# A smart urban energy prediction system to support energy planning in the residential sector

**Moulay Larbi Chalal**

A thesis submitted in partial fulfilment of

the requirements of Nottingham Trent University for the degree of Doctor of Philosophy

February 2018

School of School of Architecture, Design and the Built Environment

## Copyright statement

# List of Publications


I. MENDOZA, M., MEDJDOUB, B., CHALAL, M.L. and SUDRA, J., 2017. Documentation of Félix Candela's Jamaica Market hypar shells in Mexico City. In: A. BÖGLE and M. GROHMANN, eds., Proceedings of IASS 2017: Interfaces - Architecture. Engineering. Science. HafenCity University Hamburg, Hamburg, Germany, 25-28 September 2017. Madrid: International Association for Shell and Spatial Structures.

II. BENACHIR, M., CHALAL, M.L., 2017. Impact of household transitions on domestic energy consumption and its applicability to urban energy planning [J]. Front. Eng, 2017, 4(2): 171-183. information Modelling, 5(4), pp.39-53.

III. Chalal, M. and Benachir, M., 2017. Powerful Predictions. [ebook] London, UK: The Energy Industry Times. Available from: http://www.academia.edu/31795101/THE_ENERGY_INDUSTRY_TIMES_-MARCH_2017 [Accessed 1 Aug. 2017].

IV. Chalal, M.L., Benachir, M., White, M., Shahtahmassebi, G., Cumberbatch, M. and Shrahily, R., 2017. The impact of the UK household life-cycle transitions on the electricity and gas usage patterns. Renewable and Sustainable Energy Reviews, 80, pp.505-518.

V. Shrahily, R.Y., Medjdoub, B., Klalib, H.A. and Chalal, M.L., 2016. Construction Site Communication Study Using the RAM Management System for BIM Adaptation. International Journal of 3-D I Benachir

VI. Chalal, M.L., Benachir, M., White, M. and Shrahily, R., 2016. Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review. Renewable and Sustainable Energy Reviews, 64, pp.761-776.

VII. Shrahily, R., Medjdoub, B., Kashyap, M. and Chalal, M., 2015. Communication framework to support more effective onsite construction monitoring. Building Information Modelling (BIM) in Design, Construction and Operations.

VIII. Chalal, M.L. and Balbo, R., 2014. Framing digital tools and techniques in built heritage 3D modelling: the problem of level of detail in a simplified environment. The International Journal of the Constructed Environment, 4(2), pp.39-52.

# List of Contents

## Chapter 4

# Chapter 8

## Chapter 9



## Appendix

# List of Figures

## Chapter 2

**Chapter 3**

**Chapter 4**



**Chapter 5**

**Chapter 6**



**Chapter 7**

**Chapter 8**

**Chapter 9**

**Appendix A**

# List of Tables

**Chapter 4**



**Chapter 5**

**Chapter 6**

**Chapter 7**



**Chapter 9**

**Appendix A**

i. Abstract

The UK residential sector accounts for approximately 27% and 17% of the country energy consumption and its CO2 emission, respectively. Thus, developing appropriate polices to reduce the environmental factors, which are associated with the CO2 emissions of a rapidly growing urban population, constitutes a high priority. Moreover, ensures the creation of cities that respect the natural environment and the well-being of future generations. While a great deal of expertise on detailing and constructing low-energy buildings and cities has been developed, it is fragmented and does not consider the concept of household life-cycle demographic transitions in the prediction of residential energy consumption.

This research aimed to develop an integrated 3D urban energy prediction tool which supports decision-making for a sustainable energy monitoring and planning in the residential sector. This, while considering the UK household demographic transition patterns in the energy prediction process.

To attain the above aim, the research embraced a mixed-methods methodology with 4 stages of practical implementations. In stages 1 and 2, statistical procedures such as binary logistic regression, were applied to the British household panel data survey (BHPS) to attain the two following objectives. First, to analyse the socio-economic and demographic factors affecting the UK household transitions; consequently, predict future transition patterns in the next 10-15 years. Secondly, to investigate the impact of the predicted transition patterns on the residential energy consumption. The examination of the findings indicated that the nature of independent factors and their degree of influence on household transition patterns were not consistent across the 10-15 years. Moreover, it advised that household transitions mostly have a positive but weak effect on their energy usage. Based on those findings, a linear regression model was developed to predict the households' future electricity usage in function of their transition, demographic and socio-economic variables.

