Getting Proxies

The following creates the data required to predict EPC ratings, estimate solar PV output and determine heat pump capacity. The final output from the process outlined in this document will be a series of *.geojson* files, while another set of files will be encoded and saved as *.csv* for model training.

Data

For more details on the selected columns, see Appendix B: OS Data.

Ordinance Survey Maps

Ordinance Survey (OS) is a national mapping agency for Great Britain which produces digital map data. They have some open-source data, but for our purposes we need three licensed data products:

1. OS MasterMap Topography Layer which provides the building footprints
2. OS MasterMap Building Height Attribute which provides the heights of each building
3. AddressBase Premium which provides the address data for each house

We received 58 West Midlands tiles listed in *osmap_tiles.txt* and Appendix A: West Midlands Grid Tiles.

Merging Ordinance Survey Data

The files provided are all named according to the OS grid system, which increases by two numbers for each additional level of granularity. Each file has data for a 10km-by-10km space. If you are unfamiliar with the grid system, see Appendix A.1: How to read Grid System.

Method

1. Set 'ROOT_DIR' in the file to the main folder containing the following subfolders:
    a. 'building_height': building height data in .csv
    b. 'landbaseprem': AddressBase Premium data in *.gml*
    c. 'topology': OS Master Map Topographic Layer data in *.gml*
        i. The data we received was in *.gfs*. Make sure that the file is in *.gml* because the driver for the following method cannot read files in *.gfs*.
2. Filter OS Master Map Topographic Layer
    a. Select the layer 'TopographicArea', which contains all the polygons
    b. Filter for the 'featureCode' = 10021, which represents the building area
    c. Remove all columns that are not necessary to speed up later processes
3. Set the projection for the AddressBase Premium to the project coordinates (EPSG:27700), so it will map the coordinates correctly
4. Join the building height and shapefiles on their 'OS_TOPO_TOID', and then with AddressBase on the point location in a polygon.
    a. Discard any polygons which are not assigned to an address.
5. All outputs are saved in a temporary folder within 'ROOT_DIR', with each file named with its grid reference in *.gml*

The final output will contain the following: UPRN, address, building heights, area per floor, number of floors and polygon.

Mapping Additional Data

We need to map additional data to each house to use as proxies from:

- ONS UPRN Directory (August 2022)
  - *ONSUD_AUG_2022_WM.csv* contains all data for the West Midlands
  - Added columns: UPRN, LSOA code, MSOA code, Local Authority Code
  - Updated every six weeks on Open Geography Portal
- Sub-regional fuel poverty data 2022
  - Added columns: Number of households, number of households in fuel poverty, proportion of households in fuel poverty
  - Updated every year
- Lower and Middle Super Output Areas electricity consumption
  - Added columns: total, mean and median consumption

Final Output

The .csv file will not have the geometry column.

Column Name	Column	Type	Source
uprn	UPRN	Int	AddressBase Premium
calculatedAreaValue	Area per floor (m^2)	Float	OSMap Topology
AbsHMin	lowest point extracted from the DTM within the footprint of the building as represented in OS MasterMap Topography Layer	Float	OSMap Building Height Attributes
AbsH2	Lowest point where the roof intersects the alignment of the external vertical walls of the principal part of the building	Float	OSMap Building Height Attributes
AbsHMax	Absolute Height Maximum – highest point of building	Float	OSMap Building Height Attributes
RelH2	RelH2 = AbsH2 – AbsHMin (height until roof)	Float	OSMap Building Height Attributes
RelHMax	RelHMax = AbsHMax – AbsHMin (total building height)	Float	OSMap Building Height Attributes
postcode	Postcode	String	AddressBase Premium
geometry	Building footprint	Polygon	OSMap Topology
lsoa_code	LSOA code	String	ONS UPRN Directory
msoa_code	MSOA code	String	ONS UPRN Directory

local-authority	Local authority code	String	ONS UPRN Directory
constituency	Parliamentary constituency in which the building is located	Encoded Int	EPC database
num_households	Total number of households in LSOA	Int	Sub-regional fuel poverty data
num_households_fuel_poverty	Number of households in fuel poverty	Int	Sub-regional fuel poverty data
prop_housesholds_fuel_poor	Proportion of households in fuel poverty	Float	Sub-regional fuel poverty data
total_consumption	Total energy consumption within the LSOA code	Float	LSOA domestic electricity consumption data
mean_consumption	Mean energy consumption within the LSOA code	Float	LSOA domestic electricity consumption data
median_consumption	Median energy consumption within the LSOA code	Float	LSOA domestic electricity consumption data

Appendix

Appendix A: West Midlands Grid Tiles

SJ8500	SO9080	SP0090	SP1090	SP2570
SJ9000	SO9085	SP0095	SP1095	SP2575
SJ9500	SO9090	SP0570	SP1570	SP2580
SK0000	SO9095	SP0575	SP1575	SP2585
SK0005	SO9575	SP0580	SP1580	SP3070
SK0500	SO9580	SP0585	SP1585	SP3075
SK0505	SO9585	SP0590	SP1590	SP3080
SK1000	SO9590	SP0595	SP1595	SP3085
SO8580	SO9595	SP1070	SP2070	SP3575
SO8585	SP0075	SP1075	SP2075	SP3580
SO8590	SP0080	SP1080	SP2080	
SO8595	SP0085	SP1085	SP2085	

Appendix A.1: How to read Grid System

SJ9050/SJ95SW: Each square in the grid represents a 100x100km space

SJ9050/SJ95SW: Each square is further divided into 10kmx10km tiles. The first number represents its location on the x-axis and the second number is the y-axis

SJ09									SJ99
									SJ95
SJ01									
SJ00	SJ10								SJ90

SJ9050/SJ95SW: Each square is subdivided into 5kmx5km tiles. The OS data uses 5kmx5km tiles which they denote from the bottom left corner of the tile while Defra uses 1kmx1km tiles.

