Article:

Beck, A., Long, G., Boyd, D.S. et al. (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

Automated classification metrics for energy modelling of residential buildings in the UK with *open* algorithms

Journal Title
XX(X):1–26

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Abstract

Estimating residential building energy use across large spatial extents is vital for identifying and testing effective strategies to reduce carbon emissions and improve urban sustainability. This task is underpinned by the availability of accurate models of building stock from which appropriate parameters may be extracted. For example, the form of a building, such as whether it is detached, semi-detached, terraced etc and its shape may be used as part of a typology for defining its likely energy use. When these details are combined with information on building construction materials or glazing ratio, it can be used to infer the heat transfer characteristics of different properties. However, these data are not readily available for energy modelling or urban simulation. Although this is not a problem when the geographic scope corresponds to a small area and can be hand-collected, such manual approaches cannot be easily applied at the city or national scale. In this paper, we demonstrate an approach that can automatically extract this information at the city scale using off-the-shelf products supplied by a National Mapping Agency. We present two novel techniques to create this knowledge directly from input geometry. The first technique is used to identify built form based upon the physical relationships between buildings. The second technique is used to determine a more refined internal/external wall measurement and ratio. The second technique has greater metric accuracy and can also be used to address problems identified in extracting the built form. A case study is presented for the City of Nottingham in the United Kingdom using two data products provided by the Ordnance Survey of Great Britain (OSGB): MasterMap and AddressBase. This is followed by a discussion of a new categorisation approach for housing form for urban energy assessment.

Introduction

In the UK, the building stock imposes a major demand on the nation's energy supply with domestic buildings being responsible for 29% of use (estimated in 2015)(BEIS, 2017). Improving their efficiency is vital to reducing the country's carbon emissions. Buildings are assessed and assigned a rating as part of an Energy Performance Certificate (EPC). These ratings are derived from the Building Research Establishment's Domestic Energy Model (BREDEM). BREDEM was originally developed in the 1980s (updated in 2012) and is the most widely adopted and validated model for the calculation of domestic (space heating, hot water, lighting and electrical appliance) energy use in the UK (Hendersen and Hart, 2015). At the core of BREDEM is a thermal energy balance calculation, accounting for both thermal gains (e.g. solar and internal equipment) and losses (e.g. fabric and infiltration). This calculation requires many parameters (e.g. areas, U-values, ventilation characteristics, heating systems, and dwelling dimensions) which are time-consuming and expensive to collect (Chapman, 1994: 83).

Although it is relatively simple to manually gather property details of a small number of buildings for the purposes of energy modelling; it is difficult to apply this to larger scales. An automated approach to quantifying building stock is required to support energy conservation policy formation and decision making at the city or national scale. This is particularly relevant given the Foresight Sustainable Energy and the Built Environment Project resulting from the UK Government's Energy Review which aims to investigate the UK's transition to secure, sustainable, low-carbon energy systems over the next fifty years (Science, 2008: 8)

To address the problem of quantifying building stock at large scales, abstractions are used to create a representative typology. Every model is an abstraction of reality and one concept that guides this is the quality and availability of data to support it. Note, in this work we consider the term modelling as referring to the capture of the description of an object. Simulations may be undertaken on the resulting model in order to establish its performance. Hence, as data availability and granularity change, the model can evolve to more closely reflect reality. This is exemplified by Chapman (1994) who developed an influential approach that considered built form (detached, semi-detached, end-terraced, mid-terraced, detached bungalows, semi-detached bungalows) and construction date bands (pre-1900, 1900-19, 1919-44, 1945-64, 1965-75 and post-1976) as proxies for the BREDEM model. This approach produces 36 combinations of age and built form. By addition of a building height, these typologies can be used to infer such properties as heating volume, building material, roof pitch and number of floors. For example, Chapman (1994: 83) highlights how UK houses built in 1960 often have room heights between 2.3m and 2.4m, whereas those built between 1900 and 1930 range from 2.5m to 2.7m. The underlying assertion is that the 36 combinations of built form and age are representative of the characteristics of the UK building stock (in terms of floor plan, glazing, construction and insulation, for example). Rylatt et al. (2001) proposes six main classes of built form which differ from Chapman:

1. detached;
2. semi-detached;
3. end terrace;
4. mid terrace;
5. mid terrace with unheated connecting passageway; and
6. flat.

and nine age groups that also differ from Chapman:

1. pre 1900;
2. 1900-1929;
3. 1930-1949;
4. 1950-1965;
5. 1966-1976;
6. 1977-1981;
7. 1982-1990;
8. 1991-1994;
9. post 1994.

Their justification for this age categorisation is its alignment with major changes in construction regulations in the UK, which in turn correspond to particular building element specifications (e.g. U-values of walls and roofs materials, water heating system information and ventilation requirements) (Rylatt et al., 2001). This more precise range of age groups presents a stronger base for a building typology from which thermal characteristics can be extrapolated. The built form drives the volumetric stereotypes which are refined with the age categorisation which also drives the thermal attributes. Hence, these abstractions provide the potential to apply the BREDEM model to large areas. The challenge to achieving this is based on the ability to collect, or improve on the collection of, the abstracted calibration data. As such, we are interested in the collection of built form and construction date or age. This paper is principally concerned with the former although potential methods for the latter are discussed. Note that detailed simulation of the energy performance of buildings is not a direct focus of the paper. The work described here relates to the identification of specific built forms using typical UK spatial data sources, and how these results could be employed in the extrapolation of energy simulation results to urban housing stocks.

What follows is a summary of UK based built form extraction techniques using data from OSGB. the National Mapping Agency for Great Britain and the producer of commonly used topographic products for many urban analysis applications. After the release of precise and accurate 2D digital mapping for the UK from OSGB, a number of approaches were developed to infer built form directly from the geometry itself (Baker and Rylatt, 2008; Holtier et al., 2000; Hussain et al., 2012; Orford and Radcliffe, 2007; Rylatt et al., 2001). Such applications have evolved in line with the changing characteristics of the OSGB data product families from Land-Line Plus to the current MasterMap suite. For example, Rylatt et al. (2001) create a closed addressable building polygon by combining the vector line data from Land_line Plus with vector point data from Address Point (a product based upon the Postcode Address File (PAF) from Royal Mail). The point data are used as a seed from which an algorithm determines the lines that enclose the point. These lines are then turned into a closed polygon, which the analysis subsequently uses. This approach is unnecessary in the topographically structured MasterMap product and Rylatt et al. (2001) notes that this allows the extraction of accurate metrics on a per building basis (e.g. perimeters and shared walls). This means that some of the ‘dimensional data of the dwellings’ desired by Chapman (1994: 83) is potentially available at a national scale. Rylatt et al. (2001) also identify that automatic determination of built form purely using spatial reasoning is much more problematic and work on this aspect is not yet well-advanced.

