



This is a repository copy of *Automated classification metrics for energy modelling of residential buildings in the UK with open algorithms.*

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/129506/>

Version: Supplemental Material

---

**Article:**

Beck, A., Long, G., Boyd, D.S. et al. (5 more authors) (2020) Automated classification metrics for energy modelling of residential buildings in the UK with open algorithms. *Environment and Planning B: Urban Analytics and City Science*, 47 (1). pp. 45-64. ISSN 0265-8135

<https://doi.org/10.1177/2399808318762436>

---

Beck A, Long G, Boyd DS, et al. Automated classification metrics for energy modelling of residential buildings in the UK with open algorithms. *Environment and Planning B: Urban Analytics and City Science*. 2020;47(1):45-64. Copyright © 2018 The Author(s). DOI: <https://doi.org/10.1177/2399808318762436>. Article available under the terms of the CC-BY-NC-ND licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

---

## Appendix A - UK data sources

### *MasterMap*

Ordnance Survey MasterMap\* forms a core digital map product created by Great Britain's National Mapping Agency, the Ordnance Survey. MasterMap provides a high-level of accuracy and acts as a detailed source of high quality geographic reference data. For example, in urban areas, the Topography Layer is designed for use at a scale of 1:1250, and is stated to have a nominal positional accuracy of 1m (95% confidence interval). In rural areas, the accuracy is typically 2.5m (95% confidence interval) and is designed for a 1:2500 scale. MasterMap topographic data contains unique identifiers for addressable features called a TOpographic IDentifer (TOID). The TOID remains unchanged throughout the life of the feature and, as a unique identifier, may be linked across datasets and updated through Change Only Updates. At the time of writing, OSGB MasterMap data are available under various commercial, research-only and public sector (via the Public Sector Mapping Agreement) licences depending on the end-use.

### *AddressBase Plus*

In the UK energy analysts often use Local Land and Property Gazetteer (LLPG) data, maintained by local authorities as a basis for their housing stock analyses. The LLPG data are now exposed through the AddressBase products, provided by GeoPlace. AddressBase allows the techniques to be applied to different UK cities and respond to data changes more effectively (by applying epoch-based Change Only Updates).

GeoPlace is a public sector limited liability partnership between the Local Government Association (LGA) and OSGB. GeoPlace's role in addressing is to work with local authorities to create and maintain the national address gazetteer infrastructure, providing the definitive source of publicly-owned spatial address data. The national address gazetteer infrastructure is the data storage and set of processes bringing together the existing local authority sourced addressing datasets together with OSGB, Valuation Office Agency (VOA)

---

\*<https://www.ordnancesurvey.co.uk/business-and-government/products/mastermap-products.html>

---

and Royal Mail data. The national address gazetteer is the single source from which the AddressBase products are developed. The products include the Unique Property Reference Number (UPRN) to provide accurate identification of properties across England, Wales and Scotland.

The data are made available by OSGB through the AddressBase range of products to the public sector and beyond. At the time of writing, AddressBase Plus data are available under various commercial, research-only and public sector (via the Public Sector Mapping Agreement) licences depending on the end-use.

The AddressBase range contains three products (AddressBase, AddressBase Plus and AddressBase Premium).

This project uses AddressBase Plus<sup>†</sup> which:

- merges two address datasets together (PAF and Local Authority) to provide the best available view of addresses currently defined by Local Authorities, giving many advantages over AddressBase.
- lets you link additional information to a single address, place it on a map, and carry out spatial analysis that enables improved business practices. GeoPlace argue that further benefits include:
- A more detailed classification – allowing a better understanding of the type (e.g. Domestic, Commercial or Mixed) and function of a property (e.g. Bank or Restaurant)
- Local Authority addresses not contained within PAF – giving a more complete picture of the current addresses and properties (assuming they are in scope (see below))
- Cross reference to OSGB MasterMap TOIDs – allowing simple matching to OSGB MasterMap Address Layer 2, Integrated Transport Network or Topography Layer
- Spatial coordinates
- Unique Property Reference Number (UPRN) – which provides the ability to cross reference data with other organisations, and maintain data integrity.

Figure S3 shows an example of MasterMap and AddressBase Plus attributes. These include the coordinates, RPC = Representative Point Code which reflects the positional accuracy, UPRN = Unique Property Reference Number, OS\_TOPO\_TOID = OS Topographic TOID, STATE = a code identifying the state

---

<sup>†</sup><https://www.ordnancesurvey.co.uk/business-and-government/products/addressbase-plus.html>

---

of the property, `POSTAL_ADDRESS` = a code indicating whether the address is postal or not. Further information can be found in the technical documentation of AddressBase Plus.

Identifying addressable buildings is not a simple process of joining buildings in MasterMap with records in AddressBase Plus. This is primarily due to the multiple views of addresses taken within AddressBase Plus. AddressBase Plus represents data which have been integrated from different organisations. Each organisation has a different operational view of what an address is. For Royal Mail an address is a place where mail is delivered (a delivery point is related to a postal address). For a council an address is associated with an asset which requires its management. As a result the dataset contains records that do not represent either houses or workplaces and need filtering out.

## Appendix B - PostGIS topology details

A database topology is required by this work. Different database environments store this information in different ways. Within PostGIS (as adopted here) topology is managed in a dedicated schema. Four tables are created within this topology schema:

- `node`: a table of nodes that represent points and/or edge intersections or edge start/end points
- `edge_data`: a table of edges or polylines (all edges start and end at a node and record what face exists on their left and right side)
- `face`: a table of faces or polygons (there is always at least 1 entry in this table that represented the universal face with a `face_id` of 0)
- `relation`: a table that expresses the between nodes, edges and faces in the topological model that exist in the materialised physical model.

Topology was generated for the geometry of the set of addressable buildings in PostGIS. It should be borne in mind that the MasterMap product is derived from topologically aware back-end databases. So, here we are simply re-creating a topological data structure over a dataset which was derived from an intrinsically topological source. Hence, in this case study no cleaning was required to make a valid topology.

Calculation of the external wall ratio requires the use of data in the `edge_data` and `relation` tables. The `edge_data` table describes the physical relationships between the buildings and the `relation` table describes whether the relationship actually exists in the real world. From the set of edges (or walls) which are used to construct building footprints, some will be referenced to the universal face. In this context, the universal face represents an external or non-shared wall. As PostGIS has a hierarchical approach to constructing topology, this relationship is encoded in the `right_face` field.