AN EPIDEMIOLOGICAL APPROACH TO SIMULATION-BASED ANALYSIS OF LARGE BUILDING STOCKS

Brian Coffey, Andrew Stone, Paul Ruyssevelt, Philip Haves
Centre for Energy Epidemiology, University College London Energy Institute, UK

ABSTRACT

This paper describes a novel approach to building stock energy modelling: individual building simulation models are auto-generated for each building in the stock, and the resulting set of virtual buildings is selectively sampled, simulated and analysed in much the same way that an epidemiologist might study a population through surveys and statistical analysis. A conceptual and software framework is described, along with initial case study results for a London borough.

INTRODUCTION

The goal of the project described here, SimStock, is to develop a simulation model of the energy use of the non-domestic building stock in England and Wales for use in assessing the impact of technology improvements, tighter regulations and market interventions in order to inform government policy-making and planning.

The SimStock project aims to develop a stock model that is both:

- Interpretive - i.e. can reproduce, to some degree, actual energy consumption at a national or regional scale and help to understand the factors leading to that consumption
- Predictive - i.e. can forecast the impact of interventions in the stock

Various governmental organizations have a potential interest in this type of model. For example, in the UK, the Department of Energy and Climate Change is tasked with delivering reductions in national carbon emissions. Being able to quantify the potential savings from interventions in the stock is of clear interest, as is being able to assess whether year-to-year changes in energy use are due to external factors or fundamental improvements in efficiency.

At a regional or city level, government bodies will be able to assess what specific interventions might be applicable in their areas (what works in larger cities may not work in smaller towns). Technology providers will be able use such a tool to understand the potential market for new technologies and the potential savings they could generate. Similarly, bodies that fund R&D will be able to use the tool to help identify the most promising emerging technologies towards which they should direct resources. If expanded to include an entire national non-domestic building stock, the tool would also assist large building owners in identifying how to improve their portfolios of buildings and would assist urban planners in large-scale master planning activities.

BACKGROUND

In recent years, there have been many detailed energy models of the UK domestic sector developed (Firth et al 2010, Cheng & Steemers 2011); however, models of the non-domestic sector have been more basic, due in a large way to the heterogeneity of the non-domestic stock (compared to the domestic stock) and the difficulty in obtaining and integrating information about the non-domestic stock.

The most notable non-domestic stock models that have been supporting the UK government are the Non-Domestic building Energy and Emissions Model (N-DEEM, Pout 2000) and the CarB model (Bruhns et al 2006). In both of these models, the total energy performance of the building stock is obtained by multiplying a mean energy use per unit floor area (kWh/m^2), derived from surveys and other data sources, by the total floor area of buildings. As these models do not model the physics of building performance, they cannot assess the impact of interventions on the stock. SimStock addresses this by using a simulation-based approach.

A key characteristic of SimStock is the use of a auto-generated whole building energy simulation models to represent each building in the stock model. This approach allows explicit modelling of different technologies and operating practices to predict their effect on the energy performance of different building types in different climates. Successful application of this detailed simulation approach depends heavily on data availability for informing the auto-generation of the individual building models. These data either directly inform the model (as with building footprints from online maps), or provide a basis from which aspects of the building model description might be inferred (as with use categories from taxation offices and date of last refurbishment). Such data are increasingly becoming available, in the UK and elsewhere, along with a concomitant decrease in computing costs, which also aids in the feasibility of this approach.

Non-domestic buildings often comprise several premises occupied by different users with different operational characteristics. Premises often span adjoining buildings and sometimes span separate buildings on the same site. The association of energy meters with premises means that addressing the stock by reference to buildings can be both inadequate and misleading. For this reason, we use the concept of the Self-Contained Unit (SCU) which draws together buildings into a single geometric entity so that no single premise has to be divided. A more detailed description of SCUs is provided in Taylor et al (2014).