Geometrical spatial reasoning was in part addressed by Orford and Radcliffe (2007: 211–212). They used OSGB MasterMap and Address Layer to provide dwelling type information for individual residential addresses and also demonstrated the efficacy of using addresses for identifying residential buildings (a technique that has also been undertaken in subsequent studies such as in Orford (2010) and Mavrogianni et al. (2012)). Bespoke scripts were developed that identified first and second order relationships based upon whether a building polygon touched other building polygons or not. From these relationships, a framework was implemented that could classify the building into detached, semi-detached, mid-terrace and end-terrace (Orford and Radcliffe, 2007: 211–212). Hussain et al. (2012) used the same datasets to extract building type into detached, semi-detached, end-terraced, mid-terraced and 'complex' classes. Likewise, bespoke scripts were used.

Orford and Radcliffe (2007) also describe issues of classification in detail. While urban analysis, social science research and Government policy refer to the built form of a residential property and imply that it is a homogeneous entity for statistical derivation, the concepts are not formally defined. As we shall see in this paper, such concepts may also not be supported in the data itself. Hence, there can be conceptual agreement that the residential building stock can be classified in a manner that reflects national and regional building vernaculars, but it is not demonstrated that the informal categorisation process reflects this diversity, nor the proxied building attributes (Orford and Radcliffe, 2007: 209). This lies at the heart of any abstraction process - how much it reflects reality and how important this is to the model. The use of high quality digital OSGB data products means that in addition to identifying a built form generalisation, it is possible to extract accurate metrics about the proportion of shared walls, externally exposed walls and building volumes on a building by building basis.

Possibly the broadest analysis of the UK housing stock has been undertaken by Steadman et al. (2009). Using the Virtual London model (derived from OSGB MasterMap and a 1m resolution LiDAR elevation model) the authors consider the effect of different volume and wall area metrics and the implications they have for occupation activities. They also critically question the quality of MasterMap and that the 2-dimensional concepts involved in map creation may not be compatible with building or premise representation from other domains (Steadman et al., 2009: 464 - 466). Conceptual issues surrounding the definition of buildings or premises is more relevant to those modelling the non-domestic energy domain. There is a close relationship between the concept of a 'building footprint' in MasterMap (irrespective as to whether this actually does represent a building) and an addressable object in the Address Layer. However, the relationship between MasterMap and Valuation Office Agency (VOA) data (which is used to identify the footprint of commercial activities) is less well defined. VOA data are not related to buildings but to a concept called hereditaments, equivalent to the term premise and described by Evans et al. (2014). As such, premises are fluid and have a complex relationship with buildings and their spatial representation in MasterMap. Taylor et al. (2014), Evans et al. (2014) and Evans et al. (2017) have developed approaches to address energy modelling, however these are restricted to the non-domestic domain and use of VOA data, unlike the work we present here.

Modelling building stock is not restricted to the UK. Meinel et al. (2009) developed a system for analyzing settlements through integrating topographic, block boundary and statistical datasets. Hecht et al. (2015) investigate automatically classifying building footprints into eleven residential and non-residential categories achieving upto 95% when the most detailed data source (3D city model) is adopted. Building classification based on remotely sensed data and national mapping agency datasets has also been demonstrated (Belgiu et al., 2014). In the context of urban map generalisation, Lee et al. (2017) describe an approach to building classification based on existing map data for two districts in the South Korean capital of Seoul. Their work evaluated a range of machine learning techniques discovering that such an approach though viable, requires increased predictive accuracy. Recently, Hartmann et al. (2016) described an approach that automatically quantified the entire stock of German buildings using databases of footprints, land-use and addresses. Many issues highlighted in these articles are related to the quality of the available input data and can mean different data extraction approaches are required. Hecht et al. (2015) provide a good overview of top-down and bottom-up information extraction approaches and note that the Volunteer Geographic Information (VGI) platform OpenStreetMap (OSM) is not suitable due to discrepancies in building geometry and incomplete coverage (Hecht et al., 2015: 23). Although spatial variability in data quality is likely to be a long-term issue in VGI (Goodchild and Li, 2012; Haklay, 2010), the quality can only improve as a result of increased formal and informal engagement and improved sensing and algorithmic capabilities, quality control and conflation procedures.

Although there have been dramatic improvements in the ability to extract and combine information so that urban energy models can be automated, each of the above applications employs bespoke algorithms or software implemented within predominantly proprietary environments. The nature of these algorithms is often not publicly released and this represents a significant barrier for re-use by the research community and beyond. As such, this paper has reviewed the nature of techniques used within the UK energy domain and has developed an approach which can either be directly re-used by those sharing the same software set or re-implemented using the information described in this paper.

This paper describes the following in detail:

- How a database of residential buildings with a heating component can be automatically derived for the UK
- How these buildings can be automatically classified to identify qualitative attributes on a building by building basis
- How these buildings can be automatically analysed to extract quantitative attributes on a building by building basis

A comparative analysis evaluates how the qualitative and quantitative outputs impact on energy models focused on the city of Nottingham, UK.

The main database queries described have been placed on GitHub and are released under an open Creative Commons by-attribution CC-BY licence.*

The data

This work makes use of Ordnance Survey MasterMap as a source of 2D polygons representing building features. In addition, the address database AddressBase Plus is used as a data source to help determine which buildings in MasterMap are part of the domestic housing stock. Although these datasets are specific

*https://github.com/lucas-uk/pgBuiltForm for the UK, similar approaches might be applied to other high-quality building footprint and address point databases. For the UK context, further details on the characteristics of MasterMap and AddressBase are included as in Appendix A.

