Calderon C, James P, Urquizo J, McLoughlin A.

[A GIS domestic building framework to estimate energy end-use demand in UK sub-city areas.](#)

Energy and Buildings 2015, 96, 236-250.

Copyright:

NOTICE: this is the authors' version of a work that was accepted for publication in Energy and Buildings. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Energy and Buildings, vol. 96, 1 June 2015. DOI: 10.1016/j.enbuild.2015.03.029

Link to published article:

<http://dx.doi.org/10.1016/j.enbuild.2015.03.029>

Date deposited:

27/03/2015

This work is licensed under a

[Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License](#)

A GIS domestic building framework to estimate energy end-use demand in UK sub-city areas.

Carlos Calderón^{a,*}, Philip James^b, Javier Urquiza^a, Adrian McLoughlin^c

^a*School of Architecture Planning & Landscape, Newcastle University
Newcastle upon Tyne, UK, NE1 7RU*

^b*School of Civil Engineering & Geosciences, Newcastle University
Newcastle upon Tyne, UK, NE1 7RU*

^c*Climate Change Officer, Newcastle City Council, Civic Centre
Newcastle upon Tyne, UK, NE1 8QH*

Abstract

This paper presents the development, evaluation and application of a spatially referenced domestic building level framework (i.e. address level) to estimate domestic energy end-use demand baseline in sub-city areas. The paper core idea and conclusion is that unless knowledge and model estimating is available at an appropriate level, future UK local energy infrastructure planning will not be effective. Our framework innovatively combines a dataset, which includes detailed building surveys of 60,977 out of a total of 139,257 dwellings, with a normalised national dataset (i.e. the English Housing Survey) and applied to a BRE Domestic Energy Model (i.e. Cambridge Housing Model) so as to establish an energy consumption baseline for the domestic stock in localised areas of Newcastle upon Tyne. Our validation results show a poor alignment with existing observed data as published by the Department of Energy and Climate Change (DECC), particularly at neighbourhood scale. Our belief is that as spatial resolution is increased, local building and urban socio-economic and physical characteristics play a more important part in the estimation of dwelling energy consumption. Thus, we propose a taxonomy to holistically deal with the sources of uncertainty arising from these issues and the components of our framework.

Keywords: cities, energy, data sets, neighbourhood, districts, local energy policy,

*Corresponding author

Email addresses: carlos.calderon@newcastle.ac.uk (Carlos Calderón), philip.james@newcastle.ac.uk (Philip James), j.urquiza@newcastle.ac.uk (Javier Urquiza), adrian.mcloughlin@newcastle.gov.uk (Adrian McLoughlin)

1. Background

After a series of national policy initiatives (e.g. CERT, CESP), UK local authorities (LAs) are now in need to go beyond “low hanging fruit” energy measures such as loft and cavity insulation if they are to meet the stringent carbon targets set in the climate change act. However, as [1, p. 1] have illustrated a unique future national energy scenario would translate itself into a myriad alternative energy futures pathways at the local authority scale (see Figure 1).

As a result, LAs require a spatial understanding of end-use energy demands in targeted city areas to provide evidence for meeting future challenges in planning local energy services and infrastructure. In particular, LAs have identified the domestic sector as a priority and developed a growing interest in mapping and modelling the energy use of their housing stock at city level or below. A robust mapping and modelling methodology would enhance the effectiveness of their local energy planning of future interventions of energy efficiency measures, decentralised energy supply infrastructure, and biomass resource and carbon dioxide storage. For instance, in each city some local areas might be more suited for building centric technological interventions such as those promoted by the Green Deal whereas others may benefit from centralised neighbourhood CHP/ground source systems and networks and distributed solar and wind.

Our work is in line with “the most significant use of sub-national consumption data” as highlighted by DECC [2, p. 13] and aims at enabling LAs and stakeholders

UKERC ENERGY 2050, REF, LEVEL N

UKERC ENERGY 2050, REF, LEVEL R

UKERC ENERGY 2050, REF, LEVEL LA

Figure 1: Illustrative Level N (national), Level R (regional) and Level LA (local authority) scenarios for residential energy demand. Adopted from [1].

(e.g. built environment professionals such as planners, architects and civil engineers) to map and model localised energy demands baselines for their domestic building stock. The work presented is part of an ongoing project as to provide an evidenced based and geographically targeted carbon route map for the city of Newcastle upon Tyne. Hence, we have labelled our framework as Newcastle Carbon Route Map (NCRM).

1.1. Sub-city areas

In this study we have followed the LA’s criteria for selecting three suburban residential areas (Westgate, South Heaton, Castle) which represent the city’s diverse demographic and housing stock/ownership. These areas, in turn, could be put forward as zones of the city to demonstrate different technologies and types of retrofit [3]. Our table 1 shows that our three city zones can be categorised as suburban residential areas (i.e. as categorised by HOMES 2010 [4]) but have very different urban densities. Urban density is a key indicator for evaluating area-based retrofit energy schemes. A building centric retrofit programme may be suited to a low density suburban area such as Castle with 1.8 dwellings per hectare. A decentralised technological energy solution (e.g. district heating) may be more suited to high density suburban area such as South Heaton with 43.48 dwellings per hectare. Similarly, tenure and social demographic profile are useful urban descriptors for discriminating between future funding options and models of retrofit programmes. The funding models for social housing and private owned stock are likely to be very different. In our case, we have a representative mixture of predominant social groups: young families (Group E see 1), Castle; educated young single people (i.e. university students, Group F see 1), South Heaton; and people living in social housing (Group B see 1), Westgate. Table 1 summarises our sub-city area descriptors.

1.2. Current practice and challenges

[5] provide an excellent review of current modelling techniques used for estimating residential sector energy consumption. Due to our data availability, namely, a spatially referenced building level dataset which includes detailed building surveys of 60,977 out of a total of 139,257 domestic dwellings, we followed a bottom-up engineering approach based on an “actual sample house data as the input information to the model” [5, p.1828]. [5, p.1828] state that “the bottom-up engineering approach is the only method that can fully develop the energy consumption of the sector without any historical energy consumption information” and that “these techniques have the capability of determining the impact of new technologies” [5, p.1833] which, in turn, makes the method ideally suited for an area-based delivery approach favoured

		Sub-city Areas		
	Descriptors	South Heaton	Westgate	Castle
Urban Density	Dwellings per hectare	43.48	12.25	1.8
	City Centre	0.0	0.0	0.0
	Other Urban Centre	0.0	0.0	0.0
Area Type (%)**	Suburban Residential **	98.5	98.4	89.7
	Village Centre	1.3	1.0	8.1
	Hamlet/Isolate	0.2	0.6	2.2
	Owner occupied	31	17	78
Tenure (%)	Private rented	38	14	7
	Rented from local authorities	24	49	14
	Rented from housing associations	7	20	1
Demographics	Dominant mosaic group	Group E	Group F	Group B

Table 1: Summary of sub-city area descriptors. The selected areas are a good representation of the city’s diverse demographic and housing stock/ownership. ** HOMES 2010 [4] original categories were adapted to fit with those employed by Newcastle City Council. ** Suburban residential equates to: older urban areas swallowed up by the metropolis, outer areas of cities characterised by large housing states, and residential areas of suburban areas

by LAs. Furthermore, a recent review of mapping and modelling domestic energy demand practice by [6, p.1696] has also revealed that UK residential stock models have two key limitations: 1) lack of transparency and quantification of inherent uncertainties; and 2) resolution of spatial coverage. Our work extends recent studies on inherent uncertainties and transparencies [7], [8],[9] by providing an uncertainty taxonomy as to make our process transparent to stakeholders such as local policy makers. Our study is also based in one of the most comprehensive UK building level dataset and this has enabled us to address the second major limitation and develop a model with high spatial resolution (i.e. per dwelling or at address level).

The Centre for Sustainable Energy (CSE) address level heat map [10] is the UK’s most recent and best effort to provide high spatial resolution and coverage. CSE has acknowledged the model’s methodological and transparency limitations which, in turn, have undermined its use and validity by UK local authorities. A key limitation in CSE’s approach is the lack of building surveyed data. That means it has been made not knowing which energy efficiency measures have already been installed in the dwelling (e.g. loft and cavity insulation) and what type of upgrades or changes have been made to the dwelling heating systems. This is significant as energy insulation measures and heating systems efficiency are the two most important determinants of heating energy use after floor area [11]. Similarly, CSE Heat map’s methodology relies on “multiple key factors” such as size, age, built form, tenure, rurality, and region being *derived* from the English House Condition Survey (EHCS)

2008 and used to predict building heat demand [12]. The process of matching up specific UPRN (Unique Property Reference Number) or building addresses to EHCS records utilising these “multiple key factors” has not been made transparent and therefore we are unable to comment on its advantages and limitations. CSE heat map results have been validated using 2009 Dataset National Energy Efficiency Data Framework (NEED) cross-reference tables. It is also unclear how data from DECC 2009 has been compared with EHCS from a different year, 2008, and used to develop building heat demands. Finally, our work has taught us that the emerging picture surrounding local energy modelling and that, for example, singularities (see 2.2) such as E7 (electricity heating), group heating and district heating (decentralised energy supply), and HMOs (House in Multiple Occupation -planning changes of use-) have a great impact on final energy consumption calculations. CSE’s approach does not explain how these singularities have been dealt with in specific local areas.

