Methodology for Nesta's local heat planning data

Contents

Introduction	**2**
Definitions	**2**
Technology groups	**3**
Overview of our current methodology	**6**
1. Domestic properties	**9**
2. Property-level technology assignment	**14**
3. Building-footprint level technology assignment	**27**
4. Technology reassignment due to anchor load proximity	**29**
5. Spatial aggregation	**30**
6. District heat network recommendations	**34**
7. Contextual features	**35**
Annex	**42**

Last updated: November 2nd, 2026. Please note that, at the time of reading, the data and evaluation results contained in this document may no longer reflect the most current developments displayed on the local heat planning tool.

Introduction

This document details the methodology used to create the dataset underlying Nesta's Local Heat Planning Tool, as well as associated caveats and assumptions.

This dataset:

- supports with identifying the most suitable low-carbon heating technology category for clusters of neighbouring domestic properties
- provides contextual information about the properties in each cluster and its neighbourhood
- contains additional context layers, including location of anchor loads, ward boundaries, heat network zones and city centres

The methodology detailed in this document and data generated by it are currently in the **beta stage of testing** and cover all local authorities in Great Britain.

You can find a full list of the datasets, and their respective licenses, used in this work in the bibliography, as well as the license of Nesta's local heat planning dataset. All of the code can be accessed in Nesta's asf_heat pump_suitability repository on Github.

Definitions

Domestic property, usually used interchangeably with property: structure used wholly or mainly for private living accommodation and residential purposes. It may or may not be its own building. Examples include a flat in a building with five other flats or a detached property.

Domestic building, usually used interchangeably with building: permanent, enclosed structure with walls and a roof, for living purposes. May contain one or multiple properties. Buildings may contain domestic properties only or may be mixed use.

Building footprint: the two-dimensional horizontal area of ground that a structure covers, measured along the outside face of its exterior walls where they meet the ground.

Block of flats: a residential building with 3 or more storeys divided into several separate individual properties (e.g. tower blocks, tenements, 3 or more storeys of flats above shops and large mansions converted into 3 or more storeys of flats).

Anchor loads: are high heat demand, or potential sources of waste heat - non-domestic sites such as hospitals, schools or airports.

Technology groups

These categories have been developed through conversations with the supply chain. They look to give certainty on the types of technology that will be deployed to allow for focused delivery conversations and decisions, while acknowledging the complexity of clean heat and giving the flexibility for a range of solutions to be considered when working through delivery plans.

Individual heat sources

Each property has a standalone heat source(s), not reliant on any communal infrastructure, other than a connection to the electric grid.

Example technologies



Air source heat pump (ASHP), air-to-air (A2A), ground source heat pump (GSHP)

Most suitable for

Most suitable for properties with outdoor space (over 30m2) which can accommodate an outdoor unit.

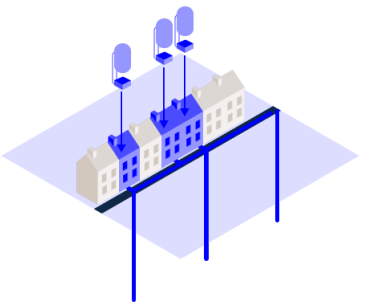
Networked heat pumps

A series of networked boreholes serving multiple properties, sometimes situated in roads or open ground, that are connected to an individual ground source heat pump in each property. The ground source heat pump is a similar size to a boiler and a hot water tank is required. They require no outdoor space.

As they are well suited to a very particular context they have a technology group of their own.

Example technologies

Networked ground source heat pumps (networked GSHP), also known as shared ground loops.



Most suitable for

Most suitable for dense properties with little to no private outdoor space (up to 30m2). Properties that are listed buildings, or in a protected area, may be more suitable for this approach due to the lack of changes to the built environment. Networked heat pumps are a subset of communal solutions and a specific type of communal solution. A networked heat pump is selected when this specific communal technology is the best fit. Where a communal solution is most suitable but the specific type can't be decided without further understanding or data on the properties and buildings, a broader communal solution is assigned by default.

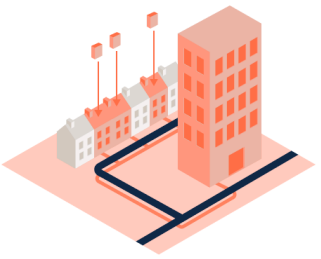
Communal heat sources

A large singular heat source, or networked heat pump, providing heat to multiple dwellings. These types of solutions are sometimes called a communal heat network.

The inclusion of networked heat pumps and singular heat sources in this category is to recognise that both may be suitable, and the decision between which solution is preferred may come down to the characteristics of the properties and the commercial model used to deliver.

Example technologies

Communal ASHP and networked GSHP.



Most suitable for

Dense properties with little or no private outdoor space (up to 30m2), with an existing communal heating system, room for a plant room or an outdoor communal heating source, or those near waste heat sources.

Blocks of flats are always assigned a communal solution. Individual houses with limited outdoor space (up to 30m2) are assigned communal solutions if within 50 metres of an anchor load (such as a hospital or school).

District heat networks

A network where heat is generated at a single central location. This heat is then distributed to various customers through insulated underground pipes filled with hot water. These networks are scalable and can be designed to serve a small group of buildings with a high heat demand or expanded to provide warmth for an entire city.

Example technologies



Networks powered by waste heat sources, industrial or large-scale heat pumps

Most suitable for

Dense properties with little or no outdoor space, with an existing communal heating system. Most suitable for city areas with a mixture of commercial, public and private buildings.

Overview of our current methodology

Identifying the most suitable technology for clusters of neighbouring properties

The most suitable low-carbon technology for each property, building, and neighbourhood is determined by the following key factors:

- Property outdoor space availability
- Building type (whether a building is a block of flats)
- District network potential (city centre locations and designated heat network zones)
- Anchor load proximity (presence of nearby heat loads)

The following diagram shows an overview of how the most suitable technology is identified:

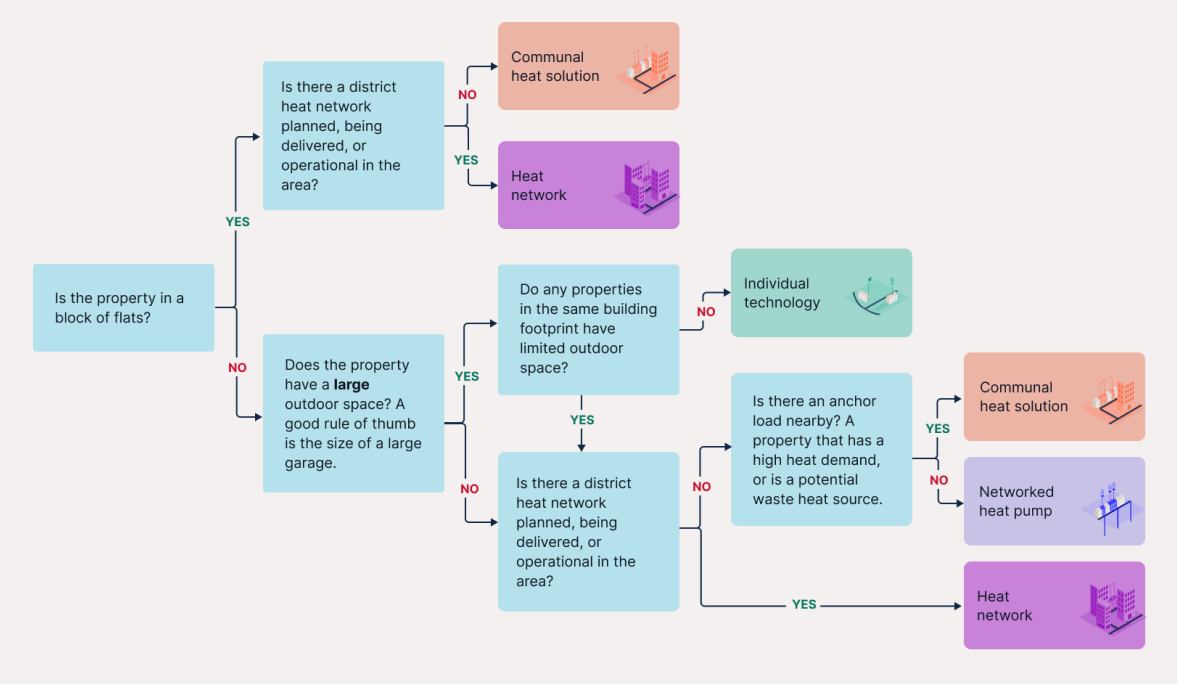


Figure 1. An overview of the methodology to identify the most suitable low carbon technology for each property. Note that we don't currently collate data on operational district heat networks.