In phases 3 and 4, a 3D urban energy prediction tool (EvoEnergy) was developed by first building a 3D semantic model of a pilot area in Nottingham city. Moreover, by integrating the research findings from stages 1 and 2 into EvoEnergy using computer scripting, open-source game technology, and 3D visualisation techniques. Finally, despite the facts that the benchmarking of EvoEnergy highlighted some areas for improvement, it has advised that EvoEnergy has the ability to predict domestic electricity consumption at the building and neighbourhood levels with a good accuracy (+/- 5% error).

## ii. Acknowledgement

Nobody walks alone on the journey of life. Indeed, the achievement of this doctoral thesis, which was a dream that has been close to my heart, was not possible without the blessing and mercy of God Almighty and the support of my family, friends, and NTU staff.

I feel very privileged to have been given the opportunity to study at the 2017's 'University of the Year'; Nottingham Trent University. Moreover, I would like to take this chance to thank Nottingham Trent University for sponsoring my PhD and for their great support.

First, my utmost gratitude to my supervisors; Professor Benachir Medjdoub, Professor Michael White, and Dr. Golnaz Shahtahmassebi for their constant support, invaluable guidance, understanding, ultimate zeal, and kindness not only during this research but also throughout my academic career. My deep appreciation goes to professor Benachir Medjdoub for providing me with cutting-edge facilities and immense support. I deeply thank him for integrating me into the research work of the CVTR lab and for giving me the opportunity to experience academic life since the very early stages of my doctoral research.

This research would not have been possible without the support of Nottingham Energy partnership. In particular, special thanks to Miranda Cumberbatch, Philip Angus, Alison McGuire, and Lucy Hull for their invaluable advice and support. Special appreciation goes to the NTU doctoral school team and my independent assessor Dr. Amin Al-Habaibeh. A very special gratitude goes to my PhD external examiners; Dr. Mohamed Gadi, who recently won an award for “excellence in engineering, science and technology”, and Professor Yacine Rezgui who is the BRE Chair in “Building Systems and Informatics”.

Finally, I would like to express my gratitude and deepest appreciation to my parents, my wife, my sisters, brother Youcef, and my friends for their inspiration, love, encouragement, and support.

# 1

# Background and Motivation

*The first chapter briefly introduces existing urban energy planning approaches and highlights the motivation behind conducting the research. Furthermore, it addresses the research aim and objectives, and the structure of the thesis.*

**Contents**

## 1.1 Background

The year 2007 was a watershed in the history of the humanity because it was for the first time in history when the majority of the globe's population lives in cities (54 % in 2014). This figure is likely to be higher in the future and is set to reach 66% in 2050 (UN, 2014). In the case of the UK, the proportion of urban population is estimated to rise from 90.1% in 2010 to roughly 92.2% in 2020 (theguardian, 2016). From this premise, the managed urban environment does not only have to cater the different needs of such a growing population but also face social, economic, and environmental risks associated with this growth. Subsequently, ensure a sustainable development of our cities.

Although tackling each of these challenges is necessary, there is a widespread concern in today's world over addressing environmental and economic issues related to the increasing energy demand in the built environment such as climate change. This is evident in the enormous pressure exerted by different international organisations on local governments translated in the form of strict CO2 emission targets. For instance, the UK is set to reduce 80% of its carbon emissions, as part of the climate change act 2008, by 2050 with effect from 1990 (GOV.UK 20 October 2014).

There is no doubt that meeting such targets necessitates a more **proactive** urban planning which takes advantage of the new possibilities offered by the technological advancement in big data and ubiquitous computing (Yeo, Yoon et al. 2013). For those reasons, many developing countries, including the UK, are allocating considerable budgets to implement and promote future/ smart cities demonstrator programmes. In this context, it is worth mentioning the Glasgow future city energy efficiency demonstrator which has particularly focused on developing energy mapping tools and building monitoring strategies to support certain physical measures more effectively. These include; building retrofit, mapping of renewables, and management of electricity demand in public buildings. It is believed that success of implementing similar strategies on the UK wider context help develop future cities that promote an excellent quality of life, while strengthening resilience to environmental change (The Farrells Review 2014).