Our work follows on CSE’s concept of matching UPRN to EHCS records for deriving “multiple key factors” but addresses current practice shortcomings. The study presented in this paper should also be seen in the context of other efforts [1] to develop a methodology for generating spatially disaggregated energy demand for a present day ‘base year’ and for future scenarios. This type of model would provide a robust description of how local energy demand would change over time, enabling stakeholders in energy demand reduction, energy supply and infrastructure development to account for spatial and temporal variation in energy demand across UK cities. Ideally, as pointed out by [1], our methodology should be embedded within national energy scenarios by utilising appropriate energy demand datasets. This, in turn, would ensure longevity and replicability as it would enable the scientific community and local authorities to develop specific localised energy demand statistics for any set of (existing or future) energy scenarios. The next section explains in detail our methodological approach.

2. Modelling framework

Newcastle Carbon Route Model (NCRM) is a spatially referenced parameterised per-dwelling domestic energy model developed with the purpose of estimating the energy consumption of sub-city areas. Two types of modelling were carried out. The first, using K means clustering of dwelling archetypes which follows [13], we assigned each observation to the cluster whose mean yields the least within-cluster sum of squares (WCSS) using Euclidean distance. The results were presented in [14] and used to benchmark the dwelling energy modelling process. The second modelling process is a sub city Domestic Energy Model (DEM) that creates individual energy

consumption estimates for each dwelling and aggregates these to sub-city areas. Both models utilise a physic based approach to energy modelling based on BREDEM 12 methodology. In our case we use the Cambridge Housing Model (CHM) which implements a BREDEM based energy model to compute energy estimates as it is the model used by DECC to underpin the 2012 Housing Energy Fact File and Energy Consumption in the UK [15].

The NCRM sub-city DEM modelling process is broken down into three separate stages: (1) per-dwelling SAP record augmentation; (2) model refinement and (3) per-dwelling energy calculation and aggregation. In this section, we describe these three steps. Figure 2 shows a schematic view of the methodology and modelling methods.

2.1. Per-dwelling SAP record augmentation

The per-dwelling SAP record augmentation process uses three imputation procedures to augment the NCRM spatially referenced per dwelling data with a complete set of values (i.e. 115 variables) needed for input into a full Standard Assessment Procedure (SAP). The SAP methodology was developed by the Building Research Establishment (BRE) and it is based on the BRE Domestic Energy Model (BREDEM), which provides the framework for calculating the energy consumption of UK dwellings. The per dwelling dataset is augmented by three sequential processes; 1) cross-scale normalisation of data sets; 2) spatial interpolation; 3) best matching algorithm.

2.1.1. Spatially-referenced per-dwelling integration

The spatially referenced per-dwelling dataset integrates a number of building information datasets and then augments these through a matching process with full SAP records from data collected as part of the English Housing Survey (EHS). A full SAP record is required as input to the BREDEM model.

The spatially referenced per dwelling dataset is built on Ordnance Survey building polygons from MasterMapTM. The building polygons are integrated with property information through shared unique identifiers in the attribute data. This data was then combined with other physical building data from three sources: social housing data; Cities Revealed housing data; and data derived from an earlier modelling project [16]. Building information was augmented with survey data collected through household surveys from Warmzone energy audits [17]. A full description of the input data and the processes used to combine them can be found in [18].

The variables that have the greatest impact on energy consumptions are: age of property, wall construction, building form (terraced, detached etc.), dwelling size,

Figure 2: Methodology and modelling methods overview.

heating fuel type and storey height [11]. In addition, through principal component and factor analysis of the data, four further important variables that affected energy consumption were identified. These included wall insulation, primary heating system, boiler type and tenure. From a possible 139,257 residential dwellings in Newcastle a high quality sample where all these variables were present of 60,977 dwellings was created.

2.1.2. Cross-scale normalisation and harmonization of samples

Matching two cross-scale datasets represents a significant statistical challenge. In this case the EHS survey and NCRM per dwelling dataset were derived from the same population namely residential buildings and were carried out at the same time period (2009-2010). In comparing the sample designs the relevant literature on the EHS was analysed [11] and a cluster analysis of the building archetypes was carried out on the NCRM dataset. The published literature on the EHS sampling and data aggregation [19, p.85] states that the full EHS household data set has a stratified random sample design and the dwelling or paired data set has been stratified according to archetypes [20]. A number of significant clusters were present in the NCRM data suggesting that like the EHS survey the high quality sample could be considered as a stratified random sample.

In this study, we utilised semantic and numerical output harmonization techniques to map NCRM data to EHS categories. Semantic output response harmonization procedures were applied to ensure that all the variables are named the same across both data sets (i.e. NCRM and EHS) and to consolidate NCRM variables into a single one when needed. Numerical procedures were applied to floor area values as they were defined differently in NCRM and EHS. NCRM uses a building footprint or Gross Floor Area (GFA) whereas EHS uses Gross Internal Area (GIA) or usable floor area of self-contained dwellings [21]. To take into account intra and inter-regional differences from EHS, cross-scale harmonization was prepared according to the procedure used by [22].

$$A_f = A_i \times \frac{mla}{mna}$$

Where A_f is the ‘national computed area’ from EHS; A_i is the ‘local computed area’ from NCRM; mla is the mean GFA area based on residential density by type in a local authority and mna is the mean GIA area for each dwelling type by region and area type (as reported by the EHS surveyor).

2.1.3. Spatial interpolation algorithms

In order to achieve energy estimates from sub city aggregate areas it is important that each dwelling has an appropriate energy calculation. However, our high quality sample with all the relevant variables accounts for roughly 44% of the total residential stock. For those dwellings without a full record set, spatial interpolation techniques are used to estimate missing data from a nearby dwellings or dwellings.

The rationale behind spatial interpolation methods is that points close together in space are more likely to have similar values than points far apart [23]. A variety of spatial interpolation techniques were used to match dwellings: Nearest Neighbour (NN), Inverse Distance Weight (IDW), and Kriging. Each interpolation technique was selected to suit the urban density of the study area and was selected through a process of experimentation. NN was selected on the basis of a high residential density and sample penetration. IDW was selected on the basis of low residential density as observed values that were closer to the point of interest were more heavily weighted. Kriging was used to help to compensate for the effects of data clustering, assigning individual points within a cluster less weight than isolated data points i.e. treating clusters like single points for areas with a large number of multiple occupancy buildings such as student accommodation.

2.1.4. Best matching algorithm

NCRM requires an input record containing 115 variables based on a full SAP survey that is then used as input into a BREDEM energy model. Our best matching algorithm identifies candidate records in EHS that match the NCRM record of interest to create a full augmented record for each dwelling with all the 115 variables, including heating, lighting, appliances, and behavioural assumptions as per the identified EHS record. The ten energy variables defined earlier were separated into three ranked groups using principal component analysis and factor analysis. The three groups were:

- Group 1 (Physical properties 1) - floor area, dwelling type and number of floors
- Group 2 (Physical properties 2) - dwelling age, wall structure and wall insulation
- Group 3 (Heating systems and tenure) - main heating fuel, primary heating system, boiler group and tenure

The matching algorithm works by attempting to find a match of the group 1 variables. This is then further refined by matching with group 2 and subsequently

MLSOA	ΔP_{DGH}	ΔP_{HMO}	ΔP_{E7}	ΔP_{SC}	ΔP_A	Max Total
Castle	1.22 %	0.00 %	1.47%	0.94 %	0.33 %	3.96 %
South Heaton	6.00 %	6.19 %	4.57%	7.03 %	0.74 %	0.74 %
Westgate	14.55 %	0.87 %	6.61%	40.34 %	7.09 %	69.49 %

Table 2: Summary of model refinement in sub-city area results by district (MLSOA). (DGH = District Heating, HMO = Houses in Multiple Occupation, E7 = Economy 7, SC = Building Classification, A = Boundary issues)

group 3. As a result of the process, the algorithm will identify a single record within EHS (i.e. an exact match), more than one match; or zero matches. When the matching process yields more than one match, we utilise minimal Weighted Hamming Distance of dimension 10 to find the most appropriate EHS record. If the algorithm produces no matches, we treat the result as a singularity and it is dealt with as part of the model refinement process (see 2.2).