A breakdown of the pipeline is shown below:

1. Domestic properties

This work is focused on identifying the most suitable low carbon tech category for domestic properties. As such, domestic unique property reference numbers (UPRNs), hereafter used interchangeably with 'properties', and their corresponding building footprints are identified.

2. Property-level technology assignment

The technology group for each property is determined using a decision tree. This decision tree has two main questions that relate to the property: This includes building type (i.e., whether the property is in a block of flats) and the property's available outdoor space (above 30m2 assigns 'individual solution' and up to 30m2 assigns 'networked heat pump').

3. Building-level technology assignment

Property-level technology assignment is aggregated to the building footprint level. The most communal solution takes precedence. eg, If a building contains a mix of individual and communal solutions for different properties, the most communal solution takes precedence.

4. Technology reassignment due to anchor load proximity

Buildings located near high heat demand sites (also known as anchor loads), such as hospitals or schools, are reassigned to communal solutions to leverage shared infrastructure. We currently set a threshold of 50m distance from anchor loads for technology reassignment.

5. Spatial aggregation

Buildings are grouped into geographic clusters based on their proximity and assigned technology.

6. District heat network recommendations

District heat networks recommendations are not assigned as most suitable by default, as these represent potential plans for district heat networks rather than confirmed/networks currently being built. Clusters most suitable for a district heat

network, can be identified through **'heat network recommendations'** which enable reassigning eligible clusters into district heat networks. Eligible clusters include clusters originally assigned **Communal solution** or **Networked Heat Pump** within a zone of high potential for district heat networks. Homes in these areas will be most suitable for a district heat network connection when/if a district heat network is constructed.

Important notes

- This work identifies the most suitable low carbon heating technology group for groups of properties, based on the properties themselves and the surrounding neighbourhood. Other technologies might still be suitable for that cluster. *Eg.* a networked heat pump solution might be identified as the most suitable solution for a row of terraces. Even so, individual solutions (such as air source or ground source heat pumps) might still be suitable for all or at least some of the properties in that cluster
- There may be a mismatch between the solution identified as most suitable for a cluster of properties, and any existing low carbon heating systems those properties already have in place. That does not mean the low carbon heating solution installed isn't suitable, but the output of this work will not be evaluated against what has been installed. *Eg*. A homeowner might decide to install air source heat pumps in their home. Our modelling might identify a different solution as most suitable, a networked heat pump for example, based on the building and surrounding neighbourhood.

Contextual information to support conversations between local authorities and supply chain and enable delivery

Additional data features are added to the clusters to give further context about the properties within them and surrounding neighbourhood.

Additional context layers

Additional context layers have been added to the data layer, to support decision making. These currently include ward boundaries, location of anchor loads, heat network zones and city centre areas. These are added by joining the relevant datasets to the local heat planning data.

1. Domestic properties

The OS Open UPRN dataset provides the baseline input data for this work. This dataset contains all addressable locations within Great Britain, so must be filtered to UPRNs which represent domestic properties only. For a UPRN to be considered domestic, a set of inclusion and exclusion criteria are defined. Domestic UPRNs are assumed to:

- Be listed in the domestic Energy Performance Certificate (EPC) register

OR

- Be located within a building footprint from the OS OpenMap Local Building layer AND;
- Not be listed in the non-domestic EPC register AND;
- Not be located within a building footprint considered to be non-domestic

i. Identifying non-domestic building footprints

Non-domestic building footprints are defined as those building footprints which are unlikely to contain any domestic properties, e.g. hospitals, large railway stations, and factories. Mixed-use (i.e. domestic and non-domestic properties in the same building) are not included in this definition.

A set of non-domestic building footprints is created from the following datasets:

- OS OpenMap Local railway station layer geopoints
- Overture Maps Foundation and GeoDS Points of Interest for the UK geopoints
- OS OpenMap Local important building layer polygons. These are buildings that fall within the extent of a functional site. They are classified into different themes such as air transport, cultural facility, emergency services.

The Points of Interest dataset contains the geographic locations and classifications of points of interest covering the UK. There are 1839 different classifications which include categories such as 'hotel', 'jewelry_store', and 'cafeteria'. Using GPT-5, these were filtered to 189 categories deemed to represent points which are unlikely to share a building with domestic properties. The initial ChatGPT prompt is shown in Annex Text 1. Further prompting with manual review was used to refine the outputted

list. The final prompt which generated the finalised list of non-domestic points of interest is shown in Annex Text 2.

The OS OpenMap Local railway station layer and the selected Points of Interest data both contain point geometries only, so must be joined with building footprint data. They are spatially joined to the OS OpenMap Local building footprints they are located within to identify railway station buildings and buildings of interest, respectively. These building footprints are then concatenated to the OS OpenMap Local Important Buildings footprints to generate a full set of non-domestic buildings. Note, all classes in the important buildings layer were retained. See the full list of classes in this dataset in Annex Table 1.

Additionally, building footprints are categorised as non-domestic if the footprint area per predicted domestic UPRN is greater than 261m2. All UPRNs within these buildings are subsequently excluded from the final dataset of domestic UPRNs. This threshold was determined by assessing different candidate thresholds for a selection of candidate features and optimising for the feature and threshold with the best Matthew's Correlation Coefficient (MCC) value, using council tax UPRNs in Plymouth as actuals. This threshold had an MCC of 0.48. The use of this threshold across other Local Authorities has not been reviewed, but it has been implemented in our pipeline for all Local Authorities in Great Britain. Further work may refine this.

Building footprints are then excluded from this final non-domestic set if they contain at least one UPRN listed in the domestic EPC register.

After applying the described methodology to OS Open UPRN data, there are 33,808,443 UPRNs across GB retained in the dataset for further analysis. For comparison, the number of domestic dwellings in Great Britain published by government departments is 30,075,100, as of March 2025. This number is calculated by combining separate estimates from the Ministry of Housing, Communities & Local Government, National Records of Scotland, and the Welsh Government. This shows that our pipeline is estimating about 1.12 times the number of true domestic dwellings in Great Britain.

ii. Caveats of domestic property identification methodology

- UPRNs which are not in the domestic EPC register will only be retained as domestic if they are joined to a building footprint. The OS OpenMap Local Building layer footprint geometries are generalised, meaning they are

simplified while preserving key geospatial information and patterns, e.g. general shape, size, and proximity. Some true domestic UPRNs are likely to be erroneously excluded due to either this generalisation or other small mapping errors where a UPRN geometry falls just outside a building footprint boundary.