## 1.2 The problem and its context

Governments "high priority" agenda on decarbonisation have encouraged researchers worldwide to explore effective solutions for energy sustainability related problems in the built environment while taking advantage of the incredible technological advancement. As a result, a great deal of expertise has been generated on detailing as well as constructing low-energy buildings and cities. This body of knowledge usually revolves around improving the thermal quality of buildings at micro and macro levels, with a lesser focus on the social and behavioural aspects of energy consumption.

The work conducted at the building scale is largely devoted to the utilisation of engineering simulation tools to assess the overall energy performance of individual buildings. The advantage of these tools lies in their suitability and great support to early design stages. Moreover, their flexibility and high interoperability with the majority of existing CAD/BIM systems. However, their major drawback lies in the non-consideration of socio-economic and behavioural aspects of energy consumption (Aydinalp and Ugursal, 2008). In contrast to building energy simulation, a substantial amount of work has been dedicated to utilising the potentials of artificial intelligence approaches such as artificial neural network (ANN), for energy forecasting and the optimisation of energy systems (Yuce et al., 2017). Although these approaches have a higher prediction accuracy and consider the effect of socio-economic factors on energy usage, they lack flexibility in assessing the impact of energy conservation measures (Guler et al., 2008; Aydinalp et al., 2002).

Furthermore, they are costly to develop since very dependent on the availability of historical data.

Unlike approaches at the building scale, the ones supporting physical improvement strategies (e.g. retrofit) at the urban scale have concentrated on utilising the mapping capabilities of GIS (geographic information system) to develop 2D or 3D-GIS based tools. So that urban planners could make decisions based on chromatically visualised energy figures, CO2 emission, or even microclimate of districts with various temporal. For instance, by mapping the CO2 emission of a given neighbourhood before and after implementing certain measures (e.g. loft insulation), it is possible to determine their effectiveness. GIS approaches have also demonstrated their great capabilities in aiding the evaluation process of different renewable energy sources in relation to climatic, geographic, and economic constraints (Aydin et al., 2013; Sun et al., 2017). However, despite this fact, the major drawback of GIS based approaches lies in their narrow scope which is still confined to visualisation purposes. In fact, the calculations of energy consumption/ demand occur in standalone software packages such as NHER plan assessor, whose outcomes are then mapped using GIS platforms. Another disadvantage consists of the negligence of occupancy and socio-economic characteristics of the households, whose impact is responsible for 4-30% of the variation in residential energy consumption (Jones et al., 2015; Brounen et al., 2012; Gill et al., 2010). This is mainly because these approaches consider dwellings either vacant or occupied by “typical users”. Thus, it could be argued, for instance, that affluent households would consume similar energy figures to the mid-class ones, which is unrealistic.

On the other hand, plenty of ink has been shed on the socio-economic aspect of residential energy consumption. At the urban scale, recent studies within this particular area are classified into aggregated and disaggregated. First, aggregated approaches are employed to predict the future energy trends in the residential sector in function of macroeconomic, technological and climatic factors such as energy price, gross domestic product (GDP), weather, appliances ownership and efficiency. For those reasons, they are adequate for developing energy policies at the national/ regional level as well as understanding their future economic implications (Herbst et al., 2012). However, despite their advantages, aggregated approaches are static since they do not capture the variation in the household demographic and socio-economic factors. This is because they initially consider the residential sector as an energy sink or single consumer which implies the homogeneity of households (Swan and Ugursal, 2009). In this respect, it could be argued that any household studied within this category can either be compared to any static object in a dwelling that continuously consumes energy or considered to behave identically to other consumers regardless of their different age, occupations, educational levels, type of dwellings, and preferences. Another limitation of aggregated approaches is their inability of representing technological progress. This is because their technological modelling is based on consumer responsiveness to buying a particular technology depending on the price change of other alternatives; known as elasticity substitution (Karjalainen et al., 2014). Apart from that, the non-consideration of physical factors such as insulation, type of wall, type of HVAC systems and size of dwelling, in the prediction of energy usage also constitutes another major limitation of this approach.

