID, Building Heights value (from which floor count was inferred, where uncertainty is less problematic for taller buildings), and additional AddressBase Premium postcode and Building Land and Property Unit (BPLU) classification code data. To account for discrepancies between NEED and EPC property age brackets, properties were assigned the NEED age bracket with the greatest overlap with its EPC bracket.

From these attributes, we inferred gas consumption for each property based on the variation of archetypes across the postcode. To improve accuracy, properties were grouped into postcodes and the sum of their consumption compared with PLADGC data. The ratio of inferred to measured postcode consumption was then applied to properties within each postcode, bringing total consumption in line with PLADGC. Estimated annual electricity consumption was based on NEED only due to limited coverage of postcode statistics and greater uncertainty of the split between heating and non-heating usage.

### 3.2. Non-domestic properties

Non-domestic properties were similarly matched from UPRN to TOID to obtain height and footprint/floor area. Each was assigned one of ten building use categories (e.g. education, factory) from the ND-NEED dataset by mapping from its BPLU classification code. 40 archetypes were then defined based on floor area with cut-off for small, medium, large and extra-large at the quartiles of the distribution. Annual gas consumption was estimated from ND-NEED based on the non-electrical energy intensity per m<sup>2</sup> from an interpolated table given by building use and floor area band, with each archetype matched to its floor area band.

Through archotyping of domestic and non-domestic properties, we had a dataset containing each property, key attributes, and approximations of annual gas and electricity consumption. Energy and carbon savings of typical measures was derived by comparing EPC energy consumption within archetypes showing the lack or presence of insulation measures. Similarly, for non-domestic

archetypes, BEES could be used to calculate indicative savings from various interventions.

### 3.3. Expanding to LSOA scale with K-means clustering

While building-level data is valuable, the number of properties in any local or combined authority area mean that analysis at that scale would be prohibitively resource intensive. To this end, we have used K-means clustering to create two sets of LSOA archetypes: one for domestic and one for non-domestic properties. A total of 17 domestic and 7 non-domestic variables were used, summarised in Table 2.

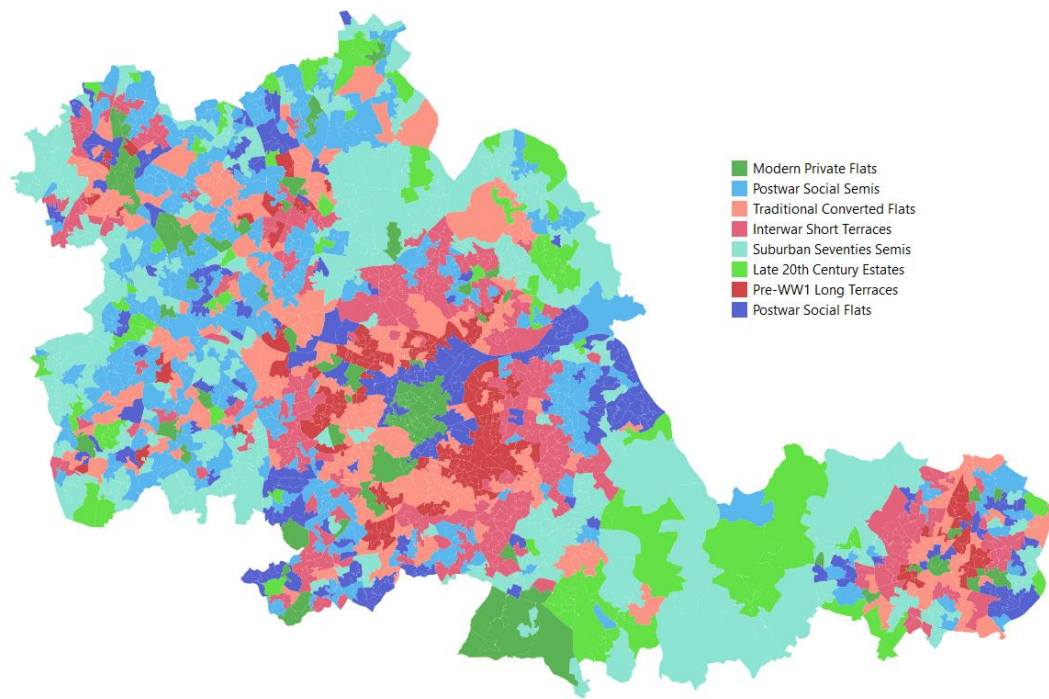
**Table 2** Variables used for K-means clustering for LSOAs, split into domestic and non-domestic

Domestic variables	Non-domestic variables
<ul style="list-style-type: none"> <li>• % of each age band (6 brackets)</li> <li>• % of each property type (5 brackets)</li> <li>• % tenure (3 brackets: social rented, private rented, owner-occupied)</li> <li>• % EPC (3 brackets: E-G, C-D, A-B)</li> </ul>	<ul style="list-style-type: none"> <li>• Total non-domestic floorspace</li> <li>• Average non-domestic footprint area</li> <li>• Average building height</li> <li>• Total factory floor area</li> <li>• Total warehouse floor area</li> <li>• Total public sector floor area (e.g. health, education, emergency services)</li> <li>• Total retail, hospitality, arts, and community leisure floor area</li> </ul>