2.2. Model refinement

Given the huge variety in building stock in the city it is not surprising that the matching procedure failed to always match real properties with EHS data. In addition certain aspects of the modelling process cannot be automated in this way due to the nature of the property or special cases of the heating systems used. This section describes the key forms of empirical model refinement that were carried out on the data. In our results we report on their significance. Table 2 shows the number of adjustments made in each of our study areas.

2.2.1. Heating systems

DECC reports separately on Economy 7 heating (E7), a type of electrical storage heating system. We have treated E7 as a singularity as the location and spatial distribution of E7 dwellings in the city makes them susceptible to a lack of coverage by energy surveys such as Warmzone and this, in turn, has an impact on the spatial interpolation procedure. In Newcastle E7 properties are virtually all social housing properties and geographically clustered. For these areas we weighted tenure more strongly to better reflect the preponderance of E7 heating. Additionally, for some social housing we have manually allocated their main heating fuel and primary heating system e.g. for properties using group heating (i.e. a group of properties shared a boiler) and district heating (decentralised heating energy supply) using local knowledge.

2.2.2. Houses in multiple occupation

In our study areas (as is common in many urban areas) we have a significant number of houses in multiple occupation (HMO) using Newcastle City Council’s planning web portal. That is, properties rented out by at least 3 people who are not from the one ‘household’, but share facilities like the bathroom and kitchen. For our calculations, we treat HMOs as one dwelling from an energy service perspective instead of several self-included dwellings. This is likely to improve our calculation as an HMO does not have a kitchen per dwelling but rather shared communal facilities which impacts on cooking energy demand. Likewise, in an HMO bathroom facilities are also shared and therefore this has an impact on hot water heating. In our opinion it is likely that shared bathrooms, WCs, washing and kitchens facilities will have impact on heating and electricity consumptions and hence these are treated as a single dwelling.

2.2.3. Building classification

A number of issues were noted with certain categories of building classification from the base dataset. We utilised the building classification attributes from previous work in Newcastle [16] to identify mixed residential and commercial properties and high rise buildings. For high rise buildings, the initial floor area is the building footprint but in reality for a block of flats there are number of flats per floor. An estimate for each dwelling within a high rise building was attained through multiplying the building footprint by the number of storeys and dividing the total by the number of dwellings within the building.

2.2.4. Boundary problem

Boundary issues are a known problem with many spatial interpolation processes that use defined geographies (i.e. Census boundaries) to limit the extent of the analysis [24]. To overcome this issue we extended the boundaries to include the neighbouring spatial unit until there were enough data points. However, in one case, the boundary was a river and interpolation on these properties was augmented by manual checking.

2.3. Energy calculation

Once the complete dwelling dataset had been aligned with the best fit SAP record from the EHS, the final stage of the process is creating an energy consumption estimate (ECE) for each individual dwelling. For each dwelling the energy calculation was carried out based on the BREDEM model to acquire estimates of gas and electricity consumption. The ECE was then added to the NCRM record allowing for

ad-hoc energy aggregates across any spatial extent or other attribute. Here we aggregated using a bottom up model similar to [8]. Aggregate estimates of energy consumption were calculated by summing up individual energy values (i) for every dwelling (j) within the study reporting areas.

$$ETotal = (\sum_{j=1}^n \sum_{i=1}^n ECE_{i,j})$$

3. Validation and Analysis strategy

In order to validate and analyse our model estimates of domestic energy consumption at sub-city level we have used two approaches. The first approach uses spatial aggregates and the second uses property types. In both cases data was analysed in comparison with published DECC data on energy consumption.

Firstly, model outputs from NCRM aggregated to neighbourhood (LLSOA) and district (MLSOA) are compared with energy consumption figures for the same geographies from DECC. In total aggregates were calculated for 3 districts composed of 18 neighbourhoods. Economy 7 dwellings are reported separately as these are also treated separately in DECC’s consumption aggregates. This process is relatively straightforward although not without its own uncertainties as discussed in section 5.

Sub-city area	Dwelling group	Frequency (%)
Castle	House	90
	Bungalow	10
South Heaton	House	96
	Bungalow	4
Westgate	House	91
	Bungalow	9

Table 3: Representative dwelling subgroups in the selected sub-city areas.

The NEED dataset disaggregates gas heating consumption by three variables: dwelling type; dwelling age; and floor area and our model utilises 10 variables to produce disaggregated results. Thus, the only conclusion that can be drawn at an aggregate level is to determine whether our results are within the distribution provided by NEED. Our results tables (see section 4.2, 10, 12, and 14) show NEED mean and percentiles (i.e. first, second (Median) and third quartiles, and 5th, 10th, 90th and 95th percentiles) which are used to test whether they are equal to those obtained from our NCRM samples. A collection of 11 test samples were

Sub-city area	Dwelling type	Frequency (%)
Castle	House semi-detached	62
	House mid-terraced	17
	House detached	11
	House end-terraced	10
South Heaton	House mid-terraced	81
	House end-terraced	12
	House semi-detached	7
	House detached	0
Westgate	House mid-terraced	65
	House end-terraced	25
	House semi-detached	9
	House detached	1

Table 4: Representative dwelling subgroups by dwelling type in the selected sub-city areas.

Sub-city area	Sample number	Dwelling type	Dwelling age	Frequency (%)
Castle	1	Bungalow	1965-82	48
	2	Bungalow	1945-1964	40
	3	Semi-detached	1945-1964	21
	4	Semi-detached	1965-82	69
	5, 6	Detached	1964-1979	87
Westgate	7	End-Terraced	Pre 1919	56
	8, 9	Mid-Terraced	Pre 1919	40
SouthHeaton	10	End-Terraced	Pre 1919	67
	11	Mid-Terraced	Pre 1919	56

Table 5: 11 samples selected. The samples are those with the highest dwelling age frequency distribution.

devised to make inferences about our domestic stock in the selected study areas of which a representative 3 are reported in the results section. Because our NCRM high quality sample (i.e. without the interpolated properties see 2.1) is a stratified random sample, the strata should adequately represent the population at large [25]. In our case, as a general rule, we selected samples from those dwelling types with the highest frequency distribution of the overall population in each MLSOA. For Castle the representative subgroups are bungalows, semi-detached and detached houses; for South Heaton and Westgate mid-terraced and end-terraced (see tables 3 and 4). Furthermore, within the selected dwelling type the samples are those with the highest dwelling age frequency distribution (see table 3). In Castle detached houses and Westgate mid-terraced were split further to match NEED categories (see table 5. Castle number 5 and 6. Westgate number 8 and 9).

4. Results

Our model provides estimations of electricity and gas consumption at sub-city level. In this section, we first report on spatial aggregates at district and neighbourhood scales: three selected districts and their associated neighbourhoods (18 in total) are compared to DECC data (all results have been rounded up or down to the nearest full percentage). Secondly, we report on dwelling type comparisons between our model estimates and NEED reported data. Our results perform consistently against both of the DECC data sets and appear to indicate that having reliable knowledge and data of the local stock and urban geography factors may benefit energy consumption estimations in a significant way.

4.1. District and Neighbourhood aggregates

Comparing the *cluster-based* and *sub-city Domestic Energy Model* aggregated estimations with DECC national statistics at district level (MLSOA) confirms the capability of our model in providing good energy consumption estimates. At district level, the NCRM cluster model estimation of combined gas and electricity consumptions for Castle, South Heaton and Westgate are respectively 5% lower, 3% higher and 27% higher than DECC statistics (see table 6 and figure 3). The NCRM sub-city DEM estimation of gas and electricity consumptions for Castle, South Heaton and Westgate are respectively 25% lower, 16% higher and 50% higher than DECC statistics (see table 7 and figure 4). One possible reason for the differences in the results of the two modelling strategies is how dwelling data is organised into meaningful structures in the cluster model, that is, how the taxonomies were developed in

MLSOA	2009 Model estimation			2009 DECC estimation			2009 Differences		
	Electricity (Kwh/yr)	Gas (Kwh/yr)	Total (Kwh/yr)	Electricity (Kwh/yr)	Gas (Kwh/yr)	Total (Kwh/yr)	Electricity %	Gas %	Total %
Castle	16,213,837	55,654,385	71,868,222	15,184,682	60,287,546	75,472,228	7	-8	-5
South Heaton	15,267,453	61,412,232	76,679,685	14,228,799	60,075,526	74,304,325	7	2	3
Westgate	22,787,300	72,449,103	95,236,403	27,211,294	47,918,836	75,130,130	-16	51	27

Table 6: Cluster model. Summary of sub-city area results by district (MLSOA).

the clusters. The cluster model uses a limited number of eight clusters made up of six key energy variables (dwelling age, wall construction, building form, dwelling size, heating and number of floors) which tend to average out energy consumption over large number of properties. Although the cluster model performs reasonably well (i.e. at district level, table 6) due to the averaging, it does not take into account, local characteristics of the building stock, socio-economic data (e.g. tenure) or surrogates (e.g. the settings of internal temperature, which closely relates to occupancy) and urban geography which we believe play an important part in determining energy consumption. Furthermore, we believe that the effect of local area characteristics on estimating dwelling energy consumption is amplified at higher spatial resolutions. Table 8 shows our NCRM cluster results at LLSOA.

