DEVELOPING A GEOGRAPHICALLY DETAILED HOUSING STOCK MODEL FOR THE NORTH EAST OF ENGLAND

Timothy Lee[1], Miaomiao He, Simon Taylor, Steven K Firth and Kevin Lomas

School of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, LE11 3TU
[1]Corresponding author: t.lee@lboro.ac.uk +44 (0)1509 565178

ABSTRACT. Housing stock models predict long term changes in the stock to inform national policy. They operate with a set of reference dwellings representing the national stock, which are changed in response to different scenarios. However, national level models do not consider geographical variations (urban location/rural surroundings, index of multiple deprivation score, etc.), so cannot aid in targeting improvement measures (eg: insulation, micro-generation, etc.) locally. A geographically varying model can identify which measures are most appropriate in a particular location. In this paper a method has been designed and implemented using information at LSOA level (c. 700 dwellings each) to introduce geographical variation for a model of the North East of England. It has been tested against DECC meter data and over 80% of LSOAs are predicted to within ±25% of DECC's data. The model allows localised policies and interventions to be tested, and is principally of interest to local government and energy efficiency initiatives.

Keywords: Domestic Stock Model, Geographic Variations, Energy Modelling

1. INTRODUCTION

There are many different types of models of housing for calculating energy demand, these vary from short term models for grid electricity management, and models of individual buildings, to national stock level models for long term planning over decades. National level stock models are typically constructed as physically based bottom up models intended for long term policy analysis [1].

These models operate by having a base set of dwellings that represent the real world stock of interest. By modelling changes to these dwellings (eg: the addition of insulation, a more efficient heating system, etc.) the change to energy demand and associated CO_2 emissions can be estimated.

Such models are used for high level policy analysis and do not contain geographic detail, and consequently can only be used for scenarios that cover the entire stock being modelled. Moving to smaller geographies will capture variations in the dwelling stock from one location to another. When moving to geographic areas of less than 1,000 dwellings each such area will have unique features. For example, a city centre might have lots of flats, a suburban area might include a high proportion of gas heated semi-detached dwellings, and a very rural area might have many old detached dwellings without gas heating being available. Due to these geographic differences, not only will the current energy demand in each such area vary, but so will the potential improvement measures applicable in each area, eg: city centre flats may not be able to have ground source heat pumps, whereas they are likely to be more appropriate in rural areas. Therefore, a model that is able to describe the variations in the dwelling stock from one geographic location to another will be particularly useful for being able to target improvement measures and incentive packages and associated policies at the local level. Such targeting will aid in more cost effectively rolling out energy efficiency improvements across the dwelling stock.

Therefore the aim of the work described in this paper is to identify and develop a method for developing a geographically rich domestic stock model that can test the impact of different policies in different locations.

2. LITERATURE REVIEW

In order to operate a national level stock model the model requires a sufficient sample of archetype dwellings in order to suitably represent the total real world stock of interest. Large scale models of this sort usually operate with up to several thousand dwelling types representing the real world stock of interest [2][3][4][5].

However, these models are restricted to one geographical level (typically the national level) and are therefore of limited use in considering more local impacts, policies and changes. Cheng and Steemers [6] have begun to address this by developing a prototype stock model (DECM) that aims to estimate the energy demand for each of the 326 local authorities in England (local authorities are governmental administrative areas, which on average contain around 65,000 dwellings).

In Cheng and Steemers' work, when modelling to local authority level they tested two methods. The first was using built form (detached, semi-detached, terraced or flat) and the second was using socio-economic class. For their dwelling stock they used the English House Condition Survey [7] (EHCS). The EHCS was an annual survey, which included a physical survey of around 8,000 dwellings per year. The physical survey included in the EHCS provided built form and the region for each dwelling. Using the physical survey data a SAP[8] based calculation of the energy demand was carried out for each of the EHCS dwellings. They could then produce average energy demand for each built form in each region. The 2001 census [9] provides a count of the built form distribution in each local authority. By combining the average energy demand for each built form and the built form distribution an estimate of total energy demand for each local authority could be calculated.

The socio-economic approach was similar in that the EHCS included a survey of the occupants of occupied dwellings, and this associated survey of occupants provided socio-economic information, identifying each household on a scale of 1-7 on the National Statistics Socio-Economic Classification (NSSeC). The census also provides a local authority level NSSeC distribution, allowing for a second estimate of total energy demand for each local authority.

These estimates could then be compared with published data from DECC based on metered gas and electricity usage. Their built form method gave mean errors of 15.09% for gas and 9.43% for electricity demand, and the NSSeC approach had mean errors of 14.1% and 9.8% respectively.

They selected built form and socio-economic classification as they are two factors with a good correlation with energy demand. From a building physics perspective it is easy to see that the detached dwelling may have higher energy demands as it has a larger proportion of its surface exposed to the outside for heat loss, whereas a mid-storey flat will be insulated above and below by other dwellings and perhaps on as many as three sides too, thus greatly reducing its potential for heat loss. A socio-economic indicator also provides useful information as there will be a correlation with dwelling size and appliance ownership and use.

Built form and socio-economics are not the only potential categories with a correlation with energy demand. Wyatt [10] found both floor area and tenure have a correlation with energy demand, although they are also related to built form and socio-economic class, eg: a larger house is more likely to be detached and have occupants with a higher socio-economic classification. Firth et al. [11] identified dwelling age as a useful factor, together with the size of the household so there are both dwelling and occupant related variables that could potentially be included. However, increasing the number of variables used to describe the spatial variation from one location to another greatly increases the amount of matching information required for each location.

Cheng and Steemers, whilst they used both socio-economic class and built form, did not combine the two together – ie: they only used one variable at a time when modelling the local authorities. It can be anticipated that a cross-tabulation to produce a multi-variable allocation method should provide an improved model, and may facilitate being able to drill down to smaller geographies than the local authority level. Being able to do this should provide a more robust model, as it reduces the reliance on a single explanatory variable that may not have the same explanatory strength in all areas.

If a model can be developed down to the LSOA level (Lower Level Super Output Area - a census geographical area containing approximately 700 dwellings), then much finer analysis is possible. Such a model can still be aggregated to provide the wider scale, high level (local authority, regional or national) analysis.