METHODOLOGY

Epidemiological Approach

A common approach to simulation-based stock modelling involves a set of representative prototype or archetype models that are produced *a priori* (and generally without reference to any one particular building), with weighting factors based on numbers of buildings or floor areas assigned to them to represent the stock. In contrast to this, SimStock is follows an "epidemiological approach" to stock modelling, which has the following characteristics:

- Automatic construction of EnergyPlus (DOE, 2015) models for any or all SCUs in the stock. This process builds on the work undertaken in the 3DStock project (Evans et al, 2014) at University College London which aims to produce 3D models of UK building stock based on data from readily available datasets. These datasets include the Valuation Office (VOA) Summary Dataset, which holds data on floor areas and use types for the purposes of setting business taxation rates, and the Ordnance Survey database of addresses, footprints and building heights. (Analogous sources of such information exist in other countries but details of such sources are beyond the scope of this paper.)
- Scenario analysis is made possible by replicating and modifying the SCU descriptions (and thus their associated EnergyPlus models) for each year over the planning horizon
- Cloud computing is used to allow for a large number of simulations in parallel
- Epidemiological sampling and statistical analysis techniques are used to selectively sample and simulate groups of SCUs to answer particular research and policy questions

The epidemiological approach avoids a common criticism of archetype-based stock model studies by basing each model on a particular building, against which its performance can be verified. This approach also avoids some of the practical difficulties associated with the archetype-based approach in projecting future scenarios, where there is a need to bifurcate the prototype/archetype bins at each time step as some portion of them get retrofitted and the remainder are left untouched. The epidemiological approach is also inherently scalable, working equally well for small building stocks or for large. The analysis is grounded in well-known statistical theory and practice, including confidence interval quantification, to inform further analysis.

The practicality and utility of the epidemiological approach hinges upon an automated process of model construction, which has thus been a major aspect of SimStock development.

Technical Overview

Figure 1 illustrates the main components of SimStock. The `simstocktools` package / library, coded in R and Python, encapsulates the automated model generation and simulation functions for use within these statistical analysis platforms, such that researchers can operate within their commonly-used platforms for their analysis, without needing to deal with EnergyPlus syntax or simulation management.

SimStock includes the following notable components:

- an extensible JSON data structure for organizing given, inferred and simulated data for each building in the stock (along with a browser-based stock viewer)
- a Python library for automatic generation of EnergyPlus models (given well-defined input in the prescribed form) and for the management of potentially vast numbers of parallel simulations (locally or remotely)
- a R package that provides interfaces to both the external data structure and the Python simulation library

The input to SimStock, denoted as `stock_data_given.json` in Figure 1, consists of a list of data objects, one for each SCU. Each of these may be conceived of as a vector of characteristics, noting that each characteristic may be either a single value or an array (or an array of arrays, or other such structure). Similarly, the main outputs from SimStock, denoted as `simulated_stock.json` in Figure 1, consist of a list of data objects, one for each SCU that has been simulated, containing annual energy consumption data, broken down by end use and fuel type (although more detailed outputs are also available).