The conceptual approach

The aim of this study is to determine qualitative and quantitative spatial and metric relationships about buildings with a significant heating component, so that derivatives can be used for urban energy modelling. A proxy for buildings with a significant heating component has been assumed as those buildings which are addressable. There are some pre-requisites for the modelling in that the input data must:

- Represent the geometry of the set of buildings and:
  - encode spatial geometry in a manner which allows the extraction of qualitative spatial relationships
  - encode spatial geometry in a manner which allows the extraction of quantitative measurements (i.e. one can extract building measurement metrics which closely reflect reality)
- Allow the extraction of a sub-set of those buildings which satisfy the problem (i.e. buildings with a significant heating component)

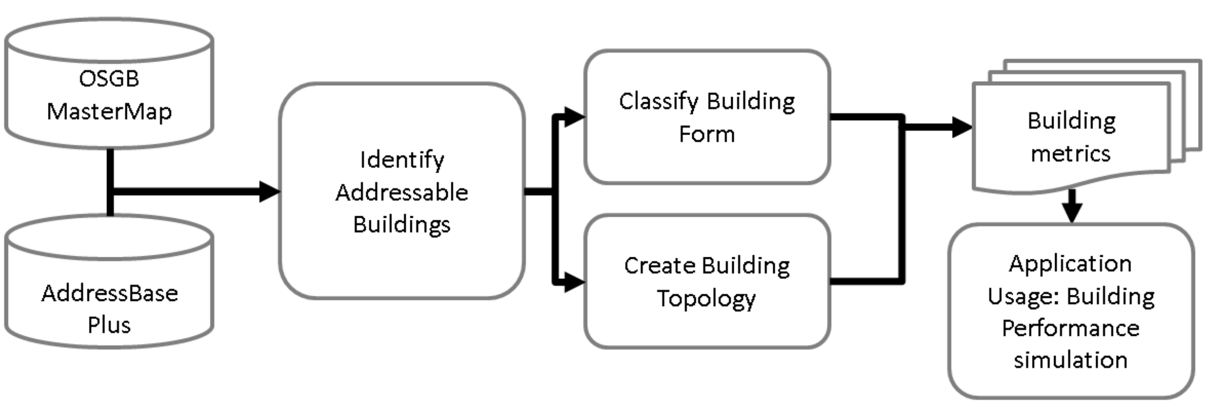

Figure 1. Workflow overview and processing steps involved

Figure 1 shows an overview of the approach developed and reported in this work.

The utility of any map is a function of the underlying data (its scope, quality and scale of capture) and the generalisation and visualization criteria developed for the cartographic process. Different map conceptualisations are required for different decision making processes. Figure 2 describes a conceptual set of addressable buildings that satisfies the above criteria. The buildings themselves are represented as polygons. Where there is an adjacency relationship between these polygons then the polygons touch (this does not mean they necessarily share a boundary in a topological manner).

For this approach OSGB MasterMap contains the spatial set of buildings. MasterMap is a high quality dataset that satisfies both quantitative and qualitative geometric requirements. However, it does not encode which buildings are addressable. AddressBase Plus is used to identify which buildings are addressable. By combining MasterMap and AddressBase Plus data our pre-requisites are satisfied.

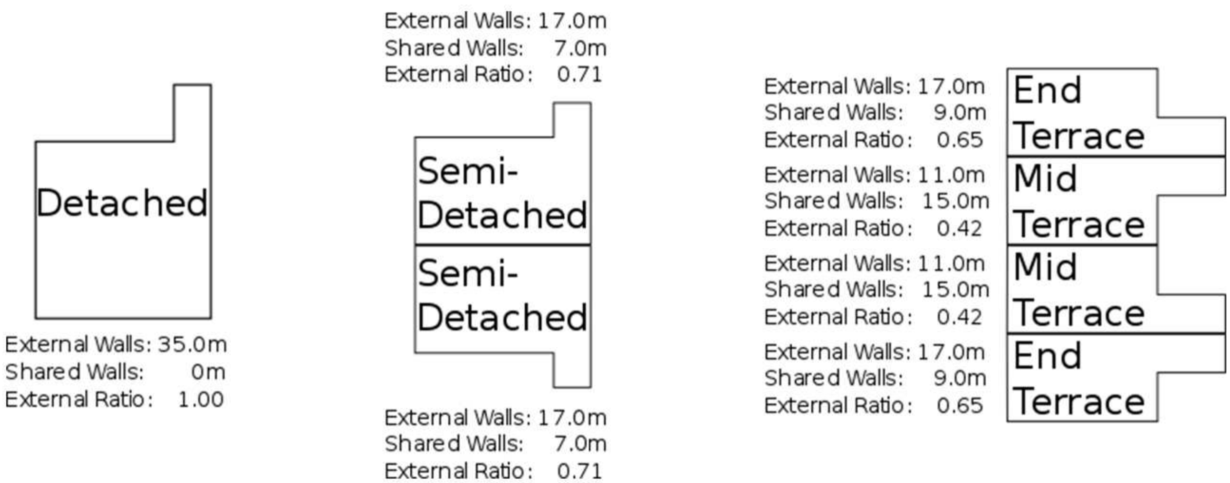

Figure 2. Conceptual examples of built form with extracted footprint metrics (Beck (2016))

Creating the set of addressable buildings

We are interested in buildings which demand energy for heating as opposed to those which do not (i.e. to differentiate between houses and garages). We are using the fact that a building can be addressable as a proxy indicator that it is heated. OSGB MasterMap is a data product that expresses a range of different topographic characteristics across the UK. AddressBase Plus is a complementary dataset which expresses address details for the UK. Aspects of both datasets are required in order to identify addressable buildings.

The following filters are applied to AddressBase Plus:

- Field state is filtered on 2 (In use) or 3 (Unoccupied)
- Field rpc (Representative Point Code which is used to describe the accuracy of the coordinates allocated to the record) is filtered on 1 (Visual Centre) or 2 (General internal point)
- Field postal_address is filtered on S (Single address) or C (Child address for properties that contain multiple addresses (flats etc.))

The next challenge is to link the Buildings from MasterMap to the Occupiable Addresses in AddressBase Plus. There are two ways to do this:

1. cross-reference between the two tables using the TOID
2. conducting a point in polygon query

In theory these should produce the same result. However, analysis has demonstrated that these yield different results, raising the question of which is the most appropriate approach. Figure S1 shows the difference in counts between spatial and TOID-based joins when linking AddressBase Plus to MasterMap. Every deviation from zero shows a mismatch in either the conceptual understanding of the data or errors in the data themselves. For example, the outlier cluster at c. 130 should be of interest to GeoPlace (but will not be considered further in this paper).

The resolution of this problem goes beyond the scope of this paper. To avoid any undue bias we chose addressable buildings as those that had any form of relationship either as a TOID link or spatial relationship between AddressBase Plus and MasterMap.