3. METHOD

The North East region of England has been chosen as the region for testing the development of a geographical housing stock model. The region contains approximately 1.2 million dwellings, 12 local authorities, 341 MSOAs (Medium Level Super Output Area – made up of four to six LSOAs) and 1,657 Lower Level Super Output Areas.

The model therefore needs to be able to approximate the real world dwelling stock in each of the 1,657 LSOAs in the North East. It needs to do this in sufficient detail that the model can later be used to calculate the effectiveness of changes to the dwelling stock in individual LSOAs. The model should be able to make predictions for physical changes to the dwellings including installation of insulation measures, innovative heating systems, micro-generation, as well as changes in occupancy patterns (eg: people working from home), in order that it can analyse different future scenarios at both the regional and small scale levels.

A suitable data set was needed that provided a sufficient number of dwellings that could provide a reasonable representation of the North East's housing stock. The English Housing Survey (EHS) [12] has been identified as the most suitable data set as it includes a detailed physical survey, sufficient to allow a SAP calculation of energy demand, together with useful socio-economic data concerning the occupants. The 2008-2009 edition includes physical surveys of 16,150 dwellings in England, of which a total of 935 are in the North East. These 935 dwellings therefore form the base stock for the model. It was decided only to use the 935 EHS dwellings that were in the North East as opposed to the entire 16,150 for the whole of England, as if the latter were used it could lead to inappropriate construction types in the model's stock (eg: it may lead to an over representation of cob walls, thatched roofs, etc.)

The Cambridge Housing Model (CHM) [13] is designed to take the EHS dwellings and carry out a SAP type calculation on each dwelling to predict the energy demand by fuel type for each EHS dwelling.

Therefore, the EHS and the CHM combined provide a physical description of 935 dwellings in the North East, the actual addresses are not revealed. Consequently there is a geographical assignment challenge as a set of dwellings for each of the 1,657 LSOAs needs to be developed from the 935 dwellings for which data are available. The intention is that the model should be full scale – ie: one dwelling in the model should represent one real world dwelling.

3.1 Built Form Method

Cheng and Steemers used a single variable to describe the housing stock distribution at local authority level. The average local authority in the North East contains 29 MSOAs and 138 LSOAs. Scaling the model down to this level needs a higher level of detail in the selection of dwellings, but using the simple built form method can give a first indication of the minimum level of accuracy that could be achieved at the smaller census geographies.

In order to do this the same simple approach has been used. Gas and electricity demand for each of the 935 North East dwellings in the EHS has been estimated using an adapted form of the CHM. Whilst the basic CHM calculation engine was used to calculate the individual dwelling energy demands, some changes were made: 902 of the 935 North East dwellings included extra information from the associated survey, so the calculation could be refined in these instances. The specific extra information for use in the calculation comes from the EHS occupant survey which asked about approximate times of occupation of the dwelling. These responses have been used to alter the standard SAP occupancy patterns and therefore heating profiles – eg: someone in the house all day will need the heating on longer than someone who is out all day. Similarly, the calculation of appliance usage has been changed to a Green Deal SAP (GdSAP) [14] base, taking into account the stated occupancy patterns. Having then used the adjusted version of the CHM to produce energy demands for each dwelling, average energy demands for each built form type could be calculated.

The 2011 census [15] provides details of the numbers of each built form type for each LSOA, MSOA and LA in the North East. These figures can then be combined with the average energy demands for each built form type to calculate total annual gas and electricity energy demands for each of the geographic areas.

DECC provide estimates of gas and electricity demand down to LSOA level, based on metered data, against which the results can be compared. However, there are some issues with this data, firstly there is an unknown misallocation of some small non-domestic units as dwellings and some large dwellings as non-domestic premises. This is because the DECC meter data simply uses a demand threshold to allocate meters as either domestic or non-domestic. Therefore small shops, offices etc. will have been classified as domestic, and in some LSOAs this will be a significant number of the properties. Also, there are some address allocation issues: DECC has been able to allocate all dwellings to the correct local authority, but 2.6% of gas demand is not allocated to the correct MSOA, and therefore not to the correct LSOA either. There is also an issue over matching geographic areas. The base year for the model is 2009 – the closest census year for that is 2011, so where possible census data from 2011 has been used, as opposed to 2001. However, the DECC meter data are from 2009, and have therefore used the 2001 census boundaries. Between 2001 and 2011 there was some redrawing of LSOA boundaries and the creation of an additional LSOA in the 2011 census. The ONS provide a conversion look-up table [16] that allows a comparison and a mapping of 2001 LSOAs to 2011 LSOA geography, However, 2.5% of LSOAs have been changed, so their results will not be directly comparable. Finally, in a small number of LSOAs new housing will have altered the population between the 2009 meter data and the 2011 census figures.

This simple built form model spilt the dwellings into just four categories. At the smaller geographies the expectation is that this will be too crude an approach, as it fails to capture the extent of variation between one LSOA and another. Therefore, the incorporation of extra variables into the categorization process should provide a finer resolution and should thus be able to capture more of the variation between geographic areas.

3.2 Detailed Method

One particular issue that needs to be addressed is the existence of areas without mains gas available. The census gives an indication of the different fuel types being used, which can be matched to the fuel types used by the dwellings in the EHS. Out of the 935 North East dwellings in the EHS 902 were occupied, and therefore included the extra occupant survey information. It is this subset of 902 that has been used for the refined geographical assignment. There are 100 dwellings available that do not have mains gas. Five alternative heating fuels were identified: bottled gas, oil, solid fuel, electric and communal. The occupant survey also identifies dwellings according to how urban or rural they are on a scale of 1-4, as does the census. Combining these gives 20 categories for the 100 off gas grid dwellings. The EHS occupant survey also identified the surrounding area of each occupied dwelling according to its decile on the IMD scale (Index of Multiple Deprivation), combining these with the 4 built form types potentially gives 160 categories for the 802 mains gas dwellings. However, in the more rural areas not every category was filled, so to account for this rural level 3 (villages) was reduced to IMD quintiles, and the most rural – 4 (hamlets and isolated) – was split into just two IMD levels. This gives 27 rural and IMD categories combined, which, together with the four built form types gives 108 categories for built form, ruralness, and IMD rating (BRIMD) for the 802 gas dwellings and 20 for the 100 non-gas dwellings. This therefore provides a much finer categorization of the dwellings than the simple four stage built form assignment. Therefore gas dwellings are categorised by built form, ruralness and IMD score, and non-gas dwellings are categorised by ruralness and fuel type.