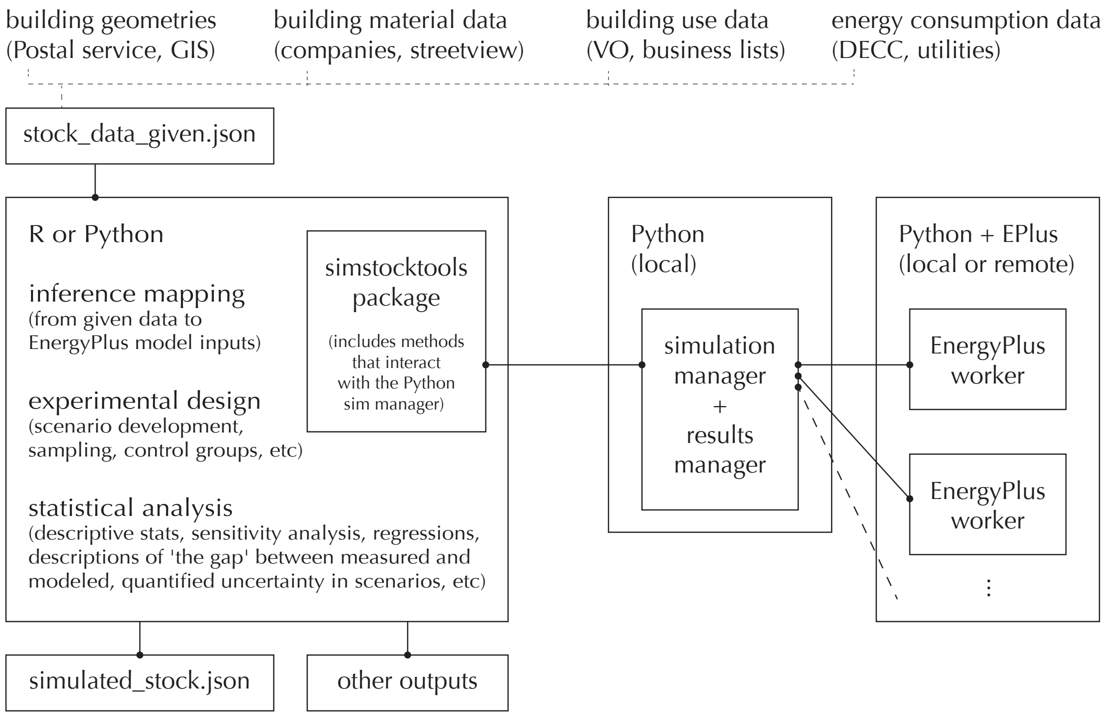

Figure 1 SimStock Technical Overview

Automated Model Generation

For each SCU in stock_data_given.json, SimStock carries out the necessary transformations to produce an EnergyPlus model for that SCU, simulates it, and reads, aggregates and presents its outputs. This process, for a single SCU, is illustrated in Figure 2. Note that the process of creating the EnergyPlus model (denoted as idf in Figure 2) consists of two major steps, one that involves a mapping from the input vector to a more detailed vector containing the information needed by EnergyPlus, which the second step then uses to produce the EnergyPlus input file with the necessary syntax. In SimStock, the vectors are all persisted in JSON format and the mapping functions are coded in Python.

This splitting of the EnergyPlus input file creation into two well-defined steps is a key aspect of SimStock, and facilitates use of its code for other projects. It separates the purely technical challenge of automatically constructing the .idf file from the statistical and/or expert-knowledge challenge of mapping from limited survey information to the more detailed information required by EnergyPlus.

SimStock allows for easy modification of its mapping functions, but the following subfunction definitions illustrate how the given_data to eplus_data mapping was carried out in the Camden pilot case study discussed below. Note that defining these mapping functions is an expert-user (or SimStock developer) task.

internalGains	$\leftarrow f$(activities)
given	array of stories, floor areas and activity codes
postulate	end use intensities and schedules
Notes: uses data from the Sheffield Hallam University surveys (Mortimer et al, 2001a, 2001b)	
constructions	$\leftarrow f$(materialsCode)
given	materials code from walk-by visual inspection data carried out by a third-party group
postulate	external construction and interior thermal mass
Notes: the materials code provides an encoding of the sub structure type, roof type, wall type and vintage. These values are decoded and mapped to EnergyPlus wall descriptions via expert engineering judgement captured in this function.	
glazingRatio	$\leftarrow f$(materialsCode, vol:area, activities)
given	materials code, volume versus wall area ratio and dominant activity type
postulate	glazing ratio
Notes: based on empirical correlations, Gakovic (2000)	

Stock-Level Analysis

With a model of each building in hand, various types of analysis are possible, including the estimation of end use breakdowns across the stock, sensitivity analysis and scenario analysis, which are all described below in the context of a case study. SimStock is structured to allow for maximum flexibility in carrying out these or other such studies, providing just the core functionality of data formatting and automated model generation, along with some example starter code for the analyst.