Figure 3: Cluster model. Summary of sub-city area results by district (MLSOA).

MLSOA	2009 Model estimation			2009 DECC estimation			2009 Differences		
	Electricity (Kwh/yr)	Gas (Kwh/yr)	Total (Kwh/yr)	Electricity (Kwh/yr)	Gas (Kwh/yr)	Total (Kwh/yr)	Electricity %	Gas %	Total %
Castle	11,675,295	44,646,725	56,322,020	15,184,682	60,287,546	75,472,228	-23	-26	-25
South Heaton	17,085,525	69,005,615	86,091,140	14,228,799	60,075,526	74,304,325	20	15	16
Westgate	24,111,129	88,752,436	112,863,565	27,211,294	47,918,836	75,130,130	-11	85	50

Table 7: Sub city Domestic Energy Model. Summary of sub-city area results by district (MLSOA).

Figure 4: Sub city Domestic Energy Model. Summary of sub-city area results by district (MLSOA).

		2009 Model estimation			2009 DECC estimation			2009 Differences		
MLSOA	LLSOA	Electricity (Kwh/yr)	Gas (Kwh/yr)	Total (Kwh/yr)	Electricity (Kwh/yr)	Gas (Kwh/yr)	Total (Kwh/yr)	Electricity %	Gas %	Total %
Castle	8306	2,549,395	10,128,714	12,700,883	2,896,451	12,512,795	15,409,246	-11	-19	-18
	8307	1,979,446	7,648,926	9,628,372	1,471,463	6,419,991	7,891,454	35	19	22
	8308	1,849,458	6,394,362	8,243,820	1,950,176	7,995,921	9,946,097	-5	-20	-17
	8311	2,593,335	8,423,198	11,016,533	2,694,528	9,372,288	12,066,816	-4	-10	-9
	8312	2,080,978	7,392,029	9,473,007	1,928,826	7,127,555	9,056,381	8	4	5
	8313	2,007,101	6,880,271	8,887,372	1,833,166	6,794,391	8,627,557	9	1	3
	8294	2,468,447	7,972,711	10,441,158	2,410,072	9,323,397	11,733,469	2	-14	-11
South Heaton	8303	5,100,328	11,752,129	16,852,457	3,062,105	7,326,969	10,389,074	67	60	62
	8359	1,993,563	7,952,238	9,945,801	2,153,830	9,644,712	11,798,542	-7	-18	-16
	8360	2,816,385	11,227,861	14,044,246	2,271,434	10,675,756	12,947,190	24	5	8
	8361	2,346,065	9,477,195	11,823,260	2,134,881	9,860,023	11,994,904	10	-4	-1
	8362	2,604,546	9,547,510	12,152,056	2,286,790	10,700,216	12,987,006	14	-11	-6
	8364	2,566,959	9,728,681	12,295,640	2,319,759	11,554,064	13,873,823	11	-16	-11
Westgate	8394	2,275,082	5,226,193	7,501,275	649,562	3,048,655	3,698,217	250	71	103
	8395	2,646,420	9,516,472	12,162,892	1,739,168	7,989,220	9,728,388	52	19	25
	8397	3,433,999	6,459,947	9,893,946	3,667,679	5,838,231	9,502,910	-6	-11	4
	8399	3,606,120	11,903,463	15,509,583	2,349,583	6,329,353	8,678,936	53	88	79
	8439	3,676,307	10,856,731	14,533,038	2,556,256	8,661,471	11,217,727	44	25	30
	8440	8,544,544	26,695,881	35,240,425	16,249,046	15,848,757	32,097,803	-47	68	10

Table 8: Cluster model. Summary of sub-city area results by neighbourhood (LLSOA).

At neighbourhood level (i.e. a higher spatial resolution using census LLSOA geographies) when comparing the NCRM cluster model and sub-city DEM estimation with DECC national statistics the DEM model provides reasonable estimates but with greater divergence (see table 9 and figures 5, 6 and 7). Since our energy estimations are derived from a dataset which ultimately corresponds to individual houses with known locations, we believe that as spatial resolution is increased local building, socio-economic and physical characteristics of the urban area play a more important part in the estimation of dwelling energy consumption. For example, in the NCRM sub city DEM estimates for low density areas like Castle, the microclimate factors (such as wind and solar) could be important. This may mean that a low residential density, which are synonymous with more open space, require more energy to maintain the same temperature as higher density areas. Our local building data also shows that most of the Castle stock uses gas for space heating. The combination of these two factors, may in turn, explain the under-estimation of gas consumption in our results for Castle which are on average 26% lower than DECC statistics at district level (see table 7 and figure 4) with a range of -16% to -33 % at neighbourhood level (see table 9 and figure 6).

In South Heaton, the implications of compact densification in linear terraces using gas for space heating potentially highlights the importance of thermal zones in energy estimates. NCRM (due to the limitations of the CHM) ignores the heat

		2009 Model estimation			2009 DECC estimation			2009 Differences		
MLSOA	LLSOA	Electricity (Kwh/yr)	Gas (Kwh/yr)	Total (Kwh/yr)	Electricity (Kwh/yr)	Gas (Kwh/yr)	Total (Kwh/yr)	Electricity %	Gas %	Total %
Castle	8306	2,391,052	10,518,903	12,909,955	2,896,451	12,512,795	15,409,246	-17	-16	-15
	8307	1,376,298	5,727,723	7,104,021	1,471,463	6,419,991	7,891,454	-6	-11	-10
	8308	1,425,907	5,938,554	7,364,461	1,950,176	7,995,921	9,946,097	-27	-26	-26
	8311	1,939,494	6,851,613	8,791,107	2,694,528	9,372,288	12,066,816	-28	-27	-27
	8312	1,398,933	4,954,461	6,353,394	1,928,826	7,127,555	9,056,381	-27	-30	-30
	8313	1,343,307	4,397,440	5,740,747	1,833,166	6,794,391	8,627,557	-27	-35	-33
South Heaton	8294	1,800,304	6,258,031	8,058,335	2,410,072	9,323,397	11,733,469	-25	-33	-31
	8303	4,377,180	9,699,069	14,076,249	3,062,105	7,326,969	10,389,074	43	32	35
	8359	2,131,968	10,919,221	13,051,189	2,153,830	9,644,712	11,798,542	-1	13	11
	8360	3,006,039	12,749,411	15,755,450	2,271,434	10,675,756	12,947,190	30	19	22
	8361	2,386,341	11,324,933	13,711,274	2,134,881	9,860,023	11,994,904	12	15	14
	8362	2,561,513	11,811,979	14,373,492	2,286,790	10,700,216	12,987,006	12	10	11
Westgate	8364	2,622,484	12,501,002	15,123,486	2,319,759	11,554,064	13,873,823	13	8	9
	8394	727,443	2,382,910	3,110,353	649,562	3,048,655	3,698,217	12	-22	-16
	8395	2,750,815	11,714,160	14,464,975	1,739,168	7,989,220	9,728,388	58	47	49
	8397	2,444,095	8,127,934	10,572,029	3,667,679	5,838,231	9,502,910	-33	39	11
	8399	5,176,186	18,152,00	23,328,186	2,349,583	6,329,353	8,678,936	120	187	169
	8439	2,960,014	10,596,295	13,556,309	2,556,256	8,661,471	11,217,727	16	22	21
	8440	10,052,576	37,779,137	47,831,713	16,249,046	15,848,757	32,097,803	-38	138	49

Table 9: Sub city Domestic Energy Model. Summary of sub-city area results by neighbourhood (LLSOA).

flow though from one terrace to the adjacent terrace and this, in turn, could be the explanation for over-estimating energy gas consumption in South Heaton, a densely populated area with a high proportion of linear terraced properties. The NCRM sub-city DEM estimation of gas consumptions for South Heaton is on average 15% higher than DECC statistics at district level (see table 7 and figure 4) with a range of 8% to 32 % at neighbourhood level (see table 9 and figure 6).