Whilst the model is aiming to describe the dwelling stock at LSOA level, there is a smaller census geography – the Output Area (OA). There are typically five or six OAs per LSOA, and they contain around 150 dwellings. For each OA the census provides its ruralness, IMD score and the proportion of different built forms; similarly fuel type used is provided, which can be combined with the ruralness to assign the non-gas dwellings. By building the LSOAs from the OA level data sufficient variation in the dwellings assigned to each LSOA is achieved, as a larger number of the different categories will be used to describe the LSOA level stock.

Potentially this method could be used to describe the dwelling stock down to OA level. However, as was seen earlier, moving to smaller geographies increases the chance of error due to the smaller sample size, also the LSOA is the smallest level for which DECC provides meter data, so there would be no way to test the accuracy of the model if operated at the OA level.

Therefore each dwelling is assigned to one of 128 categories, and each LSOA is made up of a number of those categories. The dwelling stock for each LSOA is then created by assigning dwellings to the LSOA according to the number of dwellings it should have in each category.

4. RESULTS

4.1 Built form method

Figure 1 shows the results at LSOA level for the built form method. Out of the 1,657 LSOAs in the North East, 1,620 are displayed for gas and 1,636 for electricity. The 21 LSOAs with no data for electricity are due to there being no data available from DECC once converted from 2001 to 2011 census geographies, and the larger number of missing LSOAs for gas is due to some LSOAs having very little or no mains gas supply, such that no figures are reported by DECC. Even at the LSOA level, the simple built form method provides some degree of fit: 67.9% of LSOAs are within ± 20% of DECC's gas estimates and 69.9% are within ± 20% for electricity demand.

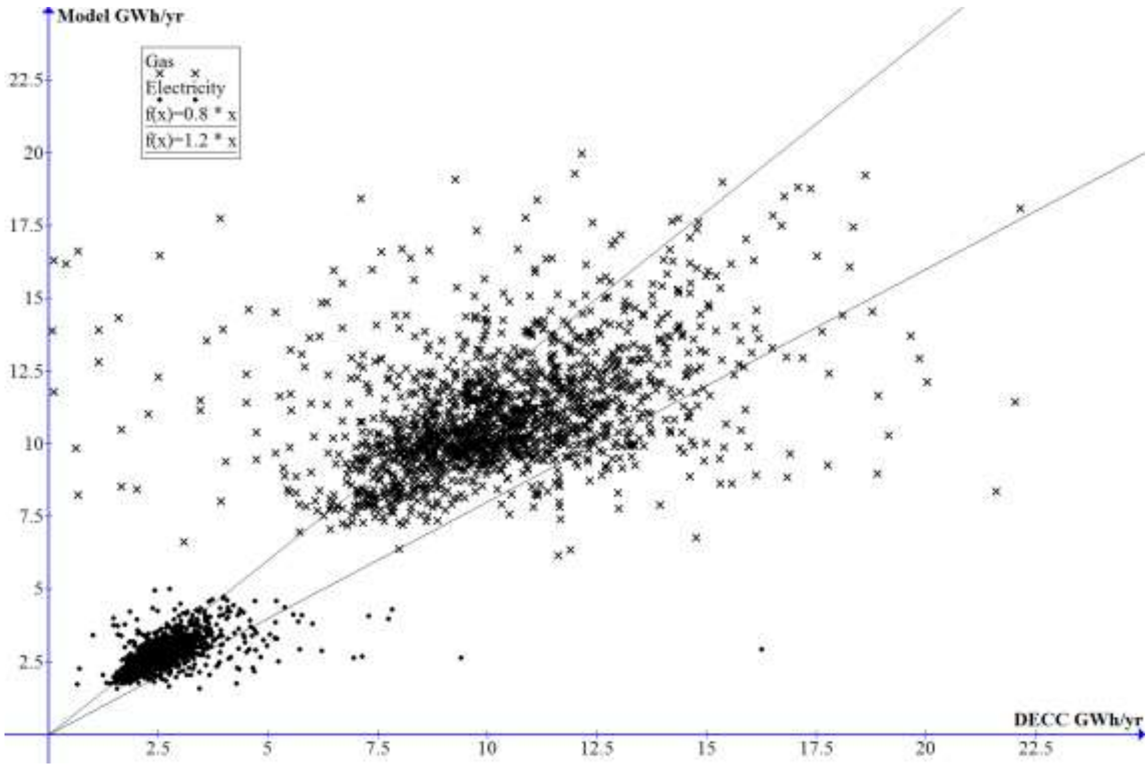

FIGURE 1. Built form based estimates of annual gas and electricity demand at LSOA level

The next figure, Figure 2, illustrates the results at MSOA level. There are 341 MSOAs in the North East, so the typical MSOA is made up of 4-6 LSOAs. As such they have a larger population, leading to a better fit, as can be seen in Figure 2: 79.8% of MSOAs are modelled within ±20% of DECC for gas and 71.6% for electricity.