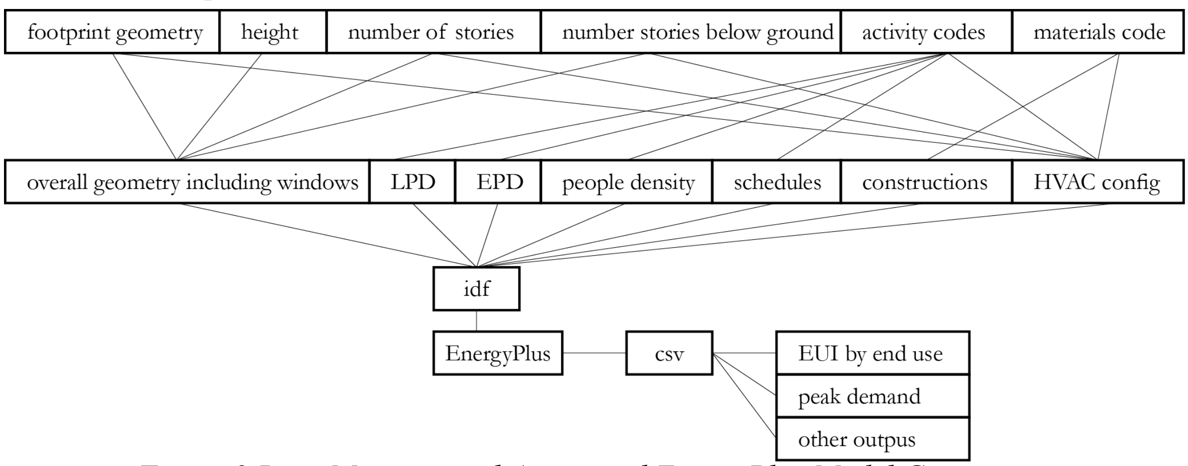

Figure 2 Data Mapping and Automated EnergyPlus Model Generation

CASE STUDY APPLICATION

As an initial case study, the SimStock framework was applied to the modelling of the non-domestic stock in the London borough of Camden. The stock model contains 143 SCUs, and the given data for each SCU includes footprint geometry, activity descriptions by floor (from the VOA), and visual materials surveys of the building envelope. SimStock was used to investigate various questions about the stock and its energy behaviour.

Energy Consumption by End Use

Perhaps the most basic application of the building stock model is to predict a more detailed breakdown of energy consumption than is readily available through measured data. Figure 3 shows the predicted end use breakdown of SCUs in the simulated Camden stock, averaged across the major activity categories for the SCU's dominant activity type. Figure 4 shows the distribution of energy intensity (annual kWh per unit of floor area) across the SCUs in the stock.

At the neighbourhood level and whole-building level, comparisons of SimStock to measured data are on-going and initial results are promising. Although the model does not accurately predict the energy consumption of each individual SCU in the stock, it does well at capturing the average behaviour and seems to do well at appropriately capturing distributions.

Sensitivity Analysis

Relationships between the many inputs and outputs of the stock model can be explored to uncover unforeseen trends and influences, which can help define the need for improved

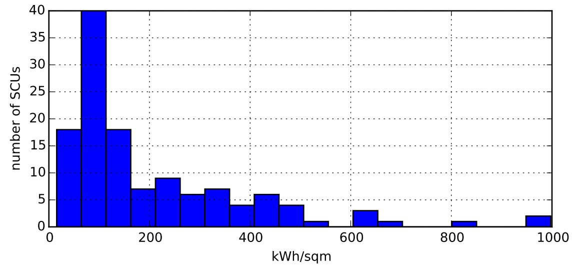

Figure 4 Distribution of Simulated Total Energy Intensity Values

information gathering and inform retrofit and/or other policy considerations.