- There are likely some cases where a true domestic UPRN is located within a building categorised by our criteria as non-domestic. This means they will be erroneously missing from the final set of domestic UPRNs.
- Non-domestic UPRNs will be erroneously retained as domestic UPRNs if they are not listed in the non-domestic EPC register AND they appear in a) a building footprint not labelled as non-domestic, or b) a building footprint labelled as non-domestic but which contains a UPRN listed in the domestic EPC register. This means that in mixed-use buildings, all non-domestic UPRNs will be retained in the final domestic UPRN dataset unless they are listed in the non-domestic EPC register. Although we intentionally retain mixed-use buildings, we would ideally remove all non-domestic properties within those buildings from counts of domestic UPRNs.
- There is incompleteness and missingness in both the EPC registers and the Points of Interest datasets. This limits our ability to completely accurately identify both non-domestic building footprints and footprints containing domestic UPRNs.
- Domestic UPRNs are later mapped to their corresponding building footprints using an intersecting geospatial join. Some UPRNs identified as domestic from the EPC register do not get joined to a building footprint. In these cases, they are joined to the nearest building footprint within 1 metre. The remaining domestic EPC UPRNs with no matching building footprint are excluded from further analysis. This rule is only applied to UPRNs in the domestic EPC register because if these UPRNs fall outside a building footprint, we know the UPRN is meant to represent a domestic property. However, if a UPRN is not in the domestic EPC register, there is no certainty that it represents a property, e.g. it could be any external address point like a postbox, bus stop, monument etc.

iii. Evaluation against council tax data

Plymouth council tax data contains a list of properties with UPRNs and point coordinates. These provide a dataset of confirmed domestic UPRNs, and were used

as a ground truth dataset to validate the UPRNs selected by the domestic UPRN identification pipeline. Some council tax records are missing UPRNs and point coordinates, so were not used for this analysis.

- Total number of council tax records: **123,478**
- Total number of UPRNs with a coordinate in council tax records: **122,315**. These are the UPRNS used for this analysis and referred to as 'council tax data'
- Number of unique UPRNs identified by the domestic filtering pipeline: **147,514**. From this point, these will be referred to as 'pipeline data'

False negatives

The UPRNs in the council tax data that are not present in the pipeline data represent an estimate of the number of domestic UPRNs that the pipeline is not correctly classifying as domestic. The number of domestic UPRNs in the council tax data not found in the pipeline data is **1693**, which is **1.38%** of all council tax UPRNs. This indicates we are successfully identifying the majority of true domestic UPRNs as domestic, and are not missing a large number of true domestic UPRNs.

False positives

The UPRNs in the pipeline data that are not present in the council tax data represent an estimate of the number of UPRNs we are incorrectly identifying as domestic. The number of UPRNs in the pipeline not found in the council tax data is **26,892** (**18.23%** of the pipeline UPRNs). This indicates that a large proportion of pipeline UPRNs are not found in the council tax data (with a small margin for error, as 1163 council tax records currently lack an associated UPRN).

At building level, the building footprints that contain at least one UPRN identified as domestic by the pipeline were compared to the council tax data to determine how many entire buildings we are identifying incorrectly as domestic. The proportion of pipeline building footprints with at least one confirmed domestic UPRN is **97.76%.** This indicates that the majority of buildings containing at least one pipeline UPRN do have at least one confirmed domestic UPRN located within them, and are present in the final dataset.

We counted the number of mixed-use buildings (i.e. buildings containing at least one domestic and at least one non-domestic property) identified by each dataset. To

create a ground-truth set of mixed-use buildings, we used the full UPRN dataset (domestic and non-domestic) to find buildings with a different number of total UPRNs to council tax UPRNs. The proportion of all building footprints which were identified as true mixed-use is **11.38%.**

11% of building footprints identified by the pipeline as containing domestic UPRNs are true mixed-use buildings, meaning they contain both domestic and non-domestic UPRNs. These buildings contain 88% of non-domestic UPRNs that the pipeline has misclassified as domestic, so false positives are mainly occurring in true mixed-use buildings. Ultimately, this indicates that, at least for Plymouth, the majority of buildings retained in the dataset contain at least one domestic property.

iv. [Not currently implemented, in development] Future improvements to domestic property identification

Further exploration may involve the creation of a machine learning classification model to identify buildings containing domestic UPRNs, or to identify UPRNs themselves as domestic.

2. Property-level technology assignment

a. Decision tree

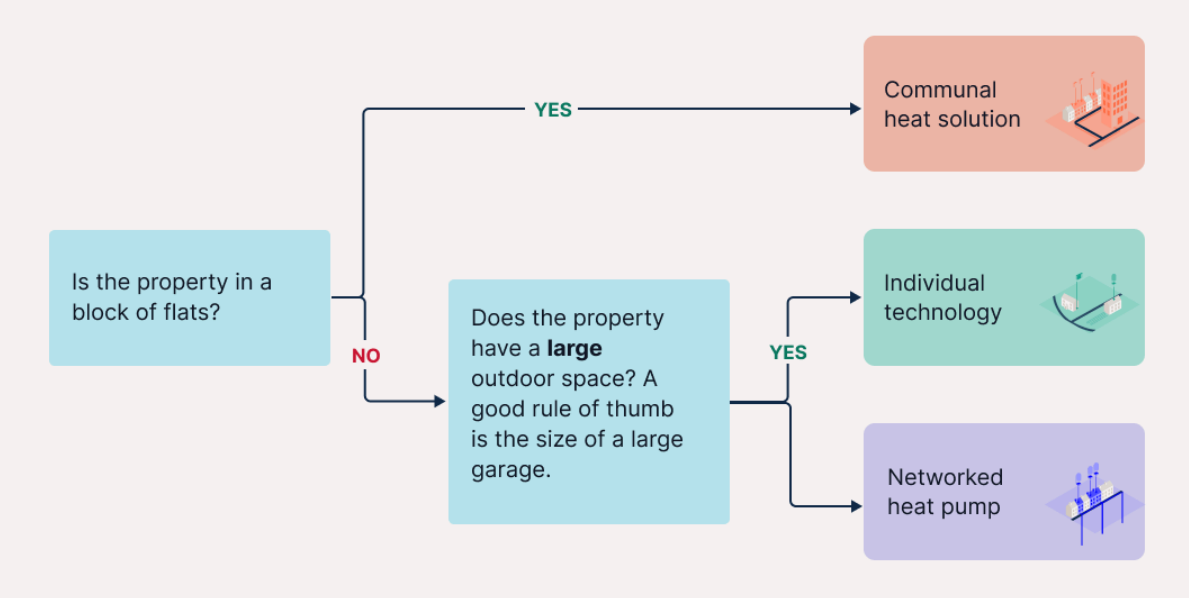


Figure 2. Decision tree to identify the low carbon heating technology group for a property.

The initial technology group assigned to each property is determined using the decision tree in **Figure 2**. This decision tree has two main questions that relate to the property and neighbouring area:

- **Building type:** Whether the property is in a block of flats. This is determined via a machine learning model (see ***Section 2.b.*** for more detail)
- **Contiguous outdoor space**: property's maximum contiguous outdoor space. To provide a practical scale, a "large" outdoor space is defined as over 30m2* (comparable to a large garage).

*The 30m2 threshold was selected as the amount of outdoor space required to consider an individual solution the most suitable low-carbon home heating option. This threshold was determined by looking at the first percentile of estimated total outdoor space (35m2) for properties where an air-source heat pump has already been installed (analysis of MCS installations data 2025), along with manual refinement through viewing specific examples and accounting for contiguous outdoor space.

In essence the decision tree assigns a technology based on the information below:

- **Multi-unit developments or blocks of flats:** Properties are assigned communal solutions, such as communal air source heat pumps systems.
- **Individual dwellings or houses:** For houses, the outdoor space is the primary determinant of the assigned technology group. Houses with large outdoor space are assigned individual solutions, such as air source heat pumps. For houses with limited outdoor space, properties will be assigned shared infrastructure solutions.

Later in the pipeline, the **proximity to heat networks zones** and **anchor loads**, may change the technology assigned to a property and building.

The decision tree was built on a foundation of data and local insight, and it has been evaluated in comparison to a manually labelled ward (the Stoke ward in Plymouth). It has also been tested on a small number of diverse properties across Great Britain. As part of the alpha testing phase, we also gathered feedback about this methodology from four additional local authorities in Great Britain, spanning rural and urban areas in England, Scotland, and Wales .

i. Caveats

- Some buildings contain multiple UPRNs which intersect with the same land parcel (e.g. this might happen for houses converted into flats). Land parcels are used to calculate outdoor space. We are unable to identify which UPRN the land parcel belongs to, so our current approach assigns the outdoor space to all intersecting UPRNs. As a result, some properties in buildings that contain multiple properties may be erroneously assigned 'individual heat pumps' as the most suitable solution. The building as a whole may have sufficient outdoor space for an individual heat pump, but a particular UPRN may not actually have access to that space (e.g. it belongs to another property in the building).