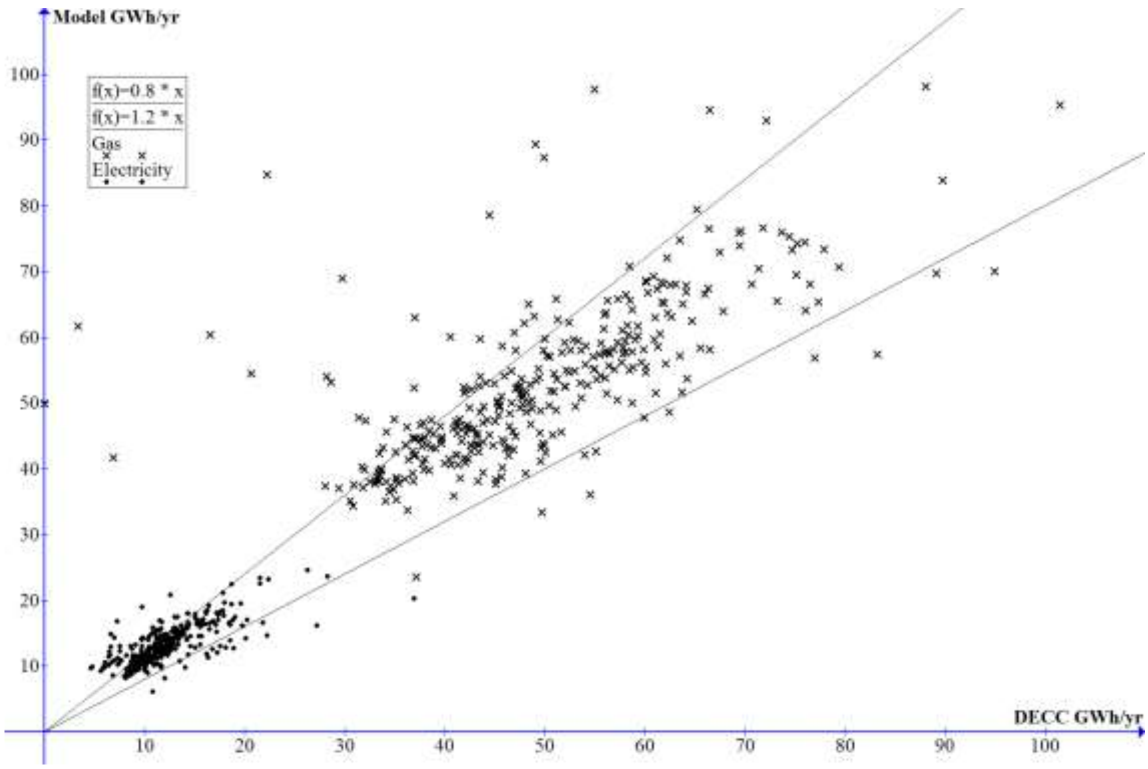

FIGURE 2. Built form based estimates of annual gas and electricity demand at MSOA level

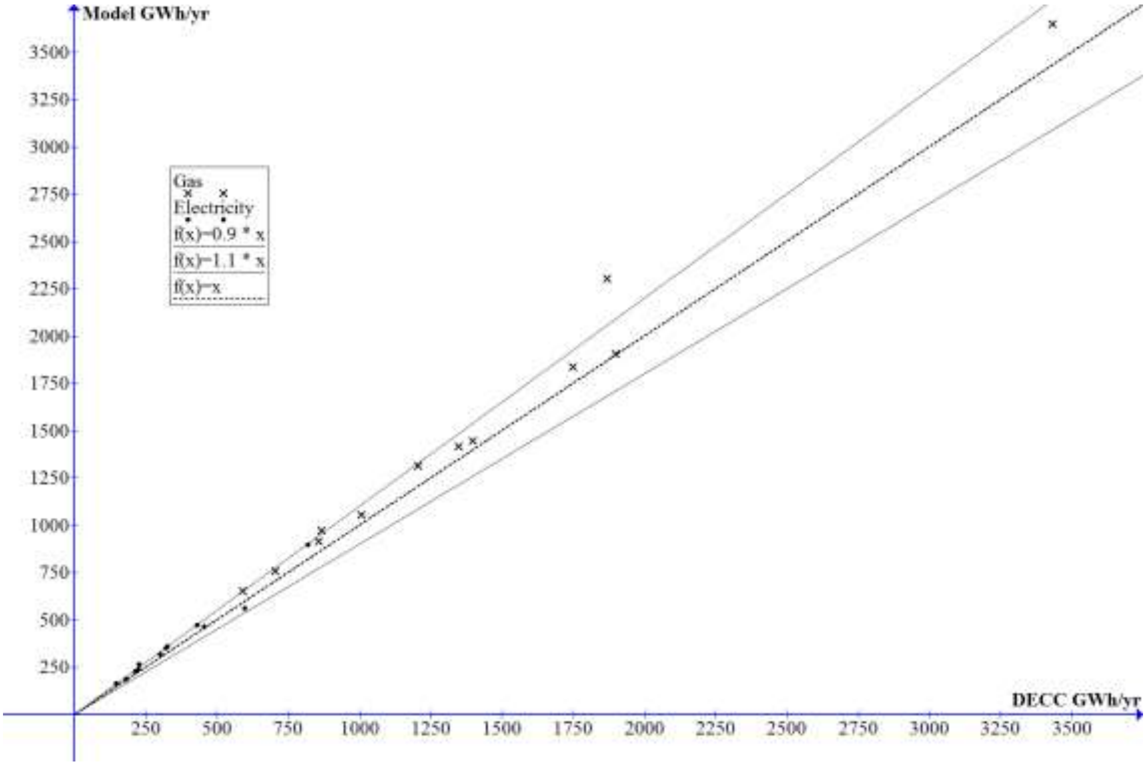

FIGURE 3. Built form based estimates of annual gas and electricity demand at local authority level

At the local authority level, with just 12 in the North East a general over prediction is observed, summing up the entire region the model is 7.7% higher for gas demand and 6.2% for electricity. The outlier for gas is Northumberland, which is over estimated by 23.4% - the most likely reason for this large error is over estimation of gas penetration. Northumberland has the lowest population density in the region with a lot of countryside, so that a larger proportion of dwellings do not have access to mains gas, and the simple built form method does not capture this level of detail.

4.2 Detailed Method

Similar graphs to those for the built form model have been produced as follows. Figure 4 shows that at LSOA level there is a noticeably closer fit: 77.6% of LSOAs were within ±20% for gas vs. 67.9% for the built form method, and for electricity 82.8% were within ±20% as opposed to only 69.9% for the built form method.