NW	NE
SW	SE

SJ9055	SJ9555
SJ9050	SJ9550

Appendix B: OS Data

The following tables show potentially useful columns and how I used them. The 'Keep?' column indicates the columns which were held in the final output.

AddressBasePremium[1]

fieldname	description	Keep?	Notes
gml_id	Unique identifier	no	'osgb' prefix removed to keep column as int type
changeType	Type of record change	no	Keep if you want to compare between data updates
lastUpdateDate	The date on which any attribute on this record was last changed.	no	
entryDate	The date on which the record was entered into the Local Authority database.	no	
uprn	Unique property reference number	yes	

[1] Other fieldnames were removed because they were repeated information or not currently useful.

country	The country in which a record can be found.	no	'E' = England, 'W' = Wales, etc.
multiOccCount	This is a count of all of the child UPRNs for this record where a parent-child relationship exists.	no	
buildingNumber	The building number is a number given to a single building or a small group of buildings, thus identifying it from its neighbours	No	
postcode		yes	
endDate	The date on which the record ceased to exist	No	Filtered out all non-existing buildings
level	Detail on the vertical position of the property if known and provided by the Local Authority Custodian.	No	
locality	A locality defines an area or geographical identifier within a town, village or hamlet	No	
townName		No	
administrativeArea	Local Highway Authority name for the area this record exists within.	No	
localCustodianCode	Unique identifier of the Local Authority Custodian responsible for the maintenance of this record.	No	
postTown	The town or city in which the Royal Mail sorting office is located which services this record. There may be more than one, possibly several, sorting offices in a town or city.	No	
thoroughfare	A thoroughfare in AddressBase is fundamentally a road, track or named access route on which there are Royal Mail delivery points, for example, HIGH STREET.	No	

OSMap Topology [TopographicArea]

fieldname	description	Keep?	Notes
fid	Unique identifier that maps to OSMap Building Height Attribute	no	'osgb' prefix removed to keep column as type int
versionDate	The date on which this version of the feature became the current version. This is the date on which the feature was changed in the database, and is not the date of any associated real-world change.	no	

theme	A theme that the feature belongs to	no	Descriptive version of other fieldnames used for filtering
calculatedAreaValue	This is the calculated area of a polygon feature in square metres	yes	
changeDate	The date a change was made to the feature by an editor	no	
reasonForChange	The reason for a change made to a feature.	no	
featureCode	Topographic features have a numerical feature code (a five-digit integer) assigned to each feature.	no	Used '10021' = Building (Area) to filter

OSMap Building Height Attributes

fieldname	truename	description	Keep?	Notes
Field_1	OS_Topo_ TOID	Unique identifier that maps to OSMap Topography	no	
Field_2	OS_Topo_TOI D_ Version		no	
Field_3	BHA_Proces sDate		no	
Field_4	Tile Ref		no	
Field_5	AbsH Min	lowest point extracted from the DTM within the footprint of the building as represented in OS MasterMap Topography Layer	yes	
Field_6	Abs H2	Lowest point where the roof intersects the alignment of the external vertical walls of the principal part of the building	yes	
Field_7	AbsH Max	Absolute Height Maximum – highest point of building	yes	
Field_8	Rel H2	RelH2 = AbsH2 – AbsHMin (height until roof)	Yes	
Field_9	RelH Max	RelHMax = AbsHMax – AbsHMin (total building height)	yes	
Field_10	BHA_ Conf		no	

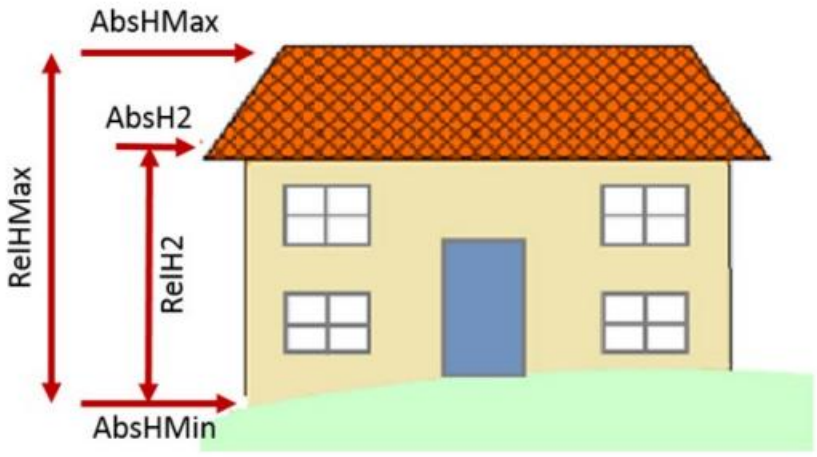

Figure 1 Illustration of the two relative and three absolute height values.