- We currently have fixed thresholds for the area of outdoor space. In the upcoming phase of work, we will test the impact of threshold fluctuations on our outputs to decide between maintaining fixed parameters or adopting a continuous modeling approach.

b. Classifying properties as 'in block of flats' or not

We trained a random forest binary classifier model which labels building footprints as either blocks of flats or not. Blocks of flats are considered to be mainly either purpose-built blocks of flats with three or more storeys (including above shops) or large buildings converted into multiple dwellings - detailed descriptions of the various archetypes considered to be blocks of flats and non-blocks can be found in Table 1. Features are generated per building from building footprint and UPRN geodata.

i. Labelling individual properties as flats

The processes of both feature engineering for the model and sampling for manual labelling require individual domestic UPRNs to be labelled with a boolean flag indicating whether they represent individual flats or not. For the purposes of this work, flats may be located within purpose-built blocks within converted houses or other buildings. A UPRN is considered to represent a flat if it is labelled by our pipeline as domestic and is located at the exact same coordinates as any other domestic UPRN. This is based on the assumption that UPRNs in the exact same location share an address point (e.g. a front door) and so are most likely to represent flats.

ii. Sampling data for manual labelling

A sample of approximately 3000 buildings from the 2026 April OS OpenMap Local building footprints was taken for manual labelling. Buildings were joined to the UPRNs they contain, and the number of UPRNs labelled as a flat was calculated per building.

The aim was to identify which buildings containing flats were blocks of flats. As such, the sampling population consisted of buildings in Great Britain containing at least two flats, as buildings with no flats are considered non-blocks by default.

We aimed to generate a labelled dataset that represents the diversity of buildings containing flats across Great Britain. We therefore selected a set of primary strata and secondary constraints to sample against which we assumed would give us a representative sample of the true population.

Train/validation and test sets of data for the model were created up front in the sampling process in order to train the model for real-world application.

The full sample was created in the following steps:

1. Building footprints were enriched with the following information:
    a. Building level attributes:
        i. Area: London; rest of England; Scotland; Wales
        ii. Number of flats in footprint: 2-6; 7-15; 16+
        iii. Proportion of UPRNs in the footprint that are flats: >80%, <=80% binary
    b. LSOA/ Data Zone level attributes:
        i. Rurality: urban/rural binary
        ii. IMD decile group: high deprivation (deciles 1-3), not high deprivation (4-10) binary
2. Buildings were grouped into 24 primary strata to stratify sampling on according to area; number of flats; and deprivation binary. The remaining binary attributes were applied as secondary constraints: urban/rural binary; and proportion of flats binary.
3. A sample was then built to meet the following conditions:
    a. Total sample size approximately equal to 3000 footprints. 75% (N~2250) of these were to form the training and validation set, and 25% (N~750) were to form the final hold-out test set. The training and validation set and the test sets were sampled separately using different methodologies.
4. To generate the training and validation sample:
    a. 50 buildings were taken from each of the 24 primary strata first, with secondary constraints applied evenly (a 50/50 split across each binary attribute). This "padding" ensures smaller strata are still represented in model training.
    b. The remaining buildings were then sampled proportionally to the total population across primary strata, with secondary constraints reflecting whole-population proportions.

c. The result is a sample that's close to whole-population proportions, but with extra representation for smaller groups.

5. We generated the hold-out test sample separately and entirely proportionally - i.e. both primary strata and secondary constraints reflect whole-population proportions with no padding.

The final dataset for manual labelling contained the full ~3000 building footprint sample (train & validation and test combined).

iii. Manual labelling

Through a previous labelling process, as well as stakeholder input, we generated a set of 24 different labels to use in manual labelling, representing the variety of archetypes we see in building footprints containing flats.

Labelling of the 3000 samples was performed by a group of three human labellers using the following information for each building:

- Count of UPRNs in the building footprint
- Count of UPRNs labelled as 'flats' in the building footprint
- Shape of the building footprint
- Satellite-, 3D-, and street-view of the building and surrounding area

Labellers were given descriptions and images of each identified archetype, and guidance and rules for labelling buildings consistently.

During labelling, buildings were given one of the 24 codes shown in the table below, which, in turn, are grouped into either 'block of flats', 'non-block', or 'excluded'. Labellers also assigned a confidence of either 1 (confident) or 0 (not confident) to their labels.

Table 1. Contains the different possible labels that could be given to building footprints during manual labelling of footprints for block of flats classification.

Label	Class	Description (abbreviated)
TB	Block of flats	Tower blocks and high-rises.
BF	Block of flats	Purpose-built block of flats including 1930s Art Deco

		blocks; deck-access estates; modern and new-builds; and social housing blocks.
BP	Block of flats	Same as above but the building footprint represents a small section of a bigger block. I.e. this represents a section of a BF with <= 6 flats in it.
IW	Block of flats	Industrial warehouse (or similar) conversions into flats.
TN	Block of flats	Traditional Scottish tenements.
TC	Block of flats	Large terrace to block of flat conversion.
SF	Block of flats	Sprawling mansion block e.g. a large old building converted into flats. These will usually be large and sprawling buildings with large garden(s) and separate entrances.
TT	Block of flats	A mixture of a block of flats and another built form combined into the same building footprint. E.g. Shopping centre with flats on one side; or terrace with a block of flats on one end.
AS	Block of flats	3+ storeys of flats above shops.
NH	Block of flats	Small number of flats across more than 2 storeys.
TS	Non-block	1-2 storeys of flats above shops.
WB	Non-block	A small wide block of flats. These are no more than 2 storeys tall (including ground floor). They can be purpose-built flats or conversions.
TE	Non-block	Terraced housing. These could contain a mix of flats and houses, or just flats.
TY	Non-block	Tyneside flats. Paired terraced houses designed specifically so that each building contains two flats.
BB	Non-block	Back-to-backs. Terraced houses built sharing back and sides with no rear outdoor space.
MT	Non-block	Multiple semi-detached building footprints merged into one.

MA	Non-block	Purpose-built tiny block of 1-2 flats.
TF	Non-block	Terraced front, flats at the back. These are terraces on one side with attached 1- or 2-storey flats with separate entrances at the back.
FC	Non-block	Old buildings, like farm houses, schools, or churches, converted into homes.
RA	Non-block	Retirement or sheltered housing with communal facilities and support services.
OF	Non-block	A building footprint that doesn't match any of the archetypes, is not a block, and contains domestic properties. E.g. semi-detached houses; active farms; etc.
CO	Excluded	Contains non-domestic properties only e.g. shops; industrial estate buildings, etc.
DE	Excluded	Derelict building.
UL	Excluded	Unable to label. This label is used either when the building type is unclear OR when there is a clear error in the number of UPRNs assigned to that footprint.

The 24 categories help to ensure a variety of built forms are captured in the labelled sample, and help to identify more specifically where the model is performing poorly.

The subsequent binary classification is ultimately based on how these different building types should be treated in the decision tree to assign a technology to, whether that be a communal solution versus a networked heat pump or individual solution.

Each labeller was given 30 samples which were independently labelled by another labeller to assess general alignment. Average agreement across the binary classification was 84%.

Manually labelled buildings which met the following criteria were **excluded** from the final training dataset:

- Buildings with a manual label confidence of 0

- Buildings which were labelled with 'UL' (unable to label), 'DE' (derelict), or 'CO' (non-domestic)
- Buildings with two labellers where labellers didn't agree at the binary level

The final dataset used for training contains 2604 buildings.

iv. Model training and evaluation

A random forest binary classifier model was trained to label building footprints as either blocks of flats or not. A random forest classifier was selected as it improved the initial F1 score compared to a baseline logistic regression model.

Features are generated per building from building footprint and UPRN geodata. See Annex Table 2 for a full list of features.