In contrast to aggregated approaches, the disaggregated ones have a great ability to handle the variation in the households' demographic and socio-economic characteristics such as income, household size, and marital status. This is since they rely on statistical models developed from analysing large-scale household survey data, including panel and cross-sectional surveys. This makes them suitable for developing and evaluating the effect of energy tax and welfare policies on residential energy consumption or/ and CO2 emission levels (Fawcett, 2016). If coupled with GIS, disaggregated approaches have the capability of targeting specific groups of homogenous households such as the ones on fuel poverty or low-income high consumers, from district to national level; consequently, develop appropriate measures. Again, like other socio-economic approaches, the disaggregated ones do not take into account physical parameters. Furthermore, despite their ability to capture the effect of heterogeneity across different households, disaggregated approaches are also static over time. This is because the energy predictions are not performed in a way that considers the households' demographic evolution over their life-cycle. In fact, households move between different family structures over their life-cycle starting from a single non-elderly through a couple with children to a single elderly. These transitions are indeed accompanied by many changes in the household demographic and socio-economic characteristics such as income, which might affect their energy expenditure patterns.

After analysing the above trends in the literature, the following problems and gaps have been identified;

1. Urban energy planning approaches supporting physical improvement strategies do not consider socio-economic and behavioural characteristics of the households. Similarly, urban energy planning approaches assisting the development of techno-socio-economic measures do not take account of physical factors.

2. Based on the above statement (1), it is believed that existing expertise on urban energy planning is fragmented and exists in various forms with no real integrated mechanism to assist urban energy planners in their sustainable decision-making (Rezgui et al., 2010). This statement was also reinforced and highlighted by Panão et al.(2008) who claimed: *"Urban planning has a considerable impact on the future energy efficiency of buildings and planners lack useful tools to support their decisions."*

3. None of the existing energy planning approaches consider household life-cycle transitions in the prediction of residential energy consumption patterns.

## 1.3 Aim and Objectives

The aim of this research is to develop an integrated 3D urban energy prediction tool which supports decision-making for a sustainable energy monitoring and planning in the residential sector. This research will explore the potential of innovative digital technologies such as GIS, Open-source game technologies, and visualisation techniques. In response to gap number (3) in section 1.2, the developed 3D urban energy prediction tool takes into account household demographic transitions from one family type to another in the energy prediction process. In order to achieve this overall aim, the subsequent objectives have been set;

1. To analyse and identify the research problems and gaps pertinent to residential urban energy planning. And;

2. To critically analyse approaches supporting physical improvement strategies and to develop a framework which underlines the modules, modelling principles, and prediction pipeline of the envisaged 3D urban energy prediction tool.

3. To critically analyse and statistically model the socio-economic factors affecting the UK household demographic transitions for a period between 10 and 15 years.

4. To investigate the impact of the modelled demographic household evolution patterns on the residential energy consumption. And;

5. To develop a statistical model which predicts the households' future energy usage based on their expected transition patterns, dwellings' physical characteristics, and change in socio-economic circumstances.

6. To develop a 3D urban energy prediction tool (EvoEnergy) using the relevant standards and technologies.

7. To Benchmark EvoEnergy on a residential pilot area in Nottingham in collaboration with urban planners and energy experts.

## 1.4 Justification and significance

In addition to the fragmentation of existing expertise on urban energy planning, which was discussed in detail at an earlier stage, another prominent motivation behind conducting this doctoral research is the **non-consideration of household life-cycle transitions** in residential energy planning. The importance of this concept in addition to its short and long-term benefits are discussed in the subsequent paragraph.

First, the notion of household life-cycle is not new; sociologists first employed it during the 1930's. It draws its principle from the interdisciplinary development approach combining rural and urban sociology, human development, and child psychology (Sorokin et al., 1930; Loomis, 1936; Waller and Hill, 1951). In 1950's, the household life-cycle received particular attention from marketing researchers to the point that a special conference on life-cycle and consumer behaviour was held in Michigan in 1954. This has in turn permitted to establish the household life-cycle as a prominent tool in market segmentation (Wells and Gubar, 1966; Du and Kamakura, 2006).