Figure 5 shows a modified scatterplot matrix with descriptive input variables based on the given data for each SCU (e.g. number of floors, interior volume to wall area ratio, dominant activity type) graphed against the resulting simulated heating, cooling, lighting and plug load intensities. (The lines in blue are smoothed distributions and the points are relationships between the variables on the vertical and horizontal axes.) These sorts of exploratory graphs can be used to uncover relationships in much the same way as they are used for measured data, but the simulated data can provide many more points of comparison.

To investigate quantitatively the relative sensitivities of the model outputs to changes in the model inputs, stepwise (backward) regression analysis was carried out. The following variables, all taken from the given data or derived from them in simple ways, were used as possible independent variables: three continuous variables of basement volume, volume to wall area ratio and floor count; and five categorical variables of dominant activity, structure type, roof type, wall type, age group.

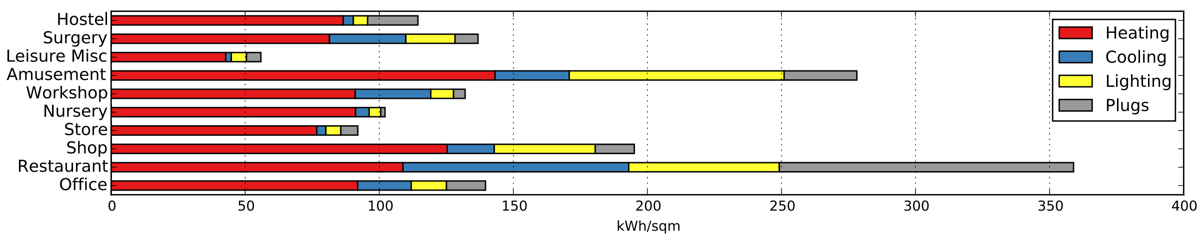

Figure 3 End Use Breakdowns by Dominant Activity Classification in Case Study Model