In a heterogeneous area like Westgate the energy consumption shows large variations. The total energy consumption at district level is over-estimated by 50% with an underestimation of electricity of 11 % and overestimation of gas of 85% (see table 9 and figures 6 and 5). At neighbourhood level there are large differences. In 8399 gas is overestimated by 187% and electricity by 120%. In 8394 gas is underestimated by 22% whereas in 8397 is overestimated by 39% (see table 9 and figure 6). These large disparities could be attributed to a very specific type of urban geography which does not fit well neither the English Housing Survey data nor the Cambridge Housing Model and that requires a large amount of model refinement (see table 2). The most important factors which contribute to refinement are related to a) building type, with the presence of large tower blocks; b) energy system configuration. A significant proportion of the dwellings in this area rely on communal heating systems (e.g 8397 and 8399 used group heating as reflected by the low number of gas meters

Figure 5: Sub city Domestic Energy Model. Summary of sub-city area electricity results by neighbourhood (LLSOA).

Figure 6: Sub city Domestic Energy Model. Summary of sub-city area gas results by neighbourhood (LLSOA).

Figure 7: Sub city Domestic Energy Model. Summary of sub-city area total energy consumption results by neighbourhood (LLSOA).

in DECC’s statistics [26]) and there is a significant amount of E7 in 8397 and 8440. There are socio-demographic factors at play as there is a significant proportion of social housing across the district. In our sub city DEM, social-demographic factors are generally represented in a crude way due to model limitations (i.e. CHM is based on the BREDEM physical model). This means, for instance, that energy consumption estimations are based on standard occupancy rather than how an individual uses the property and appliance use can vary significantly between users. This can have an important effect on electricity consumption.

4.2. Gas consumption across dwelling types

Our results by dwelling type (see tables 10, 12, and 14) suggest that reported gas domestic consumption for UK North East properties [27] may underestimate gas heating consumption when compared to NCRM sub city DEM output as mean and median heating consumption are below NEED’s values. Our NCRM sub city DEM heating gas estimates do not have the same skewed distribution that NEED data has, as shown in the columns - Mean and Median difference - giving a strong indication that the local area characteristics matter in micro-planning rather than regional averages. Tables 10, 12, and 14 also show the one-sample Wilcoxon signed

ranks, which are used to test whether the mean of the measurements (NCRM sample) are equal to the NEED mean. The Wilcoxon signed ranks illustrate different values under negative and positive rank columns meaning that the NCRM sample does not match the North-East England average composition which is a consequence again of local area characteristics. Because the NEED data has a skewed distribution, it is preferable to provide additionally to the mean and the median (i) two outer centiles, such as the 10th and 90th; (ii) the first and third quartiles (25th and 75th centiles) that define the interquartile range and; (iii) the range of the sample, the 5th centile and the 95th centile. This has been done for all the samples and in this paper we provide three indicative examples (see figures 8 9 and 10) .

For Castle, table 10 shows that mean and median NCRM values are higher than those of NEED. We believe that local area characteristics play an important part in the sample's energy efficiency. In our Castle sample, bungalows are largely made up of local authority stock and are mostly uninsulated and use standard and combi boilers. The combination of all these elements probably explains the underestimation of gas consumption (see figure 8 and table 11 to compare NEED and NCRM annual gas consumption distributions).

Table 10: Castle and NEED. Comparing heating gas consumption by property type

Sample	Property Type	Property Age	Floor Area (sqm)	CASTLE			NEED			Difference		Wilcoxon Signed Ranks		
				Sample Size	Annual Mean (Kwh)	Annual Median (Kwh)	Sample Size	Annual Mean (Kwh)	Media Mean (Kwh)	Mean %	Median %	Negative Ranks	Positive Ranks	Ties
1	bungalow	1965 to 1982	76 to 100	55	16,495	18,190	1,400	13,052	12,557	26	45	23	32	0
2	bungalow	1945 to 1964	50 to 75	37	12,625	11,764	3,020	10,905	10,326	16	14	24	14	0
3	semi-detached	1945 to 1964	50 to 75	203	15,158	13,118	950	11,070	10,574	37	24	117	89	0
4	semi-detached	1965 to 1982	50 to 75	628	13,951	10,772	290	10,161	9,748	37	11	367	282	0
5	detached	1965 to 1982	76 to 100	100	18,112	18,599	890	13,796	13,218	31	41	27	73	0
6	detached	1965 to 1982	101 to 125	18	19,431	20,897	1,410	15,365	14,870	26	41	4	14	0

Table 11: Castle. House bungalow. Sample 1 1965 to 1982 - 76 to 100 sqm

Annual Gas	Mean	Sample	Percentiles						
			5	10	25	50	75	90	95
NCRM Kwh	16,495	55	9,146	10,772	13,223	18,190	18,599	22,287	22,287
NEED Kwh	13,052	1400	6,361	7,765	9,913	12,557	15,613	19,000	21,148
Difference %	26		44	39	33	45	19	17	5

In South Heaton local area characteristics significantly shape domestic annual gas consumption. Table 12 shows that our model's mean and median are higher than NEED's. When we examine further our dataset by drilling down to individual properties, it shows that there are a high number of low efficiency dwellings. These

Figure 8: Castle. House bungalow. Sample 1 1965 to 1982 - 76 to 100 sqm

dwellings have un-insulated solid wall (i.e the ratio of uninsulated to insulated solid wall properties in the complete city sample is $258/7,396 = 0.035$, in South Heaton sample it is $12/998 = 0.012$) and has a high proportion of HMOs as South Heaton is a student area in which houses and flats have been converted to cater for a growing demand. Figure 9 and table 13 provide an example of the mismatch between NCRM and NEED distribution.

Table 12: South Heaton and NEED. Comparing heating gas consumption by property type

Sample	Property Type	Property Age	Floor Area (sqm)	SOUTH HEATON			NEED			Difference		Wilcoxon Signed Ranks		
				Sample Size	Annual Mean (Kwh)	Annual Median (Kwh)	Sample Size	Annual Mean (Kwh)	Media Mean (Kwh)	Mean %	Median %	Negative Ranks	Positive Ranks	Ties
7	End-terrace house	Pre 1919	50 to 75	39	16,084	13,919	150	11,483	10,905	40	28	1	38	0
8	Mid-terrace house	Pre 1919	50 to 75	381	14,776	13,919	1380	9,500	8,757	56	59	17	368	0
9	Mid-terrace house	Pre 1919	76 to 100	150	17,903	10,657	6050	11,070	10,657	62	34	3	148	0

Percentiles										
Annual Gas	Mean	Sample	5	10	25	50	75	90	95	
NCRM Kwh	17,903	150	13,919	13,919	13,919	14,275	22,287	22,287	34,823	
NEED Kwh	11,070	6,050	3,800	5,204	7,683	10,657	13,796	17,431	19,744	
Difference %	62		266	167	81	34	62	28	76	

Table 13: Sample 9 South Heaton House mid-terraced pre 1919 - 76 to 100 sqm.

Figure 9: South Heaton mid-terraced house. Sample 9 pre 1919 - 76 to 100 sqm

Table 14: Westgate and NEED. Comparing heating gas consumption by property type

Sample	Property Type	Property Age	Floor Area (sqm)	WESTGATE			NEED			Difference		Wilcoxon Signed Ranks		
				Sample Size	Annual Mean (Kwh)	Annual Median (Kwh)	Sample Size	Annual Mean (Kwh)	Media Mean (Kwh)	Mean %	Median %	Negative Ranks	Positive Ranks	Ties
10	End-terrace house	1965 to 1982	50 to 75	49	16,101	17,901	180	9,418	9,252	71	93	1	38	0
11	Mid-terrace house	1965 to 1982	50 to 75	106	15,285	17,306	340	8,509	8,178	80	112	16	91	0

Percentiles										
Annual Gas	Mean	Sample	5	10	25	50	75	90	95	
NCRM Kwh	15,285	106	7,297	7,297	13,118	17,306	17,901	19,313	21,926	
NEED Kwh	8,509	340	3,470	4,461	5,948	8,178	10,491	12,557	14,044	
Difference %	80		110	64	121	112	71	54	56	

Table 15: Sample 11 Westgate House mid terraced 1965 to 1982 - 50 to 75 sqm .

Finally, table 14 also shows how in Westgate the mean and median are higher than those in NEED. Similarly, when we examine local individual dwelling records, it appears that there is a large number of inefficient houses (i.e. no cavity insulation, electric heating supply) which are local authority owned. As before, we believe that the combination of these local area characteristics may explain the disparity between our model's and NEED distributions (see figure 10 and table 15 as a way of example).