Today, marketers in many industries utilise this invaluable concept to predict households' decision-making and purchasing behaviour corresponding to each phase of their life-cycle based on the various changes they undergo throughout their life-cycle (e.g. demographic, socio-economic, and psychological). As a result, effective marketing strategies, which target households at different stages of their life-cycle, can be developed. For example, banks understand that single non-elderly households have financial burdens and a high level of disposable income. Moreover, they are recreation oriented, which means they are more likely to spend a significant proportion of their income on automobiles, travelling, health clubs, and adventure sports. Therefore, the marketing strategies for this particular segment should revolve around low-cost checking, credit cards, car and holiday loans (Hawkins et al., 2013).

On the other hand, different marketing strategies are applied to target couples with young children such as mortgage, credit card, revolving credit loans, and debt consolidation loans. This is because this segment is more likely to witness a reduction in income resulting from the employment situation of one of the partners (e.g. stopped working or working part-time). Moreover, from an increase in child related expenses such as food, medicine, doctor visits, child toys and games, and school admissions and fees (Campanale et al., 2015).

Despite the prominence of the concept of household life-cycle in marketing, it **has not** been previously taken into account in urban energy planning either by local authorities or energy companies. This is evident in the strategies employed to target certain households such as the ones on fuel poverty. In fact, policies on fuel poverty are based on the proportion of the household income spent on energy bills in addition to the energy efficiency of the dwelling and its appliances. However, such policies deal with fuel poverty as a static problem (Moore, 2012). Thus, they will certainly fail to accommodate and adapt to the expected future changes in the household socio-economic and demographic characteristics accompanied by their future life-cycle transition patterns. Based on that, it is believed that integrating the concept of household life-cycle in urban energy planning and forecasting will be a game changer and could bring the following short and long-term benefits.

First, short-term benefits are manifested in the development of innovative 3D urban energy planning systems supporting residential energy modelling and CO2 emission reduction strategies. One of the possibilities of such systems is the ability to predict and monitor the energy consumption/ demand of different households, under diverse demographic and socio-economic circumstances (e.g. low-income lone parents), from building to city level. On the one hand, this will permit a smarter and proactive management of the distribution network and will facilitate the decision-making with regards the nature of implemented physical measures such as injecting power from renewable energy sources to the grid and building retrofit. On the other hand, techno-socio-economic policies, which can adapt to each particular stage of the household life-cycle, can also be developed and implemented for a more holistic problem-solving approach. However, the capabilities of these systems are not confined to predicting and monitoring energy usage/demand and policy development. In fact, they could be extended to cover the evaluation of physical and techno-socio-economic measures after implementation.

In contrast to the above, long-term benefits include improving the process of demographic projection in new urban developments by considering the change in the households' energy figures across their life-cycle. This could have several implications for research within urban energy planning. For example, by considering the behavioural aspect in addition to the physical, and techno-socio-economic, it is possible to develop energy assessment and rating tools such as BREEAM, but for household energy-related behaviour. Finally, another advantage is the possibility to address areas of public engagement relatively to energy sustainability. For instance, with the help of ubiquitous computing principles, a consumer version of 3D urban energy planning systems can be created to raise the households' awareness about their daily energy usage patterns and help them reduce their CO2 emissions.

## 1.5 Original contribution to knowledge

While this research aimed to develop a 3D urban energy prediction tool, which considers various physical and socio-economic factors, its original contribution lies in the way residential energy consumption figures are predicted. More precisely, this is indeed the first and only study to predict households' energy figures based on their transitions from one family type to another. Over a period of 10 years, the developed energy prediction model can estimate the transition probabilities of a particular household type to different ones based on the households' demographic and socio-economic characteristics. Furthermore, it could predict the annual energy usage patterns accompanied with these transitions. For example, on average, a single non-elderly household has a 53.3% chance of moving to different household types after 5 years, where the possibility of becoming a couple with children is 12.1%. The chance of consuming more than 4000 KWh of electricity annually for a single non-elderly making a transition to a couple with children over five years is 35.29%.