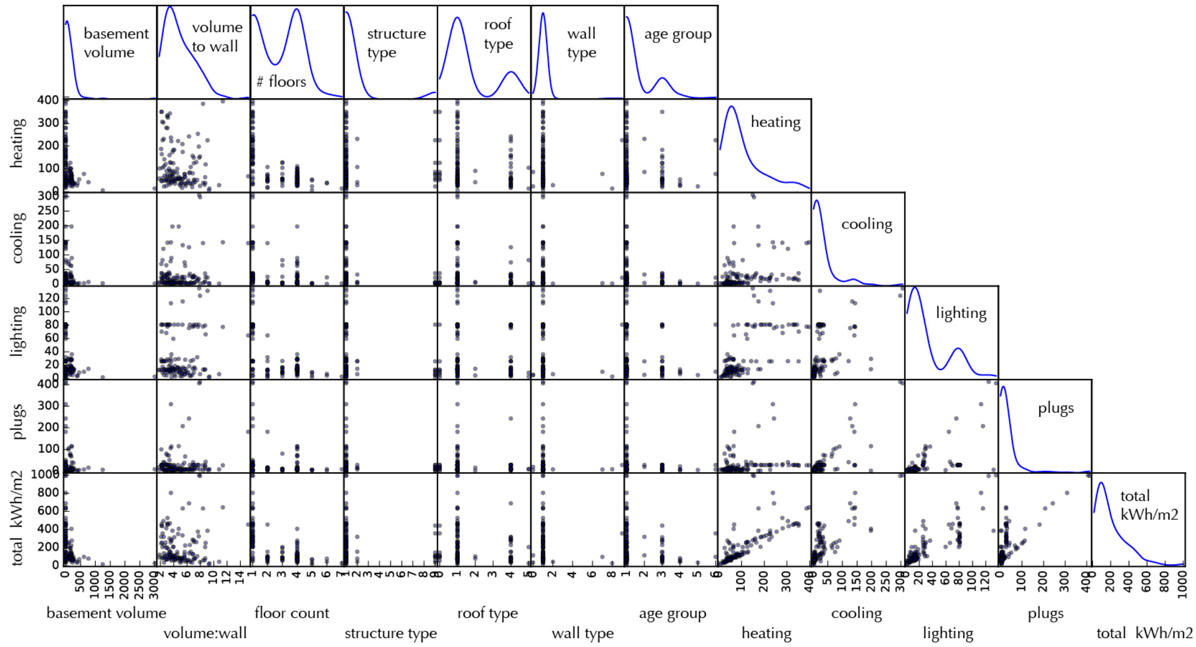

Figure 5 Scatterplot Matrix with some of the Independent (Descriptive) and Dependent (Simulation Output) Variables

With simulated total energy intensity taken as the dependent variable, the linear regression model which produces the highest adjusted R^2 value uses only the following four independent variables (with the remaining four contenders discarded): number of floors, volume to wall area ratio, dominant activity and age group. (Note that this finding is similar to recent exploratory work done independently using only measured energy data.) A linear regression model with these four independent variables has an adjusted R^2 value of 56%. With this regression model, the number of floors has a particularly strong influence on the result – it seems that in the simulated stock, when the other major variables are controlled for, the shorter buildings are using much more energy per unit of floor area than the taller ones.

The same stepwise regression method was used in turn with the simulated heating, cooling, lighting and plug loads as the dependent variable. In all four cases, the following three independent variables were always in the best-found model: number of floors, volume to wall area ratio and dominant activity. In the plug loads case, only these three are included. In the cooling and lighting cases, the roof type is added. In the heating case, the roof type is not included, but both the age group and the structure type are added to the list.

Scenario Analysis

A major use case of SimStock is for the projection of future energy consumption trajectories under different scenarios representing future changes to building regulations and other programmes that would stimulate the rate at which the existing building stock is retrofitted and refurbished. Work is ongoing to apply such scenarios to the Camden case study.

DISCUSSION

As noted earlier, comparisons of SimStock to measured data are on-going. The analysis of the differences between the model and measurements is of interest not just for validating the model. The simulation-measurement 'gap' is also a potential source of insight into the stock's behaviour – that some types of buildings (categorized by their activity types, location, or otherwise) should differ more or less broadly from physics- and data-driven expectations is perhaps indicative of other aspects of their behaviour that we do not properly understand. Investigation of such areas often leads to the uncovering of energy savings potential. Work in this direction continues.

Refining SimStock for scenario analysis will require some extra thinking, both on the part of the SimStock development team and on the part of SimStock users, around questions of the appropriate level of model fidelity for particular scenarios under consideration. For example, the mapping code in the current configuration of SimStock does not explicitly consider such things as user window-opening behaviour – for most cases this simplification is appropriate, but for some particular scenario analysis applications it should be explicitly modelled. Further research should be undertaken to better understand the trade-offs around modelling fidelity in scenario analysis.

More generally, further research is required to characterize and evaluate expert-defined mapping functions. Just as user modelling decisions in standard single-building simulation are determining factors in the model's veracity and utility, the decisions that go into defining the mapping functions are determining factors at the level of stock simulation. Building stock energy simulation is not immune to the classic engineering modelling concern of 'garbage-in-garbage-out'; more research is required to help clarify what mapping-function decisions are most important, and appropriate templates or defaults should be used in SimStock, along with guidelines for users in defining mapping functions.

One of the key early goals of the SimStock project was to solve the lower level problem of performing dynamic thermal simulations of large numbers of non-domestic buildings, and move towards addressing the much larger data acquisition and management issues: Which data sources can be incorporated into the model to improve the inference engine and hence narrow the gap between simulation and reality? How can data from disparate sources be integrated in a consistent and transparent manner? We believe that the SimStock simulation toolbox that has been developed enables these questions to be addressed by a much wider field of experts.