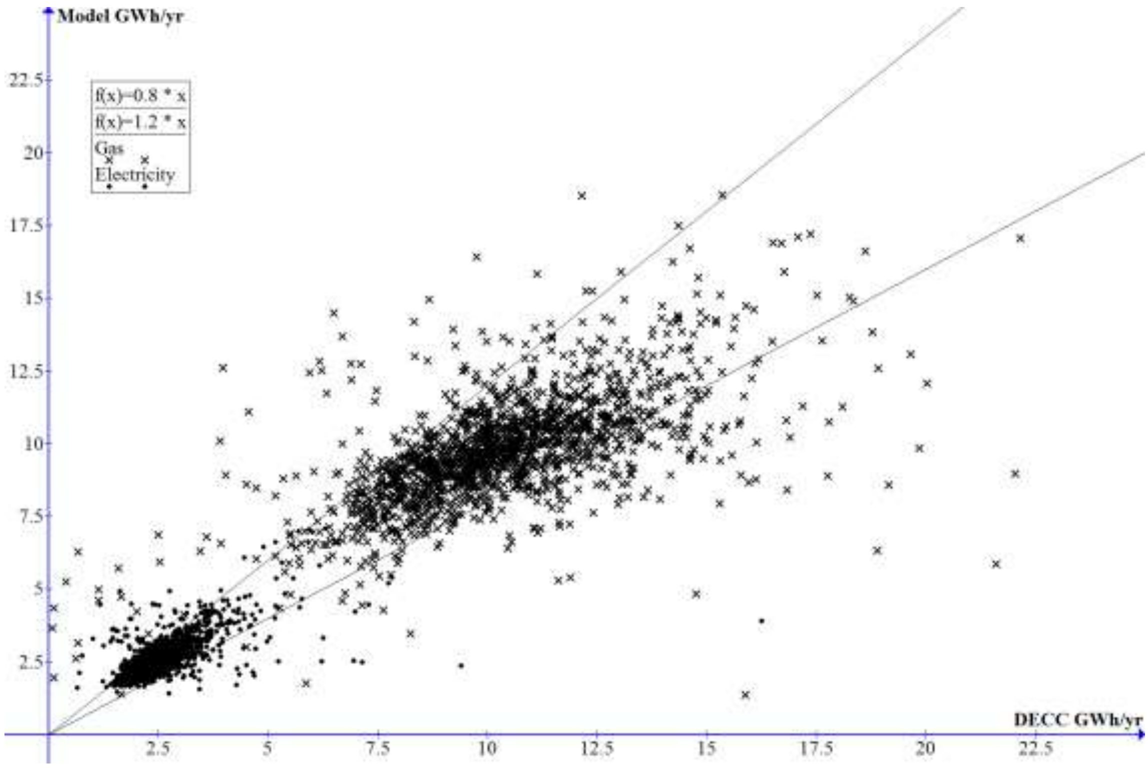

FIGURE 4. BRIMD based estimates of annual gas and electricity demand at LSOA level

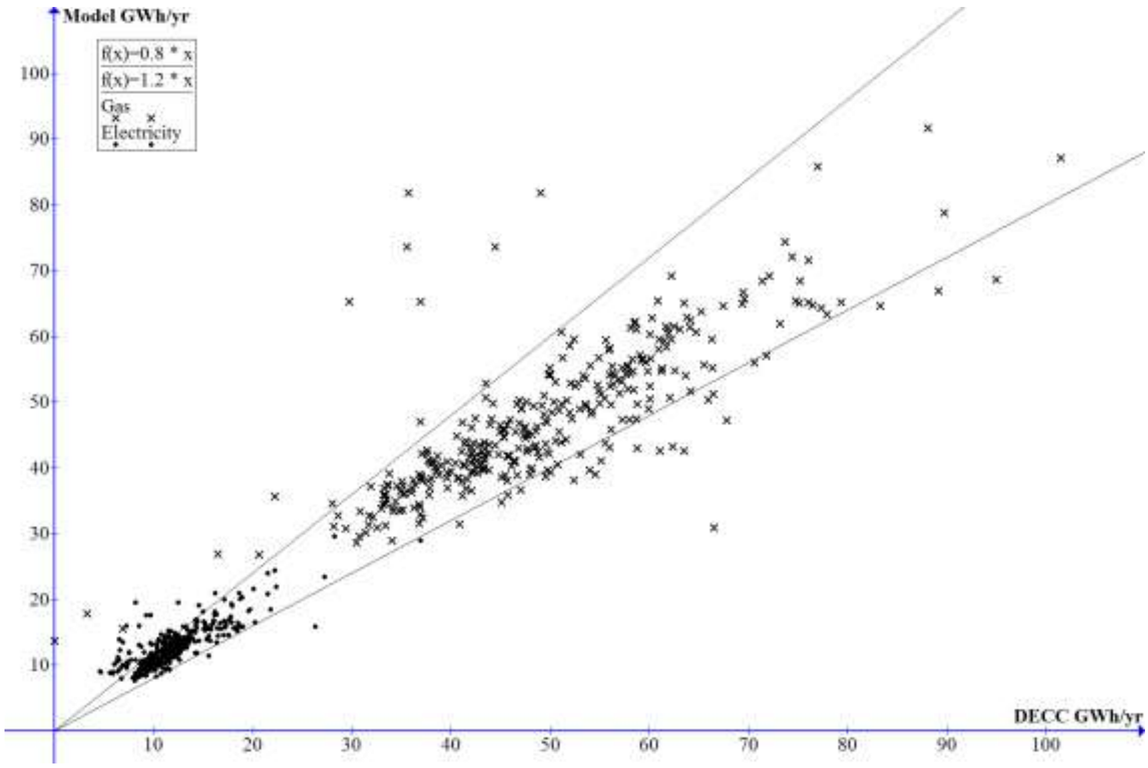

FIGURE 5. BRIMD based estimates of annual gas and electricity demand at MSOA level

At MSOA level 87.4% are within ±20% for gas and 82.5% for electricity, compared to 79.8% and 71.6% respectively for the built form method.

Figure 6 shows the results at local authority level, out of the 24 data points only one is out by more than 10% as opposed to six under the built form approach. Summing across the region the model underestimates gas demand by 5.5% and overestimates electricity by 1.8%.