Qualitative classification of addressable buildings to determine built form

Here we propose that built form can be defined directly from the geometry if the data requirements (described above) are met. This can be achieved by considering the spatial relationships of each building with all other buildings, and particularly if they touch. If the building polygons have poorer quality geometry, then a looser relationship like overlaps might be more appropriate - for example when using raw OpenStreetMap data. Alternatively, the data might be cleaned for undershoots and overshoots; however, such a process is not straightforward. The spatial relations can be determined according to Region Connected Calculus (RCC8) spatial predicates. RCC8 was initially implemented by Cohn et al. (1997) and has similar implementations from Egenhofer and Franzosa (1991), DE-9IM and the Open Geospatial Consortium (Herring, 2011).

By iterating over the set of polygons, a count of the number of other polygons it touches can be calculated. As each polygon is considered to be an addressable building with the type of relationships described in Figure 2 there is a limited range of expected results (0 = detached, 1 = semi-detached or end-terrace, 2 = mid-terrace, >2 = complex polygon).

As identified by Orford and Radcliffe (2007), the End-terraced classification requires an understanding of the second-order touch relationship (i.e. what touches the building that the source building touches). Using this approach, built form classification is directly derived from OSGB MasterMap (see figure 3). The classification does not map directly to the six classes proposed by Rylatt et al. (2001). However, the extra classes can be added with further logic. For example, connected passageway elements might be identified by establishing connections to at least two addressable buildings. The number of addresses at a property can also be used to infer flats which can be further supplemented by other information within AddressBase. These will be considered in future incarnations of the algorithm - but only after the classes are re-characterised on a fitness-for-purpose basis.

The PostGIS code for this qualitative classification is available on GitHub[†].

Quantitative implications of the qualitative classification abstraction The qualitative analysis of spatial relations, described above, is a general abstraction for identifying built form. However, as a generalisation it means that certain real-world events are poorly represented. Figure 4 compares the conceptual examples of built form with some real world scenarios. The properties identified as terraces only touch over a small proportion of their perimeter. Hence, they will have thermal characteristics that are more in-line with detached properties.

[†]https://github.com/lucas-uk/pgBuiltForm/blob/master/sql/built_form_classification.sql

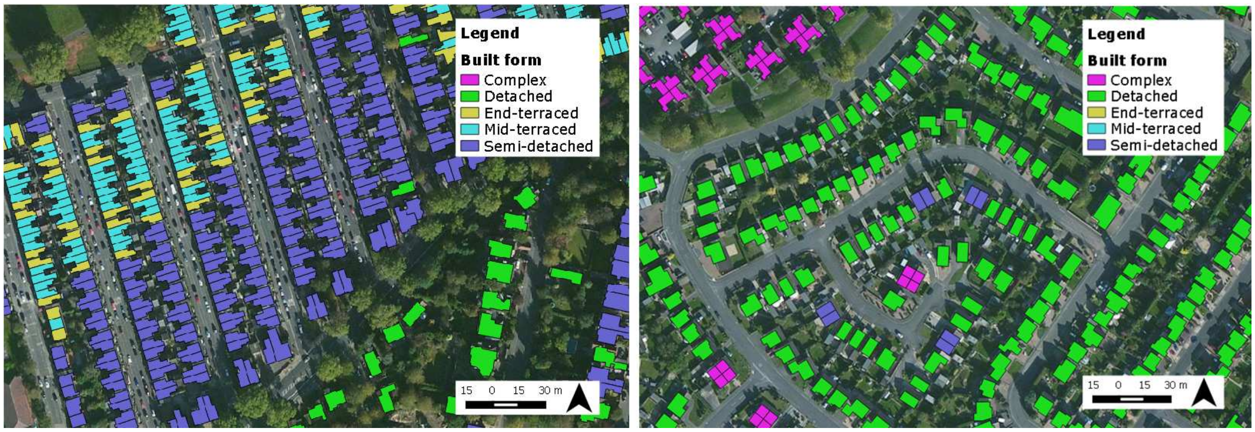

Figure 3. Map examples showing built form classification of buildings in Lenton & The Park (left) and Wollaton (right), Nottingham based on qualitative spatial reasoning © Crown Copyright and Database Right (2015). Ordnance Survey (Digimap Licence).

This is a classification issue inherent in any abstraction process. In these instances, quantitative metrics that compare the external wall ratio of each building are a better reflection of reality. Essentially all buildings are comprised of shared and non-shared walls. Shared walls between buildings represent the qualitative touching concept. Non-shared walls are therefore external walls. Shared and non-shared walls have different heat loss attributes. The external wall ratio calculation is described in Equation (1).

(1)

$$External\_Wall\_Ratio = \frac{\sum NonShared\_Wall\_Length}{\sum NonShared\_Wall\_Length + \sum Shared\_Wall\_Length}$$

It is appreciated that there will be different external wall ratios during different construction periods (due to changes in vernacular construction) and that commercial and domestic properties are conflated in this example. Figure 5 describes the relationship between the external wall ratio metric against each of the qualitative classifications. This describes 117,030 addressable domestic buildings in Nottingham.

As might be expected, the median and distribution of external wall ratio are very similar for end-terraced and semi-detached properties. In addition, and also as might be expected, the detached properties produce a clean set (i.e. the qualitative and quantitative technique produced exactly the same sub-set) with no dispersion about the median external wall ratio of 1.0. Mid-terraced and complex properties exhibit a large