Figure 10: Westgate mid-terraced house. Sample 11 mid terraced 1965 to 1982 - 50 to 75 sqm

5. Discussion

Our work demonstrates that sub-city energy estimation with a convenience building sample is technically possible and, subsequently, we could make local energy planning recommendations such as area based or dwelling type targeted energy efficiency measures or renewable energy systems. For example, in South Heaton, in general, properties are dominated by un-insulated solid wall dwellings and are rented accommodation and therefore a possible energy efficiency measure would be to reduce the air tightness in HMO properties. Similarly, community renewable heating schemes at neighbourhood level in large portions of E7 and electric heating systems in South Heaton and Westgate with solar collection schemes in Castle could be seen as a possibility. However, there is a poor agreement between our results and current DECC data, particularly at LLSOA level. We believe that as spatial resolution is increased local building and urban socio-economic and physical characteristics play a more important part in the estimation of dwelling energy consumption. Thus, we would be concerned with the “uncertainty” surrounding those recommendations. Whilst there are ways of generating data, real progress in local energy modelling depends on greater data availability and more accurate energy modelling estimation. In practice this means that statistical data, spatial interpolation techniques and substantive energy model issues overlap making local energy estimations more uncertain as spatial resolution is increased. Whilst it is clear that there is a need for high quality socio-economic and physical dwelling data sets, this is only one aspect of an emerging complex picture surrounding local energy modelling. In this section

we therefore discuss the uncertainty surrounding our sub-city energy estimations in a holistic manner to set the basis for future work in this area and present a conceptual taxonomy of our model’s uncertainty.

5.1. Uncertainty: a conceptual taxonomy

Uncertainty when developing urban energy models is an important and pervasive aspect that it is attracting attention, given the growing need for evidence-based policy, shared decision making, and highly disaggregated modelling tools. Whilst a full uncertainty analysis of our model is outside the scope of this paper, it is important to address uncertainty to provide transparency when communicating our modelling framework and predictions to other local authorities or wider stakeholders (such as national policy-makers, data holders, and researchers) interested in replicating our work. The principal value of the taxonomy we propose and examine in this section is as a way of more precisely defining the phenomenon of uncertainty so that it can be better communicated, measured and analyzed. Our taxonomy follows [28]’s framework and conceptualises uncertainty as “a multi-dimensional phenomenon with theoretically distinct domains and constructs that are potentially measurable and related to different outcomes, mechanisms of action, and management strategies” [28, p.9]. This is a departure from past studies which have either ignored uncertainty or have narrowly focused on uncertainties pertaining to particular sources or issues such as those proposed in the Cambridge Housing Model [9]. Figure 11 proposes a three dimension integrative taxonomy of the uncertainty [28, p.7] (i) the location (source); (ii) the level (issues); and (iii) the nature (locus). Location (source) dimension of uncertainty relates to where the uncertainty manifests within the complex model; (ii) the level (issues) dimension of uncertainty relates to where the substantive issues of uncertainty manifests along the spectrum between deterministic knowledge and total ignorance; (iii) the nature (locus) dimension of uncertainty relates to whether the uncertainty is due to the imperfection of our knowledge (structural (epistemic) uncertainty) or is due to the inherent variability of the phenomena being described (parametric (aleatory) uncertainty)[29]. The model outcome uncertainty in figure 11 is the accumulated uncertainty caused by all of the locations (context, model, imputation, and calibration parameters) that are propagated through the model’s issues and are reflected in the resulting estimates of annual energy consumption of interest.

5.1.1. Context

Context refers to the conditions and circumstances that underlie the choice of the boundaries of the system, and the framing of the concepts and terminology of the

Figure 11: Uncertainty integrative taxonomy: a NCRM multi-dimensional framework.

research question to be addressed within those boundaries. In our case, for instance, we have made a selection on the type of analysis (e.g. quantitative), the type of energy model (e.g. a bottom-up engineering model based on physical parameters: BREDEM/SAP.), and specific choices in variables boundaries such as daily standard occupancy (i.e. number of occupants are related to and standard daily heating regimes energy patterns -e.g. weekly 7am to 9am; 4pm and 9pm for living room areas as defined by BREDEM 12-).

Perhaps, more important when considering context in local energy modelling is understanding that our research questions are naturally driven by the spatial dimension of cities. Specifically, local authorities deliver their strategic planning, policies and implementation programmes using area-based approaches at three major scales: city-wide, district and neighbourhoods, and individual buildings. Given these options, local authorities are faced with the question: at what spatial scale should area-based energy strategies be planned, and are the proposed solutions consistent across scale? However answering this question introduces a well-known problem of geography: the modifiable areal unit problem. First identified by [30], and discussed at length by [31], the modifiable areal unit problem refers to the observation that spatially distributed data will yield different results depending on how it is aggregated. In essence, data sets are compiled for “non-modifiable” units or entities (e.g. households) whereas subsequent analysis is carried out using arbitrary and modifi-

able areal units (e.g. districts, wards, local authorities) with no intrinsic geographical meaning. The question, as formulated by Openshaw, was: “does it matter? If you change the areal basis does it have any really significant effect on the results?” His study concluded that the underlying issues of aggregation and scale variability affected different forms of analysis (i.e. correlations or complex multivariate methods). As a result, “an ecological fallacy occurs when it is inferred that results based on aggregate zonal (or grouped) data can be applied to the individuals who form the zones or groups being studied”. For instance, in our domain, a Lower Layer Super Output Area (LLSOA) contains a relatively small (minimum 400 households) and, in principle, homogenous group of households. Popular data sets in this field – such as HEED (i.e. energy efficiency measures installed in individual buildings) and DECC energy consumption statistics – have been compiled at LLSOA resolution. It is feasible to imagine that an LLSOA could be used and seen as the most spatially resolved areal unit of analysis for area-based energy strategy, analysis and delivery (e.g. a retrofit delivery initiative such as Warm-up North [32]). However an LLSOA-based analysis could conclude that a particular area is fit for a building-centric energy retrofit strategy, disregarding any potential decentralised energy solutions which might be more suited for a cluster of individual buildings. Our interest is therefore to better understand how methodologies for spatial energy studies might be “explicitly based on and around the purposeful and deliberate engineering of zoning systems” [31, p.37] and see context as epistemic uncertainty, the uncertainty due to the imperfection of our knowledge, which may be reduced by more research and empirical efforts.

5.1.2. Model

In model, there are two major categories of uncertainty (i) model structure uncertainty, and (ii) model technical uncertainty. Model structure uncertainty arises from a lack of sufficient understanding of the formulation used to represent the system, implying that any one (of many model formulations) might be a plausible representation of the system. Model technical uncertainty is the uncertainty generated by software or hardware errors. In our case model structure uncertainty has been framed by the work of [9] on the Cambridge Housing Model and we see this as ontological uncertainty. [9, p.166] uncertainty analysis results suggest that any single estimate of sub-national energy use may be subject to considerable inaccuracy compared with actual use.

Our interpretation of the NCRM energy results suggests that a better understanding of local area characteristics is needed as to aid the performance of the model. These relate to:

- Building related data and thermal modelling

- Urban geographical modelling:
 - Socio-economic
 - Landscape
 - Micro-climates

Building related data refers to the building geometry and building fabric along with its key heating systems. Urban socio-economic characteristics refers to households realistic characteristics: demographics, economic structure, state of social energy practices in the local area. Urban landscape characteristics refers to cohesive energy structures (HMO, district heating and group heating) and urban form (morphology) whereas micro-climate relates to all climatic related variables such as wind and solar. All these areas are interlinked.

Currently, NCRM should be seen as a building physic domestic energy model (DEM). DEMs have seen major methodological advances incorporating technological changes in building fabric components and heating supply systems, but far less on building thermal modelling (for local energy models, not just individual buildings and urban geographical issues such socio-economic characteristics, microclimates and building form). We believe that number of thermal zones have been simplified due to the BREDEM/SAP based nature of our model and could explain some of the overestimation of energy consumption in South Heaton which have a high number of liner terrace of converted flats. Similarly, if complexity in behaviour is not explicitly and coherently addressed in DEMs, this can lead to energy estimations that are based on incorrect individual (and aggregate) practices over a population or to ignore important specific practices by different population segments. We also believe that urban landscapes and outdoor climate conditions have a clear influence in energy consumption but the exact characterization of this link is difficult to model due to the extreme morphological heterogeneity at a fine granularity of the building and its proximate environment. For instance, our understanding of our case study areas suggests that the energy implication of compact densification (in South Heaton) is balanced between the benefits from reduced heat losses and the non-benefits from reduced solar and daylight availability. This is the opposite in Castle, a low density urban form, in which the microclimate effects may include (i) alteration of outdoor comfort conditions due to localized wind patterns, (ii) modification of the thermal balance of buildings, (iii) systematic recourse to artificial lighting inside these same buildings due to shading of neighbouring buildings. This creates a complex set of interactions between urban form, climate, buildings (fabric and heating supply systems) and urban activity and suggests a search for a new interdisciplinary analytical

direction by introducing a hybrid framework to energy consumption. This hybrid approach would require a robust spatial home energy-use data and model.