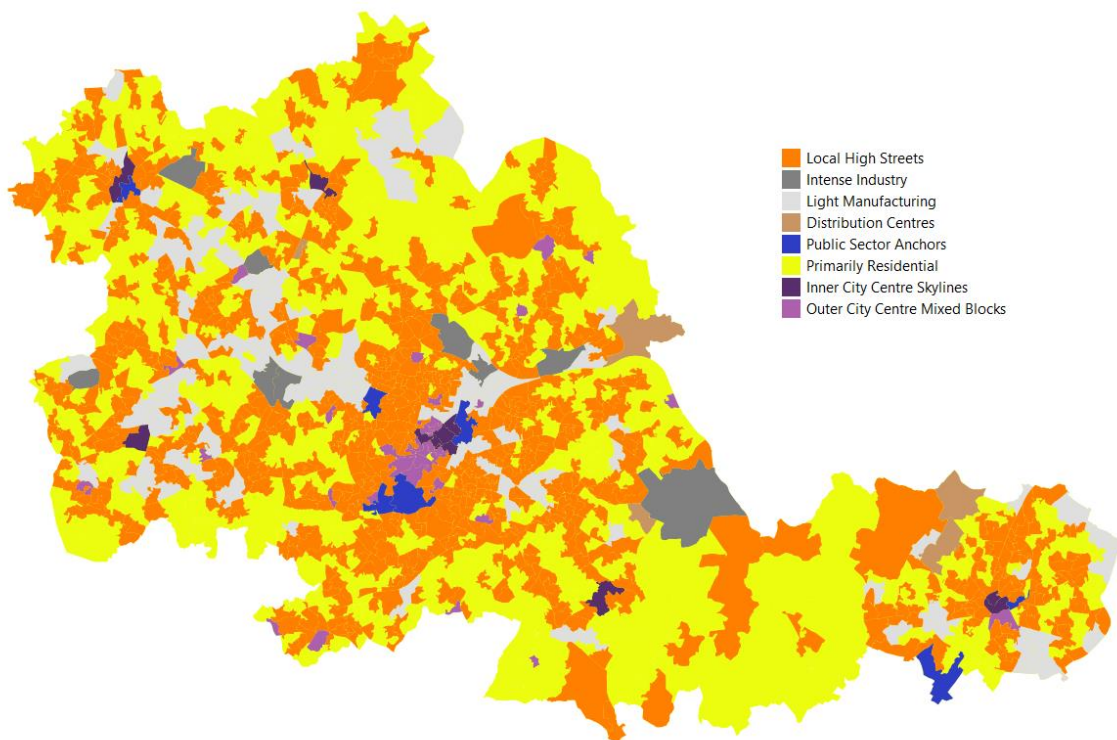
LSOAs were assigned values for each of these variables, and K-means clustering produced 16 total clusters (8 for each domestic and non-domestic). There were 1719 LSOAs in total and 8 clusters were chosen following visual inspection of the respective elbow curves. The results of this process can be seen in the next section.

## 4. Outputs and illustrative results

This methodology produced GIS data layers for domestic and non-domestic properties in the WMCA region. The 8 clusters in each case have been named to best describe the type of properties found within each LSOA archetype. These archetypes can be seen in Figures 2 and 3, along with the distribution of intersections for the domestic and non-domestic archetype of LSOAs in Table 3.



**Figure 2** Domestic LSOA archetypes across the WMCA region



**Figure 3** Non-domestic LSOA archetypes across the WMCA region

**Table 3** Distribution of intersections for the domestic and non-domestic archetype of LSOAs

<b>Archetype Intersections</b>	<b>Local High Streets</b>	<b>Intense Industry</b>	<b>Light Mfg.</b>	<b>Dist. Centres</b>	<b>Public Sector Anchors</b>	<b>Primarily Residential</b>	<b>Inner City Centre Skylines</b>	<b>Outer City Centre Mixed Blocks</b>	<b>Row Total</b>
Modern Private Flats	33	0	12	0	4	20	7	14	90
Postwar Social Semis	126	0	27	1	0	178	0	3	335
Traditional Converted Flats	147	5	19	0	2	51	2	9	235
Interwar Short Terraces	131	1	3	3	0	174	0	1	313
Suburban Seventies Semis	97	2	9	1	0	172	0	0	281
Late 20th Century Estates	21	0	3	0	1	56	0	1	82
Pre-WW1 Long Terraces	135	0	13	0	0	33	0	0	181
Postwar Social Flats	102	1	17	0	0	70	2	10	202
Column Total	792	9	103	5	7	754	11	38	1719

Additionally, we have been able to validate the estimated energy and carbon savings of each intervention by comparing against results from Cotality's Pathways tool (2026), which resulted in sufficiently similar savings across comparable archetypes (within  $\pm 20\%$  at a 2023 baseline, and within  $\pm 1\%$  out to 2050).