The training data was used for hyperparameter tuning and cross-validation with sklearn's HalvingRandomSearchCV function, with StratifiedKFold employed for cross-validation. Hyperparameters were tuned to optimise for a best F1 score on the positive class ('blocks of flats').

The resulting best hyperparameters were used to train a final Random Forest Classifier model, which was evaluated on the hold-out test set. A final F1 score for the positive class was produced (Table 2).

Table 2. Evaluation metrics for random forest binary classifier during hyperparameter tuning and cross-validation and testing.

Metric	Value
Train set best F1 score	0.71
Test set F1 score	0.71
Precision on positive class (block of flats)	0.64
Recall on positive class	0.79

The final model outputs both the predicted class and the probability of the class label for each building.

v. Model application

The trained model is applied to the subset of buildings containing any domestic UPRNs and at least one flat to classify them into 'block of flats' and 'non-blocks'. All buildings containing any domestic UPRNs with zero flats are assigned the label 'non-block'.

A corresponding boolean label is then assigned to each individual UPRN to indicate whether it is located within a block of flats or not, by using the class of the building footprint within which the UPRN is located.

vi. Caveats of block of flats classification methodology

- The 'ground truth' manual labelling is itself noisy. Manual labelling of buildings was subjective and complex and was conducted by multiple independent manual labellers. Building forms seen in the sample were extremely variable and as such, there is some disagreement in labelling between labellers. Buildings were labelled primarily according to visual inspection, using available visual information such as signage, number of bins and doorbells, and visual appearance of buildings. Images of the buildings were not standardised across labellers (i.e. labellers could freely use Google Maps satellite and 3D view from as many angles as deemed necessary for labelling) so there was also variation in the information used to make a judgement across labellers. Due to these factors, it's likely that some buildings have been erroneously mislabelled in the training dataset.

- Labellers were not given information about the spatial distribution of UPRNs within buildings and therefore had to make assumptions about this during labelling based on other available information. The result is that ambiguous archetypes, which tend to present with atypical storey counts and UPRN counts (e.g. the BP, NH, and TC labels), potentially are more often erroneously labelled than more typical archetypes like tower blocks (TB) and terraces (TE).

- Building footprints limit our classification capacity as they can contain multiple built forms and our model operates at building footprint level. This means that some individual UPRNs within a given building footprint may receive a binary classification that is inaccurate for them individually, but appropriate for other UPRNs within the same footprint. For example, our model treats TT types (a single building footprint containing a combination of a block of flats and

another built form) as a 'block of flats'. Future work may improve this and allow distinction of classifications within building footprints.

- Building footprints may represent a smaller section of a larger building, as in the case of BP types (representing a smaller section of a bigger block of flats). These sections may exhibit similar features to building types in a different binary class, and therefore may affect the accuracy of the model and the final classification of UPRNs.

vii. [Not currently implemented, in progress] Future improvements to block of flats classification model

We may undertake future work to improve the block of flats classification model, including by doing any of the following:

- Reviewing and, where required re-labelling, the training data for more consistent classification of building types.
- Engineering new features for the model, likely from new datasets that we have not previously used in feature engineering, to inject new information into the model. These may include features created from the land registry, the Points of Interest data, or census data, for example.
- Exploring instances of misclassification to identify patterns to exploit for model improvement.

c. Estimating outdoor space

i. Outdoor space estimation

Outdoor space per property is calculated using HM Land Registry polygons and OS OpenMap Local Buildings polygons. The methodology assumes that the area of a land parcel minus the total area of all buildings (or building fragments) on that land generates a good estimate of outdoor space for a property of interest. We outlined an earlier version of this methodology in more detail in our blog post here. Outdoor space is estimated via the following process:

1. An intersecting overlay operation is performed between land parcels and building footprints and the intersectional geometries are retained. This results in fragments of buildings that intersect with each land parcel.

2. Intersecting building fragments can include small slivers of building footprint that likely originate from differences in mapping, mapping errors, or boundary-touching errors. To address this, the area of both the intersecting fragments and the full building from which the fragment originates are calculated and a set of rules applied. Building intersections are kept if they meet one of the following conditions:

    a. The full building footprint area is <=30m²* and the building intersection is equal to at least 45% of the full building footprint. This assumes that small outbuildings will realistically never be split into more than two roughly equal parts. This condition aims to retain building intersections that could represent garages, sheds, or other outbuildings.

    b. The full building footprint area is >30m² and the building intersection area is at least 15m². This condition aims to retain building intersections that are large enough to represent a meaningful intersection of a main building.

    *30m² was chosen due to this being the approximate size of a small double garage building, so an intersection would need to include about half of the total footprint to be considered a valid intersection.

3. Land parcels are clipped to a variety of barriers. Barriers include: all roads; railways; public green space; woodland; rivers and bodies of water; and tidal boundaries. This aims to provide a more realistic estimate of outdoor space, for example in cases where multiple individual properties with separate gardens share the same land parcel due to a single freeholder (e.g. in some social housing).

4. After removal of unwanted building intersections and land parcel clipping, a difference overlay operation is performed between the clipped land parcels and the remaining building footprint fragments. The result is a geometry (or set of geometries) per land parcel which represents the outdoor space (i.e. the area not overlapping with a building footprint).

5. The maximum contiguous and the total outdoor space are calculated for each land parcel using the resulting geometries.

6. UPRNs are then spatially matched to the land parcel that they are located within.

7. In some instances, a UPRN is matched to multiple land parcels. This could occur, for example, due to the presence of vertical land parcels in the land registry data (i.e. land parcels that represent upper floors of buildings). In these cases, we are unable to identify which land parcel corresponds to the UPRN as neither the UPRN nor geospatial land parcel data is three-dimensional. Therefore, to be conservative, the land parcel with the smallest amount of total outdoor space is retained for those UPRNs.

ii. Caveats of estimating outdoor space methodology

- The outdoor space model provides no information about available outdoor space for heating technology installation. This may be especially relevant in the following cases:
    - Properties with outdoor space containing driveways or other outdoor features, like ponds or hedges, which may reduce the actual available outdoor space.
    - Social housing or other land rental arrangements. A single landowner may own the full extent of land on which multiple homes with separate private outdoor spaces are built. The estimated outdoor space size we calculate will actually represent the combined area of all outdoor space, clipped to any boundaries. This may result in very large outdoor space sizes for certain properties.
    - Farm houses. E.g. A farmhouse may have a separate garden area around it with physical boundaries (where a heating technology would realistically be installed), but be located within a larger extent of land (like farming fields) contained in the same land parcel. The estimated outdoor size will represent the full extent of all the farm's fields and garden combined. This may therefore overestimate available outdoor space.
- Currently we are unable to assign ownership of outdoor space to specific properties within flats and apartments (including those in converted houses) that share a single land parcel due to the lack of open access land ownership data. This can result in overestimating outdoor space, and in cases of large

apartment complexes, can result in very large outdoor space sizes for individual UPRNs.

- The Land Registry data is incomplete so there are properties in the final dataset missing outdoor space information. The Land Registry captures almost 90% of the total land mass in England and Wales as of 2019, and 63% of land mass in Scotland as of 2026. A subset of that data is then captured in the INSPIRE polygons used in this analysis.
- The OS OpenMap Local Building layer footprints are generalised and sometimes multiple buildings are combined into one footprint. The outdoor space size estimates are therefore likely to contain inaccuracies and may either over- or under-estimate outdoor space for any given building. Additional small mapping errors in either the building footprints or the land extents could exacerbate this.

3. Building-footprint level technology assignment

Property-level technology assignment is aggregated to the building footprint level. If a building footprint contains a mix of individual and communal solutions for different properties, the most communal solution will take precedence (see subsection 3.i. for more details).

We can currently only map suitability at building footprint level. While we can aggregate these building footprint polygons into larger clusters, we do not yet have the granular data required to map sub-building units. Building footprints might contain:

- One or multiple blocks of flats
- Individual houses or rows of terraced houses
- (Less common) a mixture of property types.

Throughout this section, the terms 'building' and 'building footprint' are used interchangeably; however, both refer to the building footprint. Similarly, the terms 'property' and 'UPRN' are used interchangeably, and a property always refers to 'UPRN'.