In our view, the above examples illustrate that “the entities and relations of which propositions would have been composed are *simply not known* at the time the propositions would have to be formulated” [33, p.10]. As a result, we classify uncertainties surrounding the underlying energy model structure as ontological [33]. For example, one possible approach to include social practice in NCRM data is by capturing some elements as life-cycle and dynamics between practices depending on the practice elements, the behaviour of different cohorts with respect to energy practices and the environments that encourage or discourage practices. Social practice theory, in contrast to physical models, encourages a household shifted self-understanding and it invites us to regard them as agents or carriers of routinised usage of things. The challenge for future NCRM research could be to find a unifying ontology and database to describe and store core common data objects.

5.1.3. Imputation

Imputation uncertainties are generated when converting data between different scales and geographies to fill in missing data in small geographical areas. In other words, the uncertainties were generated when multiple imputations from national data sets (e.g. EHS) were used to complete full SAP records.

All imputation processes have been explained in detail previously (see section 2) and rely on integrating different sources. For the purpose of the discussion, we will analyse one in more detail. Harmonization is a key process to merge different datasets and it is being made particularly difficult by the lack of acknowledged standard approaches to, for instance, deal with a critical variable such as building classification. In our case, Cities Revealed age definitions do not align with the EHS age bands. The NCRM building type statistics shows semi-detached properties (43%), mid-terraced (32.2%), end-terraced (13.2%), detached (6.5%) enclosed (4.2%) whereas EHS dwelling type shows end-terrace (11.4%), mid-terrace (20%), semi-detached (28%), detached (18.9%), purpose built flat (18.2%), converted to flat (3.4%). First thing to notice is that the local and national profiles are quite different. Secondly, there is a building type categories mismatch in both datasets. On the one hand, in NCRM there is no indication of “purpose built flat” or “converted to flat” whereas that in EHS there is not a category for enclosed flats. Part of the harmonisation process was to create a “purpose built flat” and “house converted to flats” type in NCRM as part of the mid-terraced count and assume enclosed flats were part of the EHS mid-terraced count. This way both datasets could be integrated and harmonised. This is particularly important in our case study areas as the most

common purpose built flat are the ‘Tyneside flats’ which was, in turn, , difficult to match with an EHS record. Tyneside flats are terraced buildings comprising pairs of self-contained flats and are below the national average in terms of space per dwelling, have different wall types (the main fabric is solid but the extensions are cavity) and are in districts within easy access to Newcastle and Northumbria Universities which means the original nineteenth century pattern of renting most Tyneside Flats still continues. Furthermore, in recent years this has led to the conversion of many lofts in upper Tyneside Flats to provide extra student bedrooms. In this case care has to be taken to update the storey height which should be an average based on the volume of the space and the internal floor area (plus the thickness of the floor if it is the upper storey of a house). Secondary information such as HMO planning permission has to be utilised to account for these changes in dwelling structure. These specific characteristics of a Tyneside flat have implications in terms of estimating its energy consumption and were reflected in each individual Tyneside flat record. Moreover, Cities Revealed individual building type and previous classification work [16] had to be used complementarily as to infer the age field and the most probable number of residential dwellings. Some building type classes are distinctive, such as the difference between detached and semi-detached but other classes are more subjective, such as the difference between small, large and very large detached houses.

In our view, imputation uncertainties could be categorised as small area estimation uncertainties [34]. Small area estimation “tackles the important statistical problem of providing reliable estimates of a target variable in a set of small geographical areas”. The main obstacle is that it is extremely difficult to “measure the target variable of all the individuals in the areas of interest and, hence, a survey is conducted to obtain a representative sample” [35, p.1].

5.1.4. Validation

Our model’s estimation have been validated against DECC’s sub-national energy consumption statistics (MLSOAs and LLSOAs) at different “statistical” geographies [2] and the DECC’s national energy efficiency data framework (NEED) [36].

The recent booklet published by DECC [2] on the methodology used to provide MLSOA and LLSOA domestic electricity and gas estimates is transparent with the regard to the caveats surrounding the released data. For instance, the sum of meter points of domestic energy consumption at MLSOA and LLSOA level does not always equal the sum of meter points of domestic energy consumption at the associated LA level. This type of caveat might not critically affect the outcomes of our model. Others, however, require closer inspection. Firstly, DECC regards LLSOA data as experimental [2, p.39]. DECC does not provide much information on the

meaning of experimental but it warns that “year -on- year analysis should be done with caution” [2, p.11]. Our experience suggests that areas such as number of meters, non-metered fuels, arbitrary gas consumption levels being regarded as domestic and E7 penetration need improvement. Secondly, DECC gas sub-national consumption figures have been weather corrected, whereas electricity consumption figures are estimates of actual consumption [2, p.10]. The DECC annualised and weather corrected Meter Point Reference Numbers (MPRN-level) gas consumption data are obtained from XoServe. The National Grid, however, apply a weather correction to the data prior to it being supplied to XoServe and DECC. Although a certain amount of information relating to the DECC process of annualisation and weather correction is available [37, p.28] the effect that this process has on modifying the gas values, and therefore the impact that the factoring has on the final results of analysis, is not fully known. We agree with [38] and believe that process of weather correction is impossible to reverse based on information currently in the public domain and introduces unknown effects into the available annualised records for each individual house that cannot be interrogated. In our case, we have only applied a weather correction factor to account for the correct year. When we carried out the analysis, NEED reference data was only available for the year 2010 whereas our NCRM model energy baseline estimates refer to 2009. To account for this, we multiply NEED’s values by the appropriate weather correction factor (i.e. as outlined by [39]) so that the NCRM and NEED heating gas consumption samples can be compared. This resulted in a weather correction factor of 0.8261 to be applied to NEED data.

In this paper, we propose that calibration is seen as parametric aleatoric uncertainty. The government data framework matches gas and electricity consumption data, collected for DECC sub-national energy consumption statistics, with information on energy efficiency measures installed in buildings, from the Homes Energy Efficiency Database (HEED). We think NCRM challenges this standard procedure by underpinning the argument for more complete and better data as the paucity of survey data and the mixed energy and building portfolio means that the model does not have enough data granularity to perform adequately. i.e. if the same measure is applied to the same property type, the energy consumption response might be different depending of the energy profile of the building.

6. Conclusion

We have presented a modelling framework for estimating domestic energy end-use demand baseline in sub-city areas and we believe that it could be replicated by other LAs and could be used to make local energy planning recommendations. However,

our validation results show a poor alignment with existing observed data as published by DECC. Our analysis seems to suggest that local area characteristics are important when systematically establishing energy baseline consumption for specific sub-city geographical areas. Subsequently, we believe that current UK Government regional and sub-city methods and data for domestic properties in its current most disaggregated form may not accurately represent energy consumption of geographically specific and homogenous urban areas in the UK and therefore be insufficient for providing evidence for meeting future challenges in planning local energy services and infrastructure. Our analysis also shows that there is a significant number of uncertainties which are not usually communicated and understood by LAs, policy development of national datasets, and other stakeholders such as planners, architects and engineers. To better understand, communicate and describe uncertainties, it is necessary to obtain a detailed local knowledge of the stock and non-filtered building or post code level consumption data. Further model validation to ascertain uncertainties in the CHM model, imputation algorithms, and DECC data is certainly needed. In our paper, we go even further to suggest that a re-think of underlying energy models to enable the integration of building and urban modelling challenges is necessary.

Acknowledgements

We would like to thank Newcastle University, Science City and Newcastle City Council for their financial and in kind support. Without that, this research would not have been possible. We would also like to thank early inputs from Prof. Tom Wagner, Mr David Alderson, and Mr Tom Bradley.

References

- [1] S. K. Firth, K. J. Lomas, M. Quddus, D. Allinson, C. Wang, S. Taylor, Disaggregated scenarios for demand studies, Website (July 2011).
URL <http://ds4ds.org/>
- [2] U. Government, Sub-national consumption statistics: methodology and guidance booklet (2014).
URL Retrieved 28, 08, 2014, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/324877/Sub-national_methodology_and_guidance_booklet.pdf
- [3] A. McCloughlin, Informal conversations on proposals for capital demonstrator to energy technology institute <http://www.eti.co.uk/about-us/> (2014).