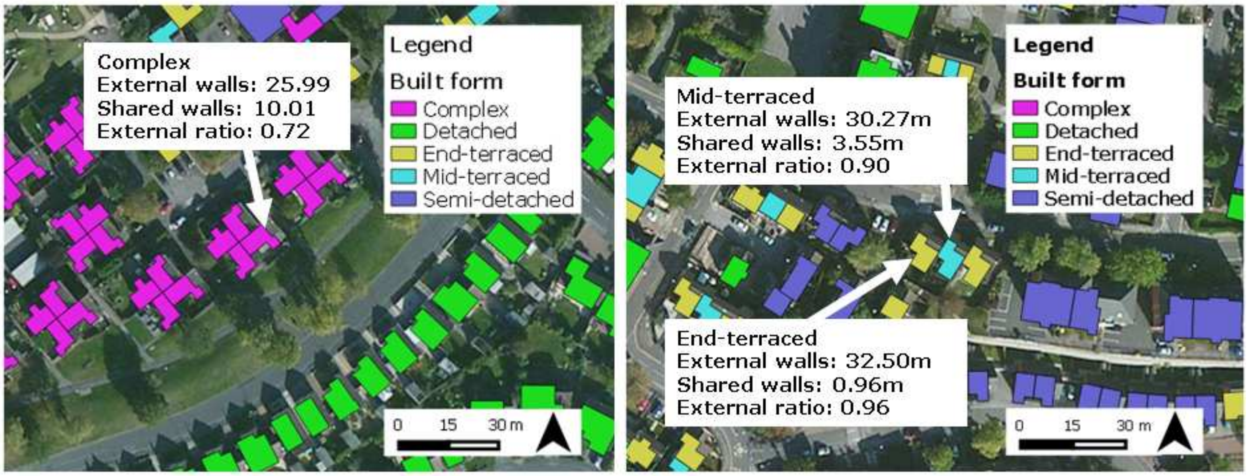

Figure 4. Real examples of building form in Nottingham (Wollaton, left and Mapperley, right) and their implications for measurements. In the new terraced scenario the shared boundary (where properties touch) represents a significantly lower proportion of the building footprint than in the stereotypical terrace. This means that these buildings have a greater proportion of external wall and should be weighted differently for their potential heat loss. © Crown Copyright and Database Right (2015). Ordnance Survey (Digimap Licence).

range of values for the external wall ratio and are clearly multi-modal. Whilst mid-terraced may have some positive correlation with construction period this may not be true for the complex buildings.

What is clear from Figure 5 is that the qualitative abstraction process can mask the range of underlying quantitative building metrics which are accessible through the geometry of OSGB MasterMap.

Extraction of quantitative metrics for each building in the set of addressable buildings

Buildings within MasterMap are represented as polygons and correspond to an accurate representation of the building footprint. This section describes a topological query for extracting the external wall ratio, and the identification of lengths of shared and non-shared walls which together comprise the perimeter of building footprints. Within a GIS, topology expresses the spatial relationships between connecting or adjacent vector features (points (nodes), polylines (edges) and polygons (faces)). Further details on the PostGIS implementation of topology are provided in Appendix B. Our query identifies every wall associated with a building's footprint, its length in metres and whether the wall is shared between building footprints. These individual wall segments then need aggregating for each building footprint. First we group the data so each value for left_face can have a maximum of two instances that reflect non-shared and shared then

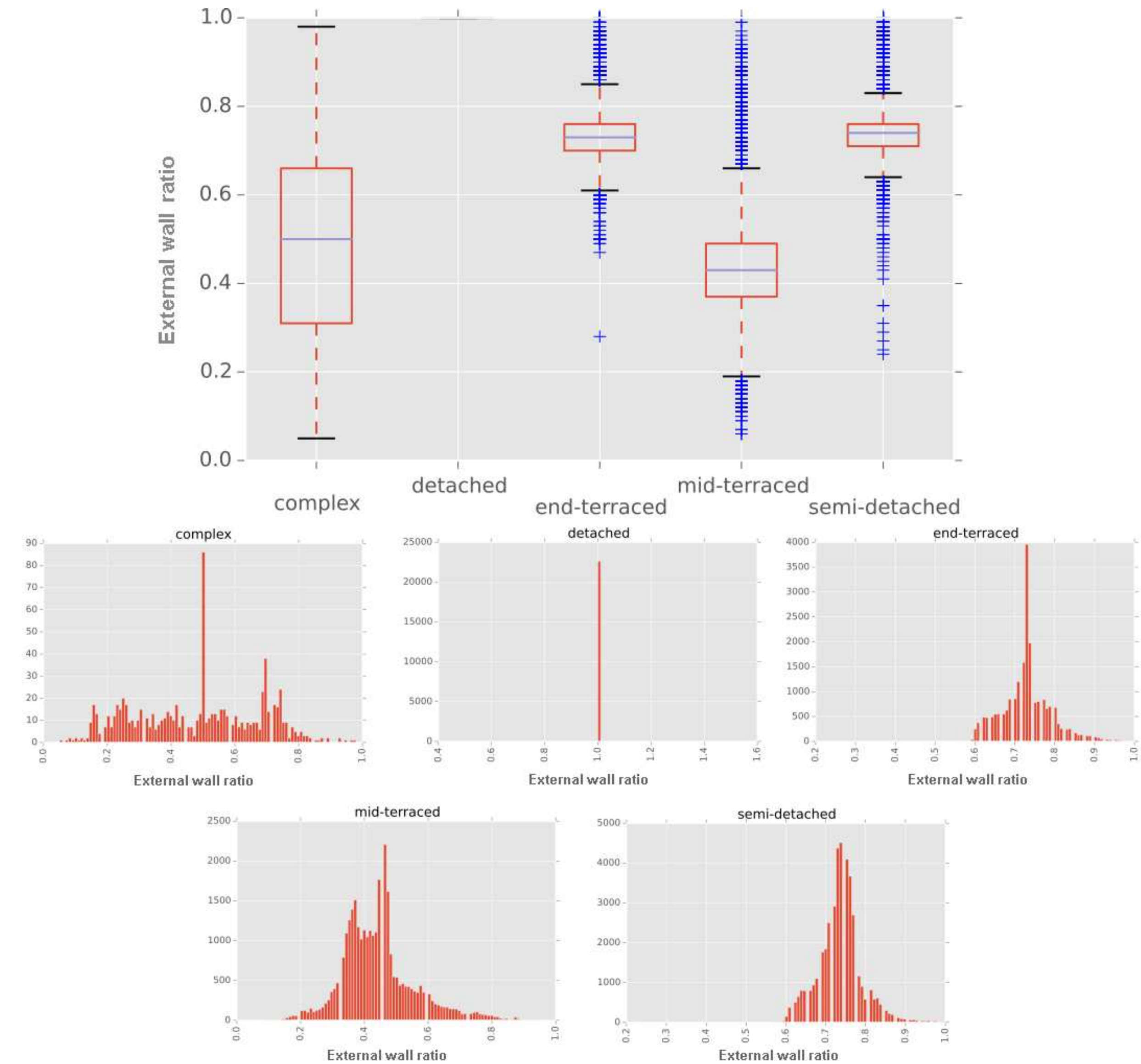

Figure 5. External wall ratios for Nottingham. Box and whisker plot of external ratio by property type (top row), histograms of external wall ratio for complex, detached, end-terraced, mid-terraced and semi-detached property types (middle and bottom rows).