## 1.6 Structure of the thesis

This dissertation has been carefully structured around nine chapters including the introduction chapter.

Chapter 2 discusses existing energy planning and forecasting approaches aiding physical improvement strategies in the residential sector. This will be according to a new proposed classification which considers approaches that were not included in previous categorisations such as CityGML. In addition to that, each approach in this chapter is critically reviewed based on its frequent intervention scale, workflow, advantages, limitations, and applicability in the building life-cycle. Based on this analysis, the main principles of the developed 3D urban energy prediction tool have been determined and analysed.

On the other hand, chapter 3 focuses on energy planning approaches supporting the development of techno-socio-economic measures in the residential sector. To maintain the clarity and consistency of this chapter with the previous one, these approaches have also been analysed in line with their frequent intervention scale, prediction pipeline, strengths, and weaknesses while referring to some key studies from the literature. Finally, the chapter concludes with the main research questions of this doctoral study.

Chapter 4 discusses and justifies in details the methodological choices made in this doctoral study to adequately answer the research questions. Following a hierarchical structure, the chapter has first discussed the philosophical stance of the research based on its ontological and epistemological assumptions. Afterwards, the research methodologies and methods have been addressed. Next, particular attention is drawn to the research design and practical implementations. Finally, the chapter concludes with a discussion on the validity and reliability of the adopted research methods as well as the ethical considerations.

Chapter 5 investigates the demographic and socio-economic factors that influence the transitions of single non-elderly households to other household types such as couples without children, with the purpose of predicting future transition patterns over the next ten years.

Based on that, Chapter 6 examines the effect of household transitions on their annual gas and electricity consumption figures. Furthermore, discusses the development of a statistical model that predicts households' electricity consumption based on their transition patterns in addition to demographic and socio-economic factors.

Chapters 7 addresses in its first part the demographic and socio-economic characteristics of the pilot area (Sneinton) households in comparison to Nottingham city residents. Furthermore, analyses certain physical features of its dwelling typologies. However, in its second part, it concentrates on the initial 3D modelling process of the 3D urban energy prediction tool.

On the other hand, chapter 8 discusses in depth the process of integrating the research findings from chapters 5-6 into the initial CityGML (3D-GIS) model of the pilot area (chapter 7) to enable the prediction of households' electricity consumption at the building and urban scales. This is followed by an analysis of findings from the participatory workshop held with urban planners to evaluate and help improve the developed 3D urban energy prediction model (EvoEnergy). Based on that, the chapter concludes with some improvements made to the user-interface of EvoEnergy.

Chapter 9 discusses the research results in comparison to the approaches reviewed in chapters 2 and 3. Furthermore, outlines key conclusions of this research by encapsulating the findings pertaining to each research question and analysing the study strengths and weaknesses. Based on that, a set of recommendations for future research has been presented.

# 2

# Energy planning approaches for supporting physical strategies

*The second chapter aims to critically evaluate the energy planning and forecasting approaches supporting physical improvement strategies in the residential sector. This examination will be based on the frequent intervention scale, the building life-cycle relevance, strength and weaknesses of each method while pointing out to key studies from the literature.*

**Contents**

## 2.1 Introduction

Studies supporting physical improvement strategies seek to provide tools that aid urban planning decision-making with regards improving the thermal quality of dwelling components at the micro or macro levels. This includes the following tasks;

- Evaluating and comparing the energy performance of new or existing dwellings.
- Identifying potential areas that necessitate physical improvement and choosing the most effective strategy (e.g. Retrofit, technology upgrade, reconstruction, and demolition).
- Determining the degree of change in the energy consumption of individual dwellings after certain physical measures have been applied.

The approaches used at the micro level are entirely different from the one adopted at the macro scale. This is mainly due to the fact that both rely on distinct levels of inputs and calculation/simulation techniques. Furthermore, their outputs are used with different applicability.