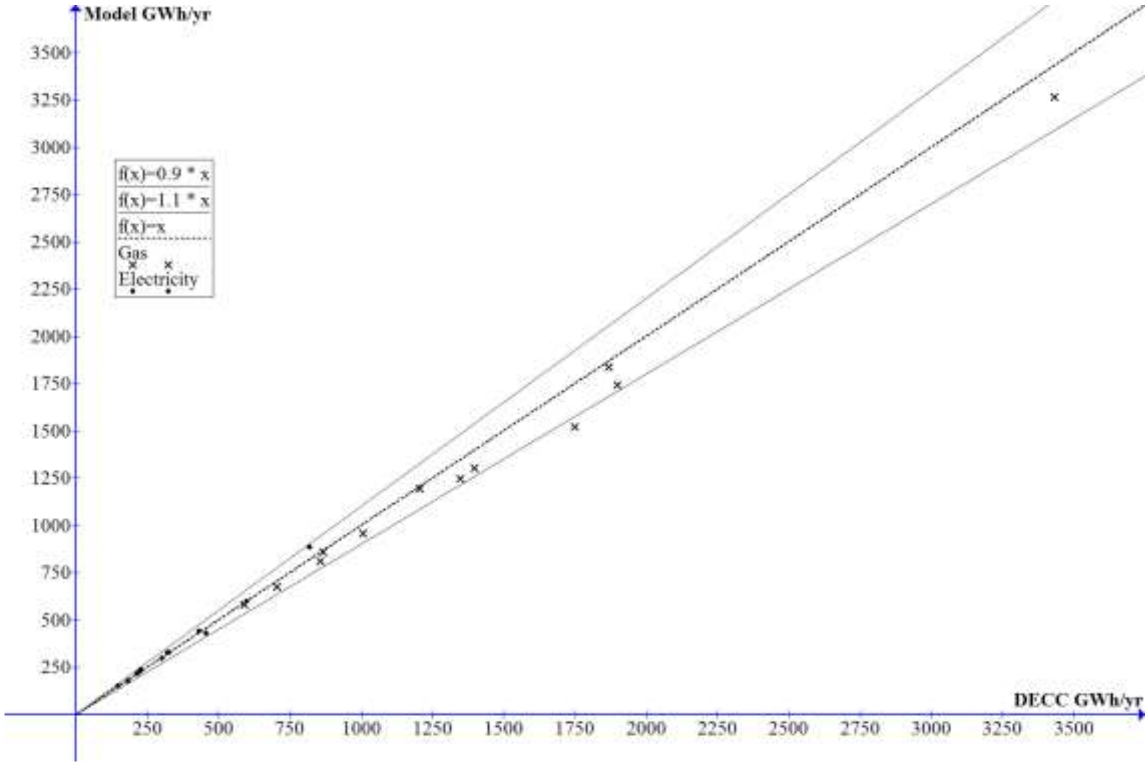

FIGURE 6. BRIMD based estimates of annual gas and electricity demand at local authority level

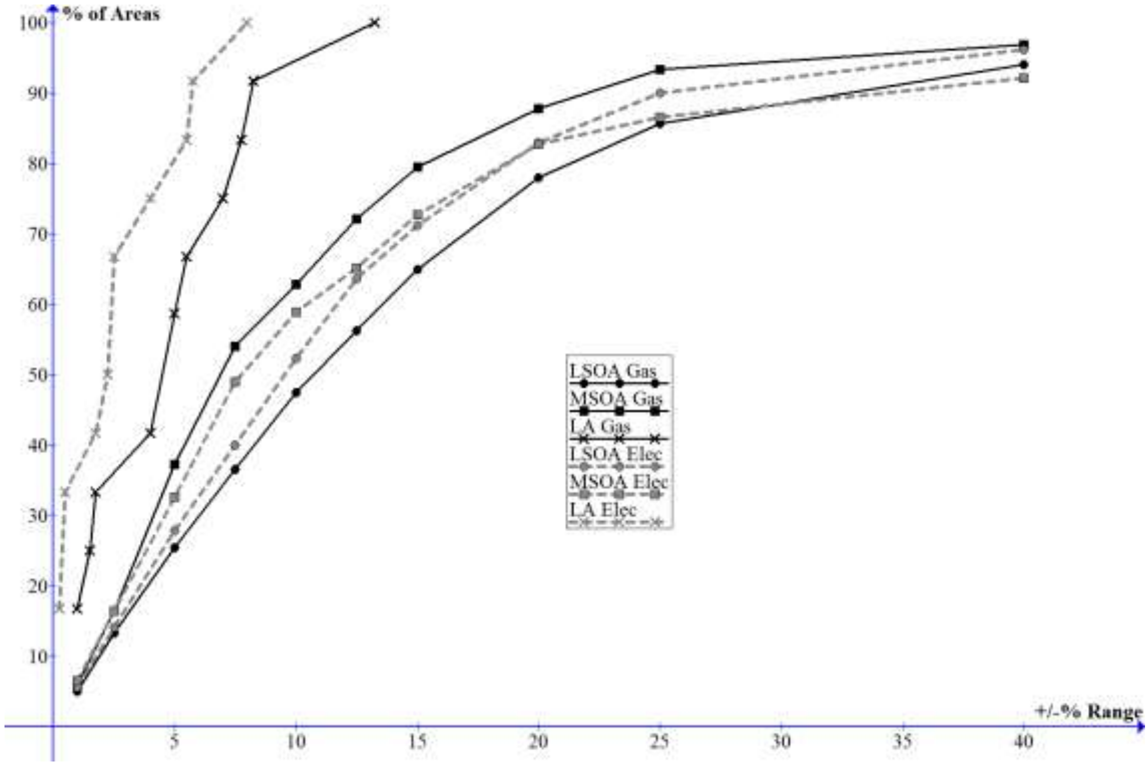

FIGURE 7. BRIMD accuracy level at different geographies

Figure 7 shows the proportion of each geographic level that estimates energy demand within a specified percentage of DECC's estimates, eg: over 90% of MSOAs are within ±25% of DECC's metered data for gas usage. It can be seen that, generally, as the population sample increases with larger geographies the accuracy increases, which is as would be expected. Figure 7 also shows that the model provides broadly similar levels of accuracy for both electricity and gas.

Figures 8 and 9 provide maps of the North East's LSOAs showing the percentage error of the model against DECC's meter data for both gas and electricity.