Equation 1 can be used to calculate the exterior wall ratio. Example results for this calculation are provided in Table 1. The query for establishing these results is located on GitHub[‡].

The topology tables have their own unique identifiers that do not directly match to the unique identifiers in the original GIS data. This topology identifier needs mapping back to the source data table which holds the data and attributes as polygons (and not topology elements). To do this we utilise the linking field in

[‡]https://github.com/lucas-uk/pgBuiltForm/blob/master/sql/compute_exterior_wall_ratio_topology_tables.sql

the original polygon table held as a topogeometry data type. For example, we can extract the id data from topogeometry field by using the GetTopoGeomElements function. This can be used to create a new table that contains the building footprint geometry, GIS linking data and extracted metrics. The query for this is located in the GitHub repository.

Table 1. Example of output from the exterior_wall_ratio query.

ID	Non-shared wall length	Shared wall length	External wall ratio
110145	17.06	8.4	0.67
110142	17.05	8.4	0.67
110148	8.64	16.8	0.34
110143	8.64	16.8	0.34
110144	8.65	16.8	0.34
110147	8.65	16.8	0.34
110146	17.06	8.4	0.67
110149	17.04	8.4	0.67

Case study of the application in the city of Nottingham, UK

To demonstrate our approach, we apply the techniques described above to a case study of Nottingham, UK. We make use of energy models and data developed within the InSmart project - a collaboration of four municipalities (Cesena in Italy, Evora in Portugal, Nottingham in the UK and Trikala in Greece) to develop rigorously formulated Sustainable Energy Action Plans (SEAPs). This involves modelling the energy use of: domestic and non-domestic buildings (with an emphasis on the former), transport of goods and people (with an emphasis on the latter), public services (e.g. street lighting) and industrial processes. Working with municipal partners, scenarios are then defined and evaluated with these models. The corresponding results, together with representations of the energy supply system, are then used to calibrate a TIMES-Markal energy system model for the identification of cost-optimal decarbonisation investments. Finally, the outcomes from this exercise are evaluated against both quantitative and qualitative criteria, using the Multi-Criteria Decision Analysis (MCDA) tool Promethee (Brans and Mareschal, 1992), to identify the highest ranking investments against the full range of criteria. These criteria include: cost, reductions in energy use and carbon emissions, technical constraints, legal issues, social acceptability and impacts on the economy and quality of life. Further details of the application of these modelling techniques can be found in De Miglio et al. (2016), and with respect to Nottingham in particular in Long and Robinson (2016b). With this information, the research and municipal partners co-develop their corresponding SEAPs. The case study in this paper relates to the workflow that was developed to simulate the energy performance of Nottingham's housing stock, identified by the local authority as having just over 130,000 dwellings (Nottingham City Council, 2016).

This exercise was informed by information derived from the Cities Revealed dataset (courtesy of the GeoInformation Group). This analysis provided the baseline data against which the new qualitative and quantitative data were compared. It was found that the method greatly reduced the time required for verification (an automated, as opposed to the previous semi-automated, approach) and, by having more control of the classification metrics, enabled the modeller to quickly highlight buildings with anomalous built form classifications.

In addition, the method identified a small but noteworthy number of residential buildings where the rule-based classification of built form and the external wall ratio value showed a mismatch, i.e. unusually high or low values of external wall ratios compared to values typically expected for that built form. Approximately 1560 buildings (1.27% of the total) were identified showing this mismatch between external wall ratio and built form. Of these, almost two thirds (985) were defined as 1970s terraced or semi-detached properties, and related to a particular architectural style of step-linked housing built across a number of areas in the city in that period. Figure 4 showed a typical example of step-linked terraced housing in the city of Nottingham. This step-linked housing is distributed throughout city of Nottingham, with over 80% of which is found in four of the city's twenty wards. Within these four wards, this type of housing has a small but significant presence. Microsimulation of these forms was undertaken to understand its impact on a building's energy use.

The InSmart energy models developed for these two typologies were modified to reflect the step-linked designs found in the city's housing stock. These models were then simulated with EnergyPlus using scripts developed to generate a synthetic stock as described by Long et al. (2015). The script ensures that all likely permutations of the key energy parameters for each energy model (identified through sensitivity analysis) are modelled. A detailed survey of over 580 residential properties in the city had been carried out in 2015. The results of this survey were used to model the distribution of typical values for the key energy parameters for each building typology. DesignBuilder was used to construct the models which were simulated in EnergyPlus, the full details of which can be found in Long et al. (2017; In prep.).

In general terms, the energy use of the mid-terraced property for a step-linked variant is similar to that observed for a traditional end-terrace version. This is not surprising since their external wall ratios are very close. As shown in Figure S2, the step-linked end terrace (equivalent to a semi-detached) property shows higher energy demand than a similar detached property of the same construction period. However, this result can be explained due to differences in the construction of the model and the properties of the specific building typologies (e.g. 1970s terraced properties have lower levels of wall insulation than 1970s detached properties).

A range of remediation solutions for the energy modelling of step-linked housing are possible. The choice of specific option would be dependent on the prevalence of this type of housing within the area modelled:

1. If the presence of step-linked housing is very low, the discrepancy in energy performance between step-linked and traditional housing forms could be ignored. Other energy parameters (e.g. occupancy levels and behaviour, wall/roof insulation or infiltration rates) would likely have a greater impact on the overall energy demand of the stock.
2. Where step-linked housing is a significant element of the housing stock, a suitable solution might be to alter the attribution of those buildings to reflect a form whose energy performance it more closely represents (i.e. mid-terrace becomes end-terrace, semi-detached/end-terrace becomes detached).

3. In instances where step-linked housing is a major element of the area modelled, it may be appropriate to create building performance simulations specifically for that type of housing and add the step-linked typologies to the set of building archetypes for the modelled area.

Discussion

The InSmart buildings dataset contains data on construction period taken from the Cities Revealed data set. These data were extensively updated using a semi-automated approach combining local knowledge, manual verification and an informal application of the geometric rules proposed above. The results of this exercise could assist in a more formal definition of a rule-based approach to the classification of construction period. Figure 6 shows examples of building footprints for a number of terraced properties of different construction periods in the city of Nottingham.