The methodology for identifying the technology for each building is as follows:

- The set of solutions for each building is identified, *i.e.* the set of solutions for all UPRNs within the same building.
- Where only 1 solution exists for a building, assign that solution as the most suitable technology for the building
- Where multiple solutions exist for a building, assign a single solution based on the combination of solutions in the set and the median outdoor space of properties within the building, as per the subsection below.

i. Identifying the most communal solution, where multiple solutions exist for a building footprint

In a building with multiple assigned technologies, i.e. different solutions for different UPRNs, the most communal solution prevails at the building footprint level. However,

this carries a caveat: while our pipeline identifies the most suitable technology for property groups, individual owners may still choose to decarbonise independently.

Opting for an individual solution does not preclude a property from joining a communal network later.

Our Clean Heat Neighbourhoods approach prioritises communal heat solutions because shared infrastructure relies on multi-property participation to succeed. Industry feedback suggests this approach improves the commerciality of installation and ongoing operational costs. Additionally, early householder insights indicate that shared infrastructure is often viewed favorably where outdoor space is limited.

1. If the set of solutions contains "Communal solution", assign "Communal solution".

Example: for buildings with a mix of communal solutions and individual solutions, the most communal solution will be the communal solution.

2. Else, if the set of solutions contains "Networked heat pump", assign "Networked heat pump".

Example: in a row of terraced houses, some properties might have a large outdoor space available (and so be suitable for an individual air source heat pump), while some others will have small or no outdoor space available (and hence assigned a networked heat pump solution). The most communal solution will be a networked heat pump solution for the whole row of terraces.

3. If the set of solutions contains "Individual heat pump" and unknown (due to missing outdoor space data) assign:
    - "Individual heat pump" if at least one property in the building footprint has outdoor space data available
    - "Individual or networked heat pump" otherwise

Example: if the outdoor space is known for at least one property in the building footprint and it is assigned an individual solution, an individual solution is assigned as most suitable for the whole building under the assumption that the remaining properties in the footprint are similar to the ones with available outdoor space. Otherwise, an 'Individual solution or networked heat pump solution' is assigned, creating a grey area.

4. Technology reassignment due to anchor load proximity

After buildings are assigned a technology type, those assigned networked heat pumps are reassigned to communal solutions if any part of the building footprint is within 50 metres (euclidean distance) of an anchor load property. Each building assigned a communal solution keeps a record of the reason for its assignment (block of flats or anchor load proximity), and the spatial aggregation step (in section 5) does not join buildings that have different underlying reasons for being assigned a communal solution.

To identify anchor load properties, we manually identified a list of building usage types from the Points of Interest dataset and the OS OpenMap Local Important Buildings layer which were likely to have regular, high heat demand and therefore act as suitable anchor load properties for communal heating systems. Anchor load properties were then mapped to their building footprints based on this list. See Annex Table 3 for the full list.

i. Caveats of anchor load properties methodology

- Anchor load properties are currently identified from building usage type only, with no consideration of size or actual heat demand. Some buildings identified as anchor loads may therefore be unsuitable in reality. To note that networked heat pumps are a subset of communal solutions and hence, the reassignment from networked heat pump to communal solution acts as a flag for a potential high heat demand commercial anchor property nearby.
- We have purposefully included a wide range of building uses as anchor properties. This is driven by uncertainty about the kinds of properties that may prove to be good anchor loads in emerging contexts. Future work may refine this identification.
- Due to incompleteness of the Points of Interest dataset, some buildings which may be suitable anchor load properties may not be included.

5. Spatial aggregation

i. Aggregating building footprints

After properties have been assigned technology type by the decision tree (section 2.a.); technology assignment is aggregated to building footprint level (section 3); and anchor property reassignment is completed (section 4), building footprints are aggregated into clusters via the process outlined below and in Fig. 3.

The purpose of this aggregation is to create clusters of buildings at what could be considered a neighbourhood level. For example:

- A group of buildings containing domestic properties which are all assigned networked ground-source heat pumps and are bordered by minor roads on three sides and a green space on the fourth side would be aggregated together into one cluster.
- A separate group of buildings located across the road which were also assigned networked ground-source heat pumps would be aggregated into a different cluster.

The result is a mosaic of small neighbourhood-level clusters of buildings suitable for the same technology type.

Aggregation process

1. Building footprint exterior edges are interpolated with additional points at a given spacing interval (default 1m). This is performed for all building footprints for the given area, both domestic and non-domestic.
2. Voronoi polygons are then generated from all of the subsequent points (i.e. those generated plus the original exterior vertices). These Voronoi polygons are dissolved by building ID to generate one Voronoi cell per building. The result is a full coverage of Voronoi polygons for the entire area of interest where one Voronoi cell corresponds to one building footprint. The methodology to generate these polygons has been adapted and extended from the methodology used in field-maps/edge-extender.
3. Voronoi cells containing non-domestic building footprints are removed. A difference overlay operation is then performed with the geometries of any

physical barriers (see below for more detail). The result is a set of Voronoi cells for buildings containing domestic properties only, with any physical barriers excised from the cells.

4. Domestic building footprints are then unioned with the remaining Voronoi cell fragments that they intersect with, while any non-intersecting cell fragments are discarded.

5. Buildings are then aggregated into clusters by dissolving touching polygons containing buildings labelled with the same technology type and, in the case of the 'communal solutions' technology type, with the same reason for assignment (a building being a block of flats versus a building located near an anchor load, see section 4 for more details).

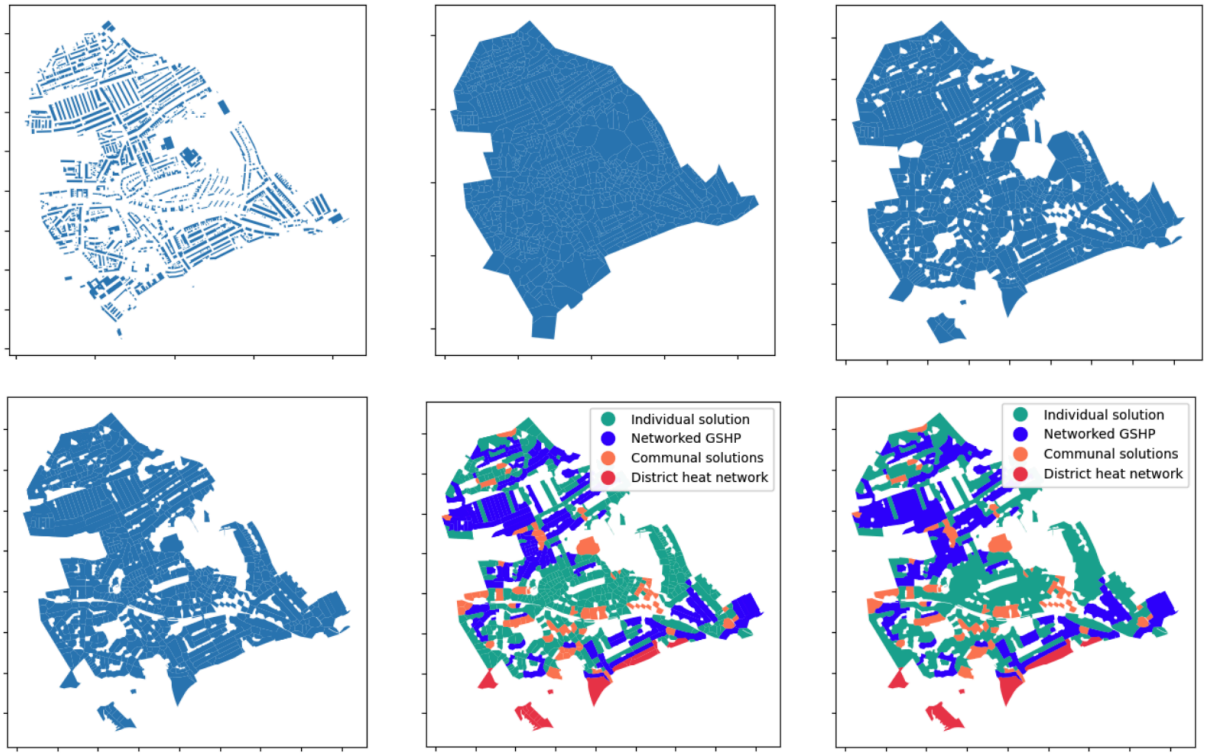


Figure 3. Shows the process of spatial aggregation for Stoke ward. **Top row from left to right:** all building footprints (domestic and non-domestic); Voronoi cells for all buildings; removal of cells containing non-domestic buildings. **Bottom row from left to right:** removal of environmental barriers; remaining cells joined to the assigned technology type; touching cells which have the same tech type dissolved to create final 'clusters'.