## 2.2 Existing and proposed classifications

Approaches supporting physical improvement measures have been subject of great interest to many scholars from various backgrounds who made invaluable effort reviewing and comparing these approaches. For example, Swan and Urgusal (2009) discussed energy prediction models used in the residential sector within two distinct categories namely; top-down (techno-socioeconomic) and bottom-up (physical). This classification was influenced by the type of envisaged measures (techno-socioeconomic or physical), the level of detail(s) and hierarchical position of data inputs in reference to the whole residential sector. On the other hand, Foucquier et al. (2013) classified different energy prediction models into three main categories namely; physical, statistical, and hybrid. Furthermore, they addressed them based on their characteristics, strengths, and limitations while providing some examples on each model. Zhao and Magoulès (2012) proposed three categories to energy prediction approaches supporting physical improvements in the building sector namely; engineering, statistical, artificial intelligence (neural networks, support vector machine), and grey models. Besides evaluating the strengths and weaknesses of each category, their level of complexity, user interaction friendliness, inputs' level of detail, computation speed, and accuracy were also addressed. Fumo (2014); however, reviewed and compared in detail previous classifications of these approaches, as shown in Table 2-1, while focusing particularly on calibrated engineering methods. However, despite the vital contribution of the above authors, there is no review that has explicitly addressed these methods based on their frequent intervention scale, nor investigated certain prominent approaches which are utilised at the urban scale such as, 2D GIS and 3D GIS (CityGML) based energy planning forecasting methods.

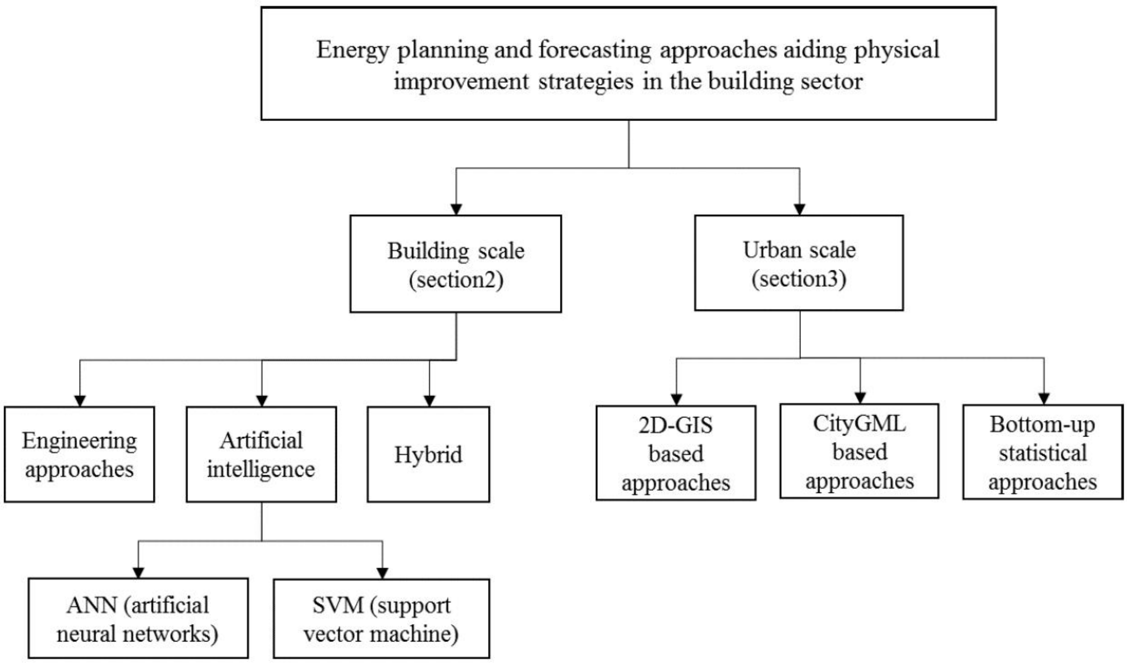



Figure 2-1.The structure of chapter 2

Table 2-1.Classification of different energy predictions approaches in previous reviews

| Authors | Classification |
|---|---|
| **(Swan and Ugursal, 2009)** | • Top Down models<br>• Bottom up models:<br>  • Statistical (regression, Conditional demand analysis, neural networks)<br>  • Engineering methods |
| **(Zhao and Magoulès, 2012)** | • Engineering methods<br>• Statistical methods<br>• Artificial intelligence<br>  • Support vector machine<br>  • Artificial neural network<br>• Grey models (hybrid models) |
| **(Foucquier et al., 2013)** | • Physical models<br>• Statistical methods (regression, artificial neural networks, support vector machine)<br>• Hybrid models |
| **(Pedersen, 2007)** | • Statistical approaches/regression analyses<br>• Energy simulation programs<br>• Intelligent computer systems ( Artificial intelligence) |
| **(Fumo, 2014)** | • Engineering methods<br>  • Calibrated<br>  • Forward(non-calibrated)<br>• Statistical approaches<br>  • Artificial neural networks<br>  • Support vector machine<br>  • Regression<br>• Hybrid approaches |