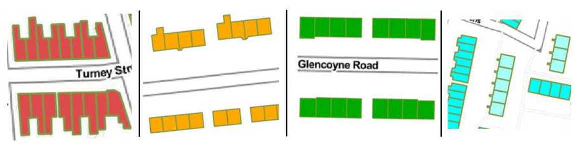

Figure 6. Examples of terraced properties in Nottingham. From left to right: Victorian terraces (built pre 1914), Inter war terraces (1915-1945), Post war terraces (1946-1964) and 60s/70s terraces (1965-1979).

The Victorian terraces (shown in red) are most distinctive with narrower front and rear walls and extensions to the rear of the property. Building height is typically greater than other terraces due to higher ceilings in each storey and the prevalence of three storey versions in some neighbourhoods.

The inter/post war terraces (shown in orange and green) are very similar in their geometric features and tend to be represented by simple rectangular shapes. The length of the terraces in these properties tends to be lower than the other examples with 3-4 properties being typical in each terrace. It is difficult to distinguish between inter and post war examples solely through their geometric features. It is likely that local knowledge would be needed to identify the specific construction period. However, energy modelling carried out on these typologies showed that the energy performance of these two types is similar, assuming that typical construction materials and methods were used.

The 60s/70s terraces are the most diverse with a number of different designs present in the city. Those shown in Figure 6 are based on the Radburn estate and include step-linked designs. There are more traditional terraces in this construction period, similar to the inter/post war examples. However, the number of properties in each terraced block is often higher than for the inter/post war period, which would assist in defining these types.

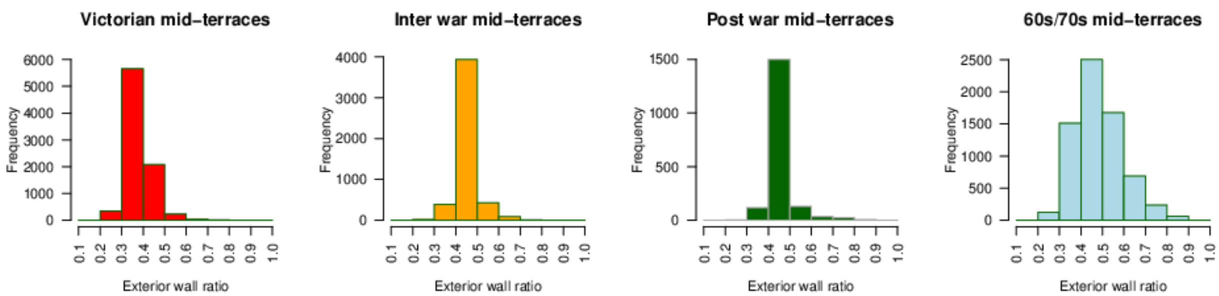

Figure 7. Distributions of mid-terraced housing exterior wall ratios by age.

Analysis of the external wall ratios, see Figure 7, using the InSmart housing stock model's construction period provides quantitative data to test the anecdotal evidence from the examples shown in Figure 6. Victorian properties have a lower average external wall ratio while inter and post war properties share a similar mean value and distribution of external wall ratio. The 60s/70s properties, which include a wider range of built forms compared to the other mid-terraced properties, show a much broader distribution of external wall ratio values.

As described earlier, typologies could be generated by conflating the built form with the construction period. Extension of the approaches discussed in this article to further enrich urban datasets offers an interesting opportunity for enabling more involved simulation e.g. using software such as CitySim to perform explicit energy micro-simulations of buildings in their true spatial context (Robinson et al., 2009). Although we have not described deriving construction period in this article, the blended use of quantitative and qualitative metrics should help the development of a system to identify construction age based on vernacular building types. For example, a rule could be described that classifies a type of building constructed between a and b which has the following characteristics: type terrace, external wall ratio of x, area of y and z no of floors.

Conclusions

This paper has demonstrated that calibration data required for domestic energy models, such as BREDEM, can be effectively inferred from high quality topographic map and address data, sourced from a National Mapping Agency. As an automated workflow, the approach could simplify the energy modelling of houses and stocks of houses at any scale and also has implications for practitioners of urban simulation interested in extrapolating predicted energy use derived from a typology across large areas. Although the approach and case study is presented utilising Ordnance Survey MasterMap and AddressBase Plus datasets of Great Britain, a similar approach could be applied to other countries providing building footprints and addresses are available. However, the positional accuracy and footprint detail would be important considerations when using other datasets. If using OpenStreetMap then significant cleaning might be required in order to create a valid and accurate topology. In this work, both inputs are spatial datasets and thus no text-based address matching is required to join the datasets, however such an approach might be applicable elsewhere.

For this work, a subset of buildings in MasterMap is identified that represents the set of buildings which have a heating requirement. In reality this is not completely true as not all buildings that demand energy for heating are addressable buildings. However, the majority are and this approach will remove all buildings which have no heating requirement (such as garages, many warehouses, sub stations etc.). Exceptions to this rule, if they significantly bias the model, can be added at a later date or can be included in a more refined model. Furthermore, although this work does not detail use of height data which is utilised by BREDEM, such data can be obtained from the Ordnance Survey Building Height Attribute dataset (Ordnance Survey, 2014) or from a LiDAR survey, providing the quality and coverage matches the study area. In the former case, this would only enable generation of single extruded footprint volumes and not take into consideration cases where a house has, for example, an extension of a different height. Further detail could be provided through a 3D city model which are increasingly planned or produced at the national level (Sargent et al., 2015; Stoter et al., 2014). 3D city models can help with both categorising the stock (Hecht et al., 2015), help identify exposed wall areas more precisely (Evans et al., 2017) and enable analyses of other contributing factors on heat loss such as building shadowing (Biljecki et al., 2015).

One of the dangers in modelling is that our abstractions can become redundant, especially when more or better quality data becomes available. In our case, there is a demand for the traditional housing form in order to calibrate the BREDEM model (or a dynamic alternative) - as that is what the model has always demanded. As such, the high quality geometric content of MasterMap can be used to infer these data. However, it can also be used to identify further quantitative metrics that could be employed to improve the underlying BREDEM model. In this case, the model abstraction can be improved so that it more closely represents reality. The external wall ratio represents just such a metric and such detail could also be better utilised within energy simulation software. Furthermore, it is also clear from Figure 5 that some of these distributions are multi-modal and the inclusion of other quantitative measures (such as area) could provide insights into construction period and is a promising line of further research. This need not be specific to any country and a wider need for further research on appropriate building typologies in the context of Germany has been noted (Hartmann et al., 2016).