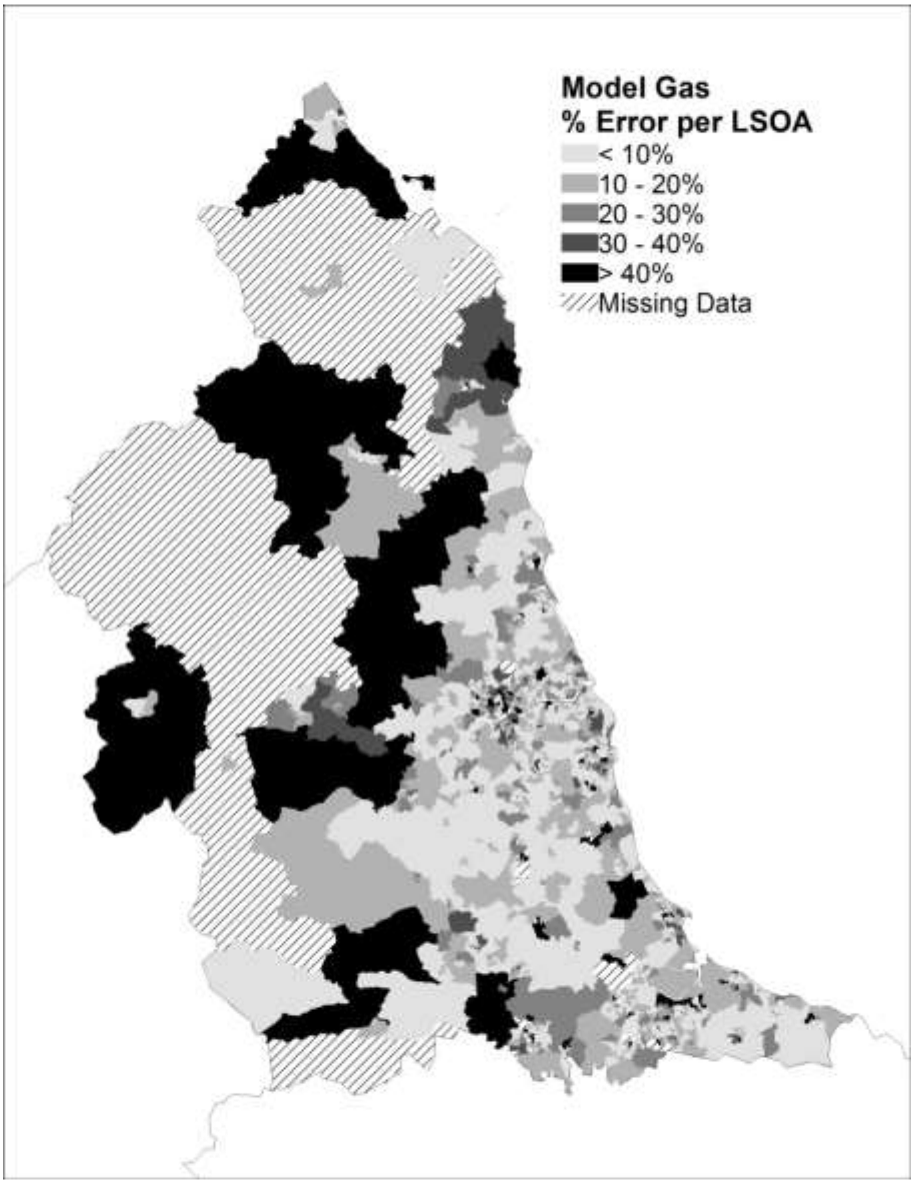

FIGURE 8. North East LSOA map of model gas accuracy

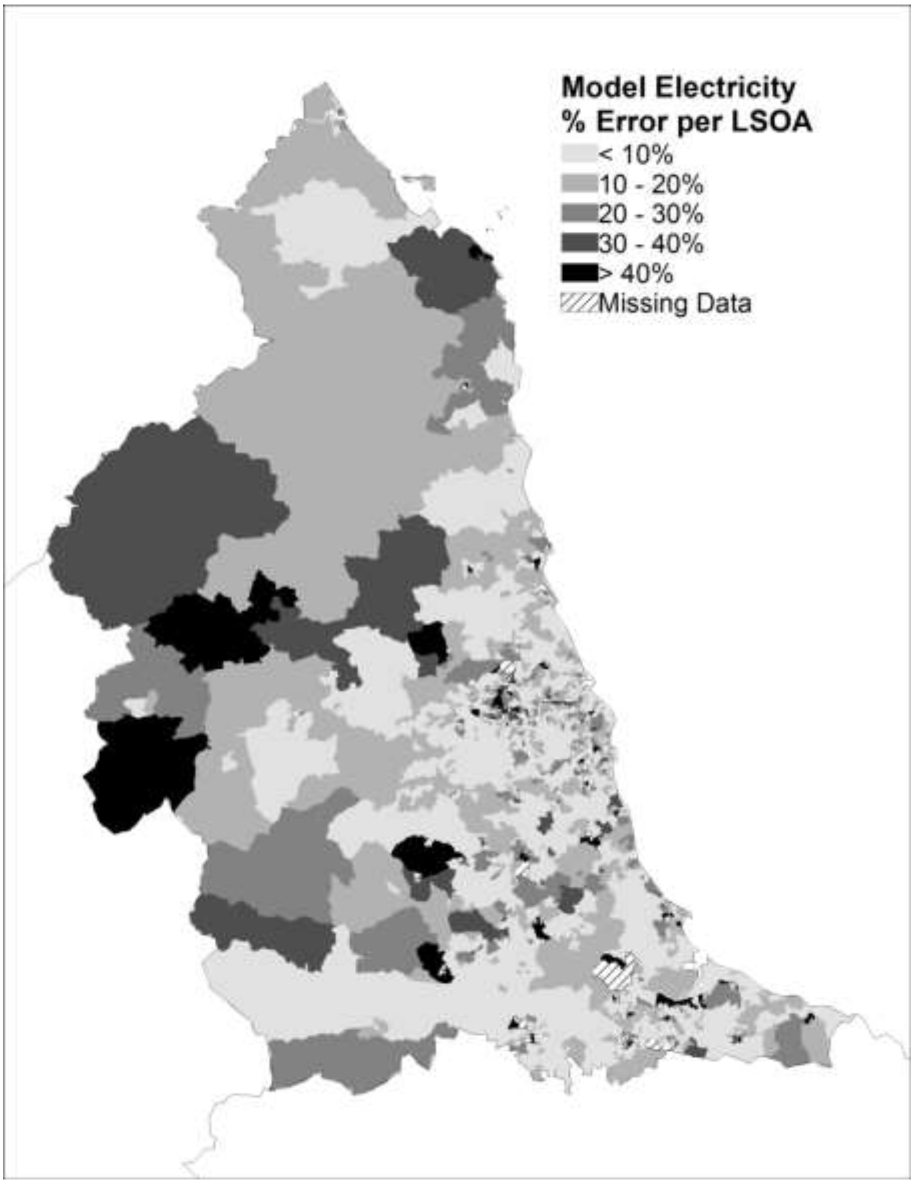

FIGURE 9. North East LSOA map of model electricity accuracy

Figure 8 shows that generally the largest discrepancies, and the LSOAs without available DECC data, are the large rural LSOAs, mainly to the north and west of the region. In Figure 9 the missing data and the largest discrepancies for electricity are mostly in more urban areas. As such some of these are going to have been caused by misattribution of non-domestic (eg: small retail units) as domestic, and also some changes in population as development has taken place, together with some boundaries being redrawn.

5. CONCLUSIONS

This work has demonstrated that it is possible to construct a housing stock model that is able to operate at different geographic levels. The base model has been constructed and tested against DECC meter data for both gas and electricity annual energy demand and found to provide a good level of agreement in most geographic areas. As this model is made up of unique LSOAs each containing around 700 dwellings, detailed analysis can take place in each of these small geographic areas.

As detailed physical information is available for each dwelling, a whole range of potential interventions can be considered at this local scale (eg: insulation, replacement heating systems, micro- heat and electricity generation, etc.) In conjunction with cost information for each technology type it will be possible to carry out localised optimisation modelling, choosing the appropriate trade-off between cost and energy and carbon savings. Since the model captures some of the spatial variation between locations, such optimisation will be different for each LSOA (eg: ground source heat pumps may be more appropriate in rural areas with lots of detached dwellings, as opposed to city centres with a high proportion of apartments).

There is the potential to refine this model further. In the areas with mains gas (approximately 83% of dwellings are connected to the mains gas supply) there are 802 dwellings available in the EHS for modelling and sufficient numbers to categorise them according to IMD score, ruralness of the area, as well as the built form of the dwellings. However, there were only 100 off-gas grid dwellings and therefore a smaller number of categories was required, based only on ruralness and main heating fuel type. Therefore, if the sample of available dwellings can be increased there is the potential to refine the categorisation process for the off-gas dwellings, in order that such areas can be more accurately represented.

Whilst it is possible to test the accuracy for gas and electricity demand, there is no reliable way to test the accuracy of the estimated demand for other fuels at this geographic level. Nevertheless, the model should provide a reasonable representation of such demand as it aims to assign dwellings in the correct proportion according to fuel type for the off-gas grid dwellings.

This same method can also be applied to the other regions of England to produce a geographically detailed stock model of the whole of England. Similarly, the same basic approach could be deployed in other countries, subject to the availability of suitable data.