ii. Physical barriers

The included physical barriers are intended to represent features of the environment that enforce separation of clusters of households. These include:

- Roads of the following types: motorway, A road, B road, minor road
- Rivers, lakes, and other terrestrial water bodies
- Tidal water
- All green spaces found in the OS Open Greenspace dataset. See the classifications here.
- Railway tracks
- Woodland

Roads and railway tracks are provided as line strings in the OS datasets. These are converted to polygons by adding a 3.5 metre buffer across all road types and railway tracks, to represent the smallest width of a road.

iii. Caveats of aggregation methodology

- It is possible that some small outbuildings belonging to domestic properties are removed from the final clusters. This is due to the methodology used to distinguish domestic from non-domestic buildings, where any building footprint not containing a UPRN considered to be domestic is assumed to be non-domestic. In any case, these outbuildings are unlikely to represent true domestic buildings in which a household lives.
- The extent of a cluster does not necessarily extend only over land which belongs to the properties contained within the cluster. In other words, land covered by a cluster polygon may include public or other privately owned land.
- The list of physical barriers is conservative in the sense that it includes features that may not truly represent a barrier to neighbouring installations (e.g. minor roads or certain public greenspaces). This list will be further refined with stakeholders.

- Buildings which intersect with a Local Authority boundary may appear in different clusters in different Local Authorities. For example, a building that crosses a boundary between two Local Authorities may appear in a certain cluster in Local Authority A, but it would be assigned a different cluster in Local Authority B.

6. District heat network recommendations

Clusters in an area of district heat network potential may be most suitable for a district heat network connection when/if a district heat network is constructed. District heat network recommendations allow for reassigning the technology of eligible clusters into district heat network. Eligible clusters include clusters originally assigned **Communal solution** or **Networked Heat Pump** within a zone of high potential for district heat networks.

There are two types of district heat network recommendations:

- **District heat network potential** defaults to showing official advanced heat network zones - DESNZ in England and LHEES in Scotland.
- **City centre areas**: uses high-density areas as a proxy for where district heat networks could potentially be deployed.

To learn more about these, refer to section 7a).

This step concludes the assignment of the most suitable technology for each cluster. A logic trace of the decision is provided as a contextual feature.

7. Contextual features

Additional features are aggregated for the properties in each cluster to provide further information relevant to planning officers and the supply chain. The table below shows the features available and where they originate from. Unless otherwise described, these features are added by joining data from the relevant datasets, with no or light processing conducted on the data points themselves. Currently, there is missingness across features and there has been no computation conducted to fill missing data.

Table 2. Cluster-level contextual features included in the tool. Features, in most cases, are first identified at UPRN level and then aggregated to cluster level. Note: you can download data directly from the Local Heat Planning Tool which contains a data dictionary with a comprehensive list of the data points available.

Cluster level features	Origin
Logic trace	Result of Nesta's approach/modelling for technology assignment. See section 6.b for more information.
Number of UPRNs	OS Open UPRN
Distribution of tenure of properties (number and percentage)	Domestic EPC register
Distribution of energy performance ratings	Domestic EPC register
Distribution of property attachment type (e.g. detached, semi-detached, etc.)	Domestic EPC register and computation of flats (see section 2.*b.i* for flat computation methodology)
Number of properties with solar PV	Domestic EPC register
Number and percentage of properties missing solar PV data	Domestic EPC register
Median estimated energy consumption for 12 months (in kWh/m2)	Domestic EPC register
Median estimated outdoor space (m2)	Land registries of Scotland, England & Wales and OS OpenMap Local Buildings (see section 2.c.i for methodology)

In heat network zone	DESNZ advanced heat network zones and LHEES heat network zoning (see section 6.a.ii for methodology)
In city centre	Spatial signatures framework (see section 6.a.i for methodology)
Number of properties in listed buildings	Geospatial listed building datasets for England, Scotland, and Wales
Number and percentage of properties missing listed building data	Geospatial listed building datasets for England, Scotland, and Wales
Number of properties off gas	Off gas postcode register
Number and percentage of properties missing off gas flag	Off gas postcode register
Within 1500m of coastline	Coastline of Great Britain
Number of properties in a protected area	Conservation areas in England and Wales Scottish World Heritage sites

a. Labelling clusters in city centres or heat network zones

i. City centre labelling

Clusters are assigned a boolean label to identify whether they are located within an area considered to be a high-density 'city centre', i.e. located in an area with potential for a district heat network to be implemented. To identify areas considered to be a city centre in our modelling, we used the Spatial Signatures in Great Britain dataset from the Urban Grammar project. This is a geospatial classification dataset which identifies 16 distinct spatial signature types across Great Britain, where each signature type is characterised by a set of form and function variables. The signatures operate across a rural to urban spectrum.

We identified six spatial signature types as city centre types (Fig. 4). To identify spatial signature types which correspond to city centres, we used the types' pen portrait descriptions and qualitatively assessed that the 'urban' signature types matched our target profile - i.e., high population and building density with high likelihood of containing high heat demand anchor load properties.

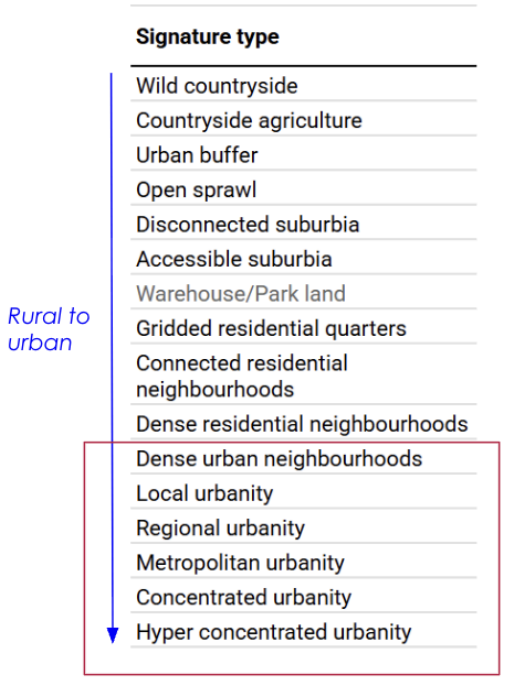


Figure 4. The 16 spatial signature types of Great Britain with selected city centre types indicated within the red box.

Clusters within or intersecting with the boundaries of any of the six city centre spatial signature types are considered to be in the city centre. All clusters that fall outside the boundaries, including those which do not intersect with any spatial signature type, are considered to be outside the city centre. In cases where a cluster intersects with multiple signature types, it is considered to be located in a city centre if one of the signature types is a city centre type.

ii. Heat network zone labelling

Clusters are assigned a boolean label to identify whether they are located within or intersecting with an existing or planned heat network zone or not. Heat network zone boundaries for a local authority district are taken from the DESNZ Advanced Heat Network Zoning in England and the LHEES heat network zoning in Scotland, where available, and clusters within or intersecting with the boundaries are considered to be heat network zone candidates. All clusters that fall outside the boundaries are not considered to be heat network zone candidates.

iii. Caveats of city centre and heat network zone labelling

- The spatial signatures dataset is not complete with some small areas missing signature information. These areas are currently considered to be outside the city centre.