- [4] U. Government, English housing survey. homes 2010, Online (July 2012).
URL Retrieved 07, 11, 2013 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6748/2173483.pdf
- [5] L. Swan, V. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques., *Renewable and Sustainable Energy Reviews* 13 (2009) 1819–35.
- [6] M. Kavgić, A. Mavrogianni, D. Mumović, A. Summerfield, Z. Stevanović, M. Djurović-Petrović, A review of bottom-up building stock models for energy consumption in the residential sector, *Building and Environment* 45 (2010) 1683–1697.
- [7] V. Cheng, K. Steemers, Modelling domestic energy consumption at district scale: A tool to support national and local energy policies, *Environmental Modelling & Software* 26 (10) (2011) 1186–1198.
URL <http://www.sciencedirect.com/science/article/pii/S1364815211001034>
- [8] S. K. Firth, K. J. Lomas, A. J. Wright, Targeting household energy-efficiency measures using sensitivity analysis, *Building Research & Information* 38 (1) (2010) 25–41. arXiv:<http://www.tandfonline.com/doi/pdf/10.1080/09613210903236706>, doi:10.1080/09613210903236706.
URL <http://www.tandfonline.com/doi/abs/10.1080/09613210903236706>
- [9] M. Hughes, J. Palmer, V. Cheng, D. Shipworth, Sensitivity and uncertainty analysis of england’s housing energy model, *Building Research & Information* 41 (2) (2013) 156–167. arXiv:<http://www.tandfonline.com/doi/pdf/10.1080/09613218.2013.769146>, doi:10.1080/09613218.2013.769146.
URL <http://www.tandfonline.com/doi/abs/10.1080/09613218.2013.769146>
- [10] CSE, Address-level data from the national heat map, Tech. rep., Department of Energy and Climate Change and created by The Centre for Sustainable Energy (CSE) (July 2013).
URL Retrieved 07, 11, 2013 from <http://www.cse.org.uk/news/view/1748>
- [11] J. Palmer, I. Cooper, United kingdom housing energy fact file, Tech. rep., Department of Energy and Climate Change (2012).
URL https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/201167/uk_housing_fact_file_2012.pdf

- [12] Z. Redgrove, National heat map modelling, Retrieved 10,12,2013. Webminar: <http://www.youtube.com/watch?v=MMvFBxPIN7c>. (October 2012).
- [13] S. Lloyd, Least squares quantization in pcm, Information Theory, IEEE Transactions on 28 (2) (1982) 129–137. doi:10.1109/TIT.1982.1056489.
- [14] C. Calderon, P. James, A. McCloughlin, J. Urquizo, An area-based GIS energy model for cities and neighbourhoods, in: 21st GIS Research UK conference (GISRUK 2013), 2013.
- [15] U. Government, Cambridge housing model and user guide, Part of: Domestic energy fact file and housing surveys (August 2010).
URL Retrieved, 8.09.14 from <https://www.gov.uk/government/statistics/cambridge-housing-model-and-user-guide>
- [16] D. Triantakou, S. Barr, Computational Science and Its Applications - ICCSA, Springer, 2009, Ch. A Spatial Structural and Statistical Approach to Building Classification of Residential Function for City-Scale Impact Assessment Studies, pp. 221–236.
- [17] WarmZone, Warmzone (2015).
URL Retrieved 03.03.15 <http://www.warmzones.co.uk/what-we-do>
- [18] C. Calderon, P. James, D. Alderson, A. McLoughlin, T. Wagener, Data availability and repeatability for urban carbon modelling: A carbon routemap for newcastle upon tyne, in: Retrofit 2012, Salford, UK., 2012.
- [19] U. Government, English housing survey household report 2008 and 09, Tech. rep., Department of Communities and Local Government (2010).
URL Retrieved 07, 11, 2013 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/6695/1750765.pdf.
- [20] A. Rafferty, Introduction to complex sample design in uk government surveys, Tech. rep., Economic and Social data service (2011).
URL <http://www.esds.ac.uk/government/docs/complexsampledesign.pdf>
- [21] W. S. U. WSU, Building area measurement, Tech. rep., Capital Planning and Development space management (December 2007).
URL Retrieved 07, 11, 2013 from <http://access.ewu.edu/Documents/Construction%20and%20Planning/Campus%20Space%20Data/BuildingAreaMeasurement.pdf>

- [22] S. W. Lorimer, Potential for a neighbourhood income-based domestic energy model for ordinary electricity use in England, in: 1st international conference Central Europe towards Sustainable Building (CESB10), Prague, 2010.
URL Retrieved from http://cesb.cz/cesb10/papers/2_energy/117.pdf.
07, 11, 2013
- [23] D. Z. Sui, Tobler's first law of geography: A big idea for a small world?, *Annals of the Association of American Geographers* 94 (2) (2004) 269–277.
URL <http://dx.doi.org/10.1111/j.1467-8306.2004.09402003.x>
- [24] A. S. Fotheringham, P. Rogerson, Gis and spatial analytical problems, *International Journal of Geographical Information Systems* 7 (1) (1993) 3–19.
- [25] A. Tversky, D. Kahneman, Judgment under uncertainty: Heuristics and biases, *Science* 185 (4157) (1974) pp. 1124–1131.
- [26] U. Government, Middle layer super output area (mlsoa) domestic gas estimates 2009: Great Britain (2009).
URL Retrieved 07, 11, 2013 from http://webarchive.nationalarchives.gov.uk/20121217150421/http://decc.gov.uk/en/content/cms/statistics/energy_stats/regional/mlsoa_llsoa/mlsoa_llsoa.aspx
- [27] U. Government, National energy efficiency data-framework (need) table creator (2013).
URL Retrieved 07, 11, 2013 from <https://www.gov.uk/government/statistical-data-sets/need-table-creator>
- [28] P. K. J. Han, W. M. P. Klein, N. K. Arora, Varieties of uncertainty in health care: A conceptual taxonomy, *Medical Decision Making* 31 (6) (2011) 828–838. arXiv:<http://mdm.sagepub.com/content/31/6/828.full.pdf+html>, doi:10.1177/0272989X10393976.
URL <http://mdm.sagepub.com/content/31/6/828.abstract>
- [29] J. e. a. Hickman, Pra procedures guide: A guide to the performance of probabilistic risk assessments for nuclear power plants: Chapters 9 till 13 and appendices a till g (January 1983).
URL <http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr2300/vol12/>
- [30] C. E. Gehlke, K. Biehl, Certain effects of grouping upon the size of the correlation coefficient in census tract material, *Journal of the American Statistical*

- Association 29 (185) (1934) 169–170. doi:10.2307/2277827.
 URL <http://www.jstor.org/stable/2277827?origin=crossref>
- [31] S. Openshaw, The Modifiable Areal Unit Problem, GeoBooks, 1984.
- [32] . Newcastle City Council, Warm-up north (Autumn 2013).
 URL Retrieved07,11,2013[http://www.newcastle.gov.uk/
 environment-and-waste/climate-change-and-energy-saving/
 warm-up-north](http://www.newcastle.gov.uk/environment-and-waste/climate-change-and-energy-saving/warm-up-north)
- [33] D. A. Lane, R. R. Maxfield, Ontological uncertainty and innovation, Journal of Evolutionary Economics 15 (2005) 3–50. doi:10.1007/s00191-004-0227-7.
- [34] J. Rao, Small area estimation, John Wiley & Sons, Ltd, Hoboken, New Jersey, 2003.
- [35] V. Gomez-Rubio, N. Best, S. Richardson, G. Li, P. Clarke, Bayesian statistics small area estimation, Technical report (2010).
 URL <http://eprints.ncrm.ac.uk/1686/>
- [36] Government, National energy efficiency data-framework (November 2013).
 URL Retrieved07,11,2013[https://www.gov.uk/government/collections/
 national-energy-efficiency-data-need-framework](https://www.gov.uk/government/collections/national-energy-efficiency-data-need-framework)
- [37] N. Grid, Gas demand forecasting methodology (February 2012).
 URL Retrieved28,08,2014,[http://www.nationalgrid.com/
 NR/rdonlyres/71CFD0F6-3607-474B-9F37-0952404976FB/52071/
 GasDemandForecastingMethodologyFeb12.pdf](http://www.nationalgrid.com/NR/rdonlyres/71CFD0F6-3607-474B-9F37-0952404976FB/52071/GasDemandForecastingMethodologyFeb12.pdf)
- [38] I. Hamilton, P. Steadman, R. Lowe, A. J. Summerfield, H. Bruhns, Report on the exploratory analysis of the homes energy efficiency database and energy demand., Tech. rep., University College London (UCL) (2011).
 URL Retrieved28,08,2014,<http://discovery.ucl.ac.uk/1388914/>
- [39] BRE, Commitment to reduce carbon emissions on the government estate by 10methodology (April 2011).
 URL Retrieved10,02,2014[http://data.gov.uk/sites/default/files/
 10percent_analysis_methodology_v5_10.pdf](http://data.gov.uk/sites/default/files/10percent_analysis_methodology_v5_10.pdf)