Although there is the potential to refine the model still further, the base model is working and producing reasonable results. Therefore scenarios can begin to be developed that can be tested in the model to calculated their effect both regionally and locally.

ACKNOWLEDGEMENTS.

This work is supported by the Engineering and Physical Sciences Research Council (EPSRC) under their Sustainable Urban Environment Programme (grant EP/I002154/1) to SECURE: Self Conserving Urban Environments, https://www.secure-project.org/. SECURE is a consortium of four UK universities: Newcastle University, the University of Sheffield, the University of Exeter and Loughborough University.

REFERENCES

[1] T. Lee and R. Yao, “Incorporating technology buying behaviour into UK-based long term domestic stock energy models to provide improved policy analysis,” *Energy Policy*, vol. 52, no. 1, pp. 363–372, Oct. 2013.

[2] L. Shorrock and J. Dunster, “The physically-based model BREHOMES and its use in deriving scenarios for the energy use and carbon dioxide emissions of the UK housing stock,” *Energy Policy*, vol. 25, no. 12, pp. 1027–1037, Oct. 1997.

[3] M. Hinnells, B. Boardman, R. Layberry, S. Darby, and G. Killip, *The UK Housing Stock 2005 to 2050: Assumptions used in Scenarios and Sensitivity Analysis in UKDCM2 Report*, no. March. Oxford: Environmental Change Insitute, 2007.

[4] L. G. Swan, V. I. Ugursal, and I. Beausoleil-Morrison, “Occupant related household energy consumption in Canada: Estimation using a bottom-up neural-network technique,” *Energy and Buildings*, vol. 43, no. 2–3, pp. 326–337, Feb. 2011.

[5] T. Lee, R. Yao, and P. Coker, “An analysis of UK policies for domestic energy reduction using an agent based tool,” *Energy Policy*, vol. 66, pp. 267–279, Mar. 2014.

[6] V. Cheng and K. Steemers, “Modelling domestic energy consumption at district scale: A tool to support national and local energy policies,” *Environmental Modelling and Software*, vol. 26, no. 10, pp. 1186–1198, Oct. 2011.

[7] CLG, “English House Condition Survey,” 2009.

[8] BRE, “SAP 2005 The Government ’ s Standard Assessment Procedure for Energy Rating of Dwellings,” *Energy*, 2008.

[9] Office for National Statistics, “2001 Census.” 2001.

[10] P. Wyatt, “A dwelling-level investigation into the physical and socio-economic drivers of domestic energy consumption in England,” *Energy Policy*, vol. 60, pp. 540–549, Sep. 2013.

[11] S. K. Firth, K. J. Lomas, and A. J. Wright, “Targeting household energy-efficiency measures using sensitivity analysis,” *Building Research and Information*, vol. 38, no. 1, pp. 25–41, Feb. 2010.

[12] CLG, “English Housing Survey,” 2011.

[13] M. Hughes, “A Guide to The Cambridge Housing Model,” no. November, 2011.

[14] BRE, “GDsap Occupancy Assessment,” 2013. [Online]. Available: https://www.gdsap.org.uk/. [Accessed: 09-Mar-2013].

[15] Office for National Statistics, “2011 Census.” 2011.

[16] Office for National Statistics, “Local Government Reorganisation 2009 District/Unitary Names and Codes Conversion Listing,” 2013.