- The DESNZ heat network zones are outputs from the Heat Network Zoning Pilot Programme. The Pilot was undertaken prior to full details of the Heat Network Zoning policy being available. Therefore, the zones are likely to change and shall not be relied upon for any business decisions. (Text taken and adapted from page 2 of the Plymouth Heat Network Zoning Zone Opportunity Report, 2025).
- We have only collected LHEES heat network zones for two areas in Scotland (Midlothian and Glasgow local authority councils). Data for other local authorities will be added to the tool in the next phase of work.

b. Logic trace

The logic trace is a piece of text that we add to each cluster to provide information about why a specific technology has been assigned as most suitable. Below you can see the logic trace for each individual possible case.

CASE 1: Cluster assigned communal solution due to being within 50m of anchor load. Cluster is in a district heat network recommendation area.

This cluster:

- *is in an area of 'district heat network' potential,*
- *contains homes that have no or little outdoor space (up to 30m2),*
- *and it is within 50m of an anchor load (e.g. a hospital or school),*

so homes are most suitable for a 'district heat network' connection when/if a district heat network is constructed.

A 'communal solution' could also be considered as an alternative solution because the homes have no or little outdoor space (up to 30m2) and it is within 50m of an anchor load (e.g. a hospital or school). This 'communal solution' could be integrated into a 'district heat network' in the future.

There might also be one or multiple blocks of flats in the cluster.

CASE 2: Cluster assigned communal solution due to being within 50m of anchor load.

This cluster is assigned 'communal solution' because the homes have no or little outdoor space (up to 30m2) and it is within 50m of an anchor load (e.g. a hospital or school).

There might also be one or multiple blocks of flats in the cluster.

CASE 3: Cluster is assigned communal solution because buildings are blocks of flats. Cluster is in a district heat network recommendation area.

This cluster is in an area of 'district heat network' potential and it contains one or multiple blocks of flats, so homes are most suitable for a 'district heat network' connection when/if a district heat network is constructed.

A 'communal solution' could also be considered as an alternative solution, because it contains one or multiple blocks of flats. This 'communal solution' could be integrated into a 'district heat network' in the future.

CASE 4: Cluster is assigned communal solution because buildings are blocks of flats.

This cluster is assigned 'communal solution' because it contains one or multiple blocks of flats.

CASE 5: Cluster is assigned networked heat pump. Cluster is in a district heat network recommendation area.

This cluster is in an area of 'district heat network' potential and the homes have no or little outdoor space (up to 30m2), so homes are most suitable for a 'district heat network' connection when/if a district heat network is constructed.

A 'networked heat pump' could also be considered as an alternative solution,

because at least one property within each building has little or no outdoor space (<30m2).

CASE 6: Cluster is assigned networked heat pump.

Cluster is assigned 'networked heat pump' because at least one property within each building has little or no outdoor space (up to 30m2)

CASE 7: Outdoor space unknown/missing. Cluster is in a district heat network recommendation area.

Outdoor space is unknown for all properties in buildings within this cluster, so the model is unable to assign a technology group.

> ***If, in reality, there is little (up to 30m2) contiguous outdoor space:***
>
> *This cluster is in an area of 'district heat network' potential and if homes have no or little outdoor space (up to 30m2), then homes would be most suitable for a district heat network connection when/if a 'district heat network' is constructed.*
>
> *A 'networked heat pump' could also be considered as an alternative solution, because at least one property within these buildings has little or no outdoor space (up to 30m2).*
>
> ***If, in reality, there is sufficient outdoor space for each property (above 30m2):***
>
> *'Individual solutions' will be the most suitable option for properties in this cluster because outdoor space is above 30m2 for all properties.*
> *The cluster is also in an area of 'district heat network' potential. If a district heat network is built, connection to the network could be offered.*

CASE 8: Outdoor space unknown/missing.

Outdoor space is unknown for all properties in buildings within this cluster, so the model is unable to assign a technology group.

- *If there is little (up to 30m2) contiguous outdoor space, a 'networked heat pump' solution is most suitable.*
- *If there is sufficient outdoor space for each property (above 30m2) an 'individual solution' will be most suitable.*

CASE 9: Cluster is assigned Individual solution. Cluster is in a district heat network recommendation area.

This cluster is assigned 'individual solution' because outdoor space is above 30m2 for all properties in this cluster.

The cluster is also in an area of 'district heat network' potential. If a district heat network is built, connection to the network could be offered.

CASE 10: Cluster is assigned Individual solution.

This cluster is assigned 'individual solution' because outdoor space is above 30m2 for all properties in this cluster.

Annex

Text 1. Initial ChatGPT prompt.

Which of the following points of interest, if any, is likely to be located within a building that does not contain any residential properties, including flats and apartments. The answer should be UK-specific. Please exclude any points which would not be located within a building. Please output the answer in csv format in alphabetical order with no duplicates. The categories are in the attached csv.

Text 2. Final ChatGPT prompt.

Please conduct the filtering again according to these refined criteria:

✅ *Include:*

Large commercial (e.g. superstores, shopping centres)

Industrial and logistics (factories, warehouses, mills, plants, refineries, data centres, depots, etc.)

Raw material refinement (e.g. oil and gas fields, steel manufacturing, coal plants etc.)

Manufacturing facilities Utilities and infrastructure (power stations, water treatment, substations, etc.)

Large public/institutional (hospitals, schools, universities, police/fire stations, government offices, courts, etc.)

Large recreational (stadiums, arenas, theatres, cinemas, museums, swimming pools, concert venues etc.)

Transport (airports, stations, depots)

Industrial waste disposal

Historic or cultural buildings which are unlikely to include residential (e.g. cathedrals, castles)

Places housing or exercising large numbers of live animals, except animal kennels or animal sitting businesses (e.g. zoos, equestrian centres)

🚫 *Exclude:*

Small or medium-scale retail, food & drink, or service businesses (shops, cafes, pubs, salons, gyms, restaurants, etc.)

POIs commonly located on high streets or beneath flats.

Table 1. Important Building Classes in the OS OpenMap Local Important Building layer

Important Building Classes
Port Consisting of Docks and Nautical Berthing
Fire Station
Hospital
Non State Secondary Education
Secondary Education
Higher or University Education
Primary Education
Place Of Worship
Medical Care Accommodation
Museum
Special Needs Education
Further Education
Non State Primary Education
Coach Station
Police Station

Sports And Leisure Centre
Vehicular Ferry Terminal
Hospice
Bus Station
Road User Services
Passenger Ferry Terminal

Table 2. List of features engineered from UPRN and building footprint data per building footprint for the block of flats random forest binary classifier

Model features
Count of domestic UPRNs per building
Count of flats per building
Count of total UPRNs per building
Building area (m2)
Building perimeter (m)
Proportion of UPRNs which are flats
Proportion of domestic UPRNs which are flats
UPRNs per building area (m2)
Area of concave hull of all UPRNs in building (m2)
Count of UPRNs per hull area (m2)
Count of flats per hull area (m2)
Average count of UPRNs sharing the same coordinates per building

Standard deviation of the count of UPRNs sharing the same coordinates per building
Maximum count of UPRNs sharing the same coordinates per building
Count of flats in nearest neighbouring building
Distance in metres of nearest neighbouring building

Table 3. List of building usage types assumed to be suitable anchor load properties and their dataset of origin.

Dataset of origin	Building usage type
OS OpenMap Local Important Building layer and UK Points of Interest	Secondary education
	Sports and leisure centre
	Hospital
	Higher or university education
	Airport
OS OpenMap Local Important Building layer only	Primary education
	Museum
	Library
	Further education
	Fire station
	Special needs education
	Medical care accommodation
	Non-state primary education
	Non-state secondary education
	Art gallery

	Police station
	Hospice
UK Points of Interest only	Central government office
	Government services
	Local and state government offices
	Public service and government
	Swimming pool
	Shopping centre
	Department store
	Water park
	University housing
	Superstore
	Airport terminal
	Jail and prison
	Ice skating rink