Unfortunately, as the complexity of the products increases it becomes increasingly difficult to filter out anomalies in the data (i.e. only buildings not addressable properties). This represents an opportunity for an organisation (e.g. OSGB or GeoPlace) to provide domain specific enhancements on core data that can be included to enhance and improve downstream modelling activities. This would represent a shift for organisations like OSGB from a provider of mapping products to a provider of domain/task orientated data products/enhancements.

This would require an interoperable element which, in this instance, could be the Unique Property Reference Number (UPRN). However, there are a number of issues with this:

- There are difficulties when linking UPRN to other representations of space (such as hereditaments from the VOA).
- Restrictive licensing (and mixed licensing) may be an issue for product derivation and re-use.

In terms of the latter point, the UK Government recognises this as an important issue and has released funding to create an open address framework (Open Data Institute, 2014). This would, by necessity, include some form of interoperable addressing element. This has the potential to streamline endeavours to model building stocks in support of decarbonisation and fuel poverty alleviation policies.

References

Baker KJ and Rylatt RM (2008) Improving the prediction of uk domestic energy-demand using annual consumption-data. Applied Energy 85(6): 475–482. Available from: http://www.sciencedirect.com/science/article/pii/S030626190700147X.

Beck A (2016) House classification by form - conceptual approach. Available from: https://commons.wikimedia.org/wiki/File:House_Classification_By_Form_-_Conceptual_Approach_01.svg (accessed 30 October 2016).

BEIS (2017) Energy consumption in the uk (ecuk) 2017. London. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/633503/ECUK_2017.pdf (accessed 22 October 2015).

Belgiu M, Tomljenovic I, Lampoltshammer TJ, et al. (2014) Ontology-based classification of building types detected from airborne laser scanning data. Remote Sensing 6(2): 1347–1366. Available from: http://www.mdpi.com/2072-4292/6/2/1347.

Biljecki F, Stoter J, Ledoux H, et al. (2015) Applications of 3D city models: State of the art review. ISPRS International Journal of Geo-Information 4(4): 2842–2889. Available from: http://www.mdpi.com/2220-9964/4/4/2842.

Brans JP and Mareschal B (1992) Promethee v: Mcdm problems with segmentation constraints. INFOR: Information Systems and Operational Research 30(2): 85–96.

Chapman PF (1994) A geometrical model of dwellings for use in simple energy calculations. Energy and Buildings 21(2): 83–91. Available from: http://www.sciencedirect.com/science/article/pii/0378778894900019 (accessed 22 September 2015).

Cohn AG, Bennett B, Gooday JM, et al. (1997) RCC: A calculus for region based qualitative spatial reasoning. GeoInformatica 1(3): 275–316.

De Miglio R, Chiodi A, Simoes S, et al. (2016) New methodological approach for planning cities sustainable and resilient energy futures: The case of the insmart project. Ireland: University College Cork; University

1(1): 18–31. Available from: http://dx.doi.org/10.1080/23729333.2015.1055644 (accessed 22 September 2015).

Hendersen J and Hart J (2015) BREDEM 2012 specification. Building Research Establishment. Available from: http://www.bre.co.uk/filelibrary/bredem/BREDEM-2012-specification.pdf (accessed 23 September 2015).

Herring J (2011) OpenGIS implementation standard for geographic information-simple feature access-part 1: Common architecture. OGC Document 4(21): 122–127.

Holtier S, Steadman JP and Smith MG (2000) Three-dimensional representation of urban built form in a gis. Environment and Planning B: Planning and Design 27(1): 51–72. Available from: http://epb.sagepub.com/content/27/1/51 (accessed 22 September 2015).

Hussain M, Barr R and Chen D (2012) Building-based urban land use classification from vector databases in manchester, uk. In: 2012 20th International Conference on Geoinformatics (GEOINFORMATICS), pp. 1–7.

Lee J, Jang H, Yang J, et al. (2017) Machine learning classification of buildings for map generalization. ISPRS International Journal of Geo-Information 6(10).

Long G and Robinson D (2016b) InSMART - report on the multi-criteria methodology, the process and the results of the decision making.

Long G, Robinson D and Alwany M (2015) Building typologies simulation report. Nottingham. Available from: http://www.insmartenergy.com/wp-content/uploads/2014/12/D.2.1.-Simulation_Report_for_building_Typologies-Nottingham.pdf (accessed 8 August 2016).

Long G, Alalwany M and Robinson D (2017; In prep.) Decarbonisation scenario modelling of city housing stocks by typological sampling.

Mavrogianni A, Wilkinson P, Davies M, et al. (2012) Building characteristics as determinants of propensity to high indoor summer temperatures in london dwellings. Building and Environment 55: 117–130. Available from: http://www.sciencedirect.com/science/article/pii/S0360132311004197.

from: http://www.sciencedirect.com/science/article/pii/S0198971506000731 (accessed 22 September 2015).

Robinson D, Haldi F, Kämpf J, et al. (2009) CitySim: Comprehensive micro-simulation of resource flows for sustainable urban planning. In: Proc. building simulation, Citeseer, pp. 1614–1627.

Rylatt M, Gadsden S and Lomas K (2001) GIS-based decision support for solar energy planning in urban environments. Computers, Environment and Urban Systems 25(6): 579–603. Available from: http://www.sciencedirect.com/science/article/pii/S0198971500000326 (accessed 22 September 2015).

Sargent I, Holland D and Harding J (2015) The building blocks of user-focused 3D city models. ISPRS International Journal of Geo-Information 4(4): 2890–2904. Available from: http://www.mdpi.com/2220-9964/4/4/2890.

Science TGO for (2008) Sustainable energy management and the built environment project (2008). final project report. London. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/293085/powering-our-lives-final-report.pdf (accessed 22 October 2015).

Steadman P, Evans S and Batty M (2009) Wall area, volume and plan depth in the building stock. Building Research and Information 37(5-6): 455–467.

Stoter J, Vosselman G, Dahmen C, et al. (2014) CityGML implementation specifications for a countrywide 3D data set. Photogrammetric Engineering & Remote Sensing 80(11): 1069–1077.

Taylor S, Fan D and Rylatt M (2014) Enabling urban-scale energy modelling: A new spatial approach. Building Research and Information 42(1): 4–16. Available from: http://dx.doi.org/10.1080/09613218.2013.813169.