## 5. Discussion: value and next steps

In the creation of this scalable and widely applicable framework, we have developed a methodology by which areas across a region might be compared for retrofit planning and design. Additional GIS layers



can be easily applied to this base layer to target, for example, areas of high fuel poverty, or those with large swathes of unused rooftop for solar PV installation. This is a process that is already underway, aided significantly by the work presented here.

Clusters of the same LSOA archetype can also be visually identified from the map. For example, the Black Country industrial cluster or swathes of hard-to-treat properties in East Birmingham. These align with existing plans for decarbonisation policies, but new geographical priority areas for schemes and supply chains could potentially be identified through this dataset. The intersection between the spatial variation of domestic and non-domestic could also be further explored to create new clusters, drawing up on more data such as electricity or heat consumption profiles, while more computational methods could be applied for the purposes of validation or spatial cluster identification. For validation this could include visualisation of clusters via Principal Component Analysis and for spatial clusters, techniques such as Moran's I could be utilised.

Another advantage of this archotyping approach is the reduced computational cost of granular analysis. Cotality's Pathways tool can provide data for every home in the UK. However, analysing the potential impact of retrofit measures on a building-by-building basis is extremely time-consuming and requires substantial data storage even for smaller council areas. Moreover, building-level data is only accessible through costly licenses which are not accessible to all local authorities given the financial challenges experienced across the sector. It is the hope of the authors that this methodology might significantly ease the resource intensity of place-based retrofit planning.

## 6. Conclusion

In this work, we have described a methodology by which properties and larger areas can be simply archotyped, allowing for comparative design decisions to be made for place-based retrofit planning. This framework can be extended nationally with little adjustment and allows for additional comparative data to be laid over top of the base layer we have created.

We carried out building-level analysis of carbon saving potential across the West Midlands to compare against archotyped results. We found estimates within  $\pm 1\text{--}\pm 20\%$ . This indicates that our choice of variables for archotyping capture most detail needed to understand housing stock implications for decarbonisation. Since this simpler analysis proves to be an appropriate substitute for high-level carbon savings potential, analytical capacity can be unlocked to investigate more specific implementation considerations.

## 7. Acknowledgements

The authors would like to thank the West Midlands Combined Authority for funding the present work and the WM-Net Zero project which is supported by the Wellcome Trust (227150\_Z\_23\_Z) under the Advancing climate mitigation policy solutions with health co-benefits in G7 countries scheme.

## References

- Bennadji, A., Seddiki, M., Alabid, J., Laing, R. and Gray, D. (2022). Predicting Energy Savings of the UK Housing Stock under a Step-by-Step Energy Retrofit Scenario towards Net-Zero. *Energies*, 15(9), p.3082.
- Cotality (2026). *Pathways: housing analytics for warm homes*. Available at: <https://www.cotality.com/uk/products/pathways> (Accessed: 25<sup>th</sup> March 2026)
- Innovate UK (2022). *Accelerating Net Zero Delivery*. Available at: <https://www.ukri.org/wp-content/uploads/2022/03/IUK-090322-AcceleratingNetZeroDelivery-UnlockingBenefitsClimateActionUKCityRegions.pdf> (Accessed: 25<sup>th</sup> March 2026)

Kaveh, B., Mazhar, M.U., Simmonite, B., Sarshar, M. and Sertyesilisik, B. (2018). An investigation into retrofitting the pre-1919 owner-occupied UK housing stock to reduce carbon emissions. *Energy and Buildings*, 176, pp.33–44.

Li, X., Arbabi, H., Bennett, G., Oreszczyn, T. and Densley Tingley, D. (2022). Net zero by 2050: Investigating carbon-budget compliant retrofit measures for the English housing stock. *Renewable and Sustainable Energy Reviews*, 161, p.112384.

National Retrofit Hub (2024). *Retrofit Pattern Book: Flat-fronted Victorian Terraced House*.

Available at: <https://nationalretrofithub.org.uk/wp-content/uploads/2025/01/NRH-Retrofit-Pattern-Book.pdf> (Accessed: 25<sup>th</sup> March 2026)

## **Biographies**

Dr Joseph Day is a knowledge exchange fellow at the University of Birmingham for the West Midlands-Net Zero project. He has experience working with local government and the non-profit sector. His current research interests are energy systems data and the interactions between climate and health policies.

Lawrie Swinfen-Styles is a research analyst in the Net Zero team at the West Midlands Combined Authority. He has a research background in renewable energy and energy storage and is interested in GIS data, holistic decarbonisation solutions and systems modelling.

Dr Laurie Duncan is a Programme Manager for net zero and energy policy, research and engagement at the West Midlands Combined Authority. His PhD research investigated the use of evidence in the development of net zero policies in local and regional government.