Another key challenge is to start to understand how the results of such simulations can most usefully be presented. The quantities of raw output data produced can be overwhelming and extracting the important patterns is a major undertaking. One approach that is being pursued is the automatic (or perhaps semi-automatic) generation of a small number of building archetypes that are in some way representative of a significant proportion of the stock. These archetypes can then be investigated in detail with some confidence that the results will be applicable to the stock as a whole. This would allow more rapid scenario analysis, and perhaps more comprehensible simulation results, without requiring the *a priori* allocation of buildings to archetypes that is inherent in typical archetypal models.

In principle, the approach outlined in this paper could be applied to the domestic building stock as well and this is likely to be a more tractable challenge. However, as discussed at the beginning of this paper, the domestic building stock has been more extensively researched in other work, at least in the UK, and therefore an extension of SimStock to cover domestic buildings is seen as a lower priority, although the inter-twining of domestic and non-domestic buildings in older urban areas may ultimately require an integrated approach.

Work to further develop SimStock and to test its ability to predict the performance of sub-sectors of the non-domestic stock is underway and will be reported in 2016.

CONCLUSION

An epidemiological approach to stock modelling has been developed and embodied in a beta software framework. The approach takes advantage of recent increases in both computational power and data availability, and allows for another mode of analysis for building stocks. An initial case study has been carried out for a London borough, including sensitivity studies and scenario analyses. Analysis of the similarities and differences between the stock simulation outputs and measured data is on-going. In future work, the stock model will be refined to address some of the issues raised here, and will be extended in applications to consider much larger stocks of buildings.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of the University College London Energy Institute's Centre for Energy Epidemiology for the research presented herein. The authors also wish to acknowledge the tremendous contributions of Philip Steadman, Rob Liddiard and Steve Evans to this work.

REFERENCES

Bruhns, H; Steadman, P; Marjanovic, L (2006) A preliminary model of non-domestic energy use for England and Wales. In: COBRA 2006 - Proceedings of the Annual Research Conference of the Royal Institution of Chartered Surveyors.

Cheng, V., & Steemers, K. (2011). Modelling domestic energy consumption at district scale: A tool to support national and local energy policies. *Environmental Modelling & Software*, *26*(10), 1186–1198. doi:10.1016/j.envsoft.2011.04.005

US DOE (2015) EnergyPlus Documentation http://apps1.eere.energy.gov/buildings/energyplus

Evans, S., Liddiard, R., Steadman, P. (2014). A 3D geometrical model of the non-domestic building stock of England and Wales. Proceedings of Building Simulation and Optimisation 2014

Firth, S. K., Lomas, K. J., & Wright, a. J. (2010). Targeting household energy-efficiency measures using sensitivity analysis. *Building Research & Information*, *38*(1), 25–41. doi:10.1080/09613210903236706

Gakovic B "Areas and types of glazing and other openings in the nondomestic building stock" Environment and Planning B: Planning and Design 2000, vol. 27, pp. 667-694

Mortimer, N. D., Ashley, A. & Rix, J. H. R. 2000a. Detailed energy surveys of nondomestic buildings. Environment and Planning B: Planning and Design, 27, 25-32.

Mortimer, N. D., Elsayed, M. A. & Grant, J. F. 2000b. Patterns of energy use in nondomestic buildings. Environment and Planning B: Planning and Design, 27, 709-720.

Pout, C (2000). N-DEEM: the national nondomestic buildings energy and emissions model, Environment and Planning B: Planning and Design 2000, volume 27, pages 721 – 732, DOI:10.1068/bst12

Taylor, S., Fan, D. & Rylatt, M. 2014. Enabling urban-scale energy modelling: A new spatial approach. *Building Research and Information,* 42**,** 4-16.