In the light of the above analysis in Existing and proposed classifications, we suggest to classify the reviewed approaches based on their intervention scale into; building and urban scale approaches (Figure 2-1). Nevertheless, there are no clear boundaries since some approaches can be applied at both levels. Indeed, the extent of which urban scale approaches can be applied varies from the parcel level to the city level, although the majority of intenventions in studies of the nature are at the district scale (Erhart et al., 2013; Strzalka et al., 2011). That is mainly attributed to the bottom-up nature of these approaches.

In all sections, each energy planning model is critically analysed based on its structure, workflow (prediction process), advantages and limitations while providing some examples on its applicability in the building life-cycle.

## 2.3 Building scale approaches

This section reviews the urban energy planning strategies supporting physical improvement at the building scale namely; engineering, artificial intelligence, and hybrid approaches. As indicated at an earlier stage, this analysis highlights their structure, prediction pipelines, strengths and limitations.

### 2.3.1 Engineering methods

Table 2-2.Classification of building energy simulation tools in some important reviews

| | **Classification** | **Criteria of categorisation** |
|---|---|---|
| **(Zhao, Magoulès 2012)** | • Simplified<br>• Elaborate | • Input requirement<br>• Level of complexity<br>• Degree of knowledge about thermal /physical equations<br>• Accuracy of prediction |
| **(Foucquier, Robert et al. 2013)** | • CFD tools<br>• Single-zone tools<br>• Multi-zones tools | • Nature of collected data<br>• Degree of knowledge about thermal /physical equations<br>• Level of complexity<br>• Accuracy of prediction |
| **(Maile, Fischer et al. 2007)** | • One dimensional<br>• Two dimensional<br>• Three dimensional | • Level of complexity<br>• Accuracy of prediction<br>• Input requirements<br>• Degree of knowledge about thermal /physical equations |

Engineering methods use physical principles and thermodynamic relationships to predict the energy performance of the whole building or some of its components. They can be further categorised into sub-groups based on the type and complexity of the used physical or thermodynamic models. However, there is no clear classification found in the literature. For example, Zhao and Magoulès (2012) classified them based on the complexity of the employed physical/thermodynamic equations into; simplified and elaborated. Others like Maile et al. (2007) categorised them depending on the type of heat transfer problem and its corresponding prediction accuracy into; one-dimensional, two-dimensional, three-dimensional models as shown in Table 2-2.

Due to the potential capabilities of information technologies, a significant number of computer simulation tools, which rely on the 3D geometry of the building, are employed to evaluate building energy performance through solving different physical\thermodynamic problems (Clarke, 2001). Owing to the prominence of building energy simulation tools in supporting engineering approaches, their structure, input and prediction pipeline will be first discussed. This is while spotting light on the heat transfer modes as well as steady-state calculation procedures utilised to forecast the building cooling and heating loads with a consideration of relevant variables (e.g. Environmental). Once analysed, the applicability of engineering tools will be discussed afterwards.

#### 2.3.1.1 Building energy performance modelling

*A-structure*

Prior to addressing the "modus operandi" of building energy simulation tools in predicting heating and cooling loads of various building types, it is important to examine their common architecture.

Overall, most of the existing simulations tools have a very similar architecture which is composed of two main components namely; Engine and graphical user interface, as illustrated in Figure 2-2. First, the simulation engine component requires an input file, usually simple text-based, to perform and write its output into one or several files. While the output file(s) comprises the simulation results, they also contain information on the simulation run. This includes warning messages or additional inputs, for instance. On the other hand, the graphical user interface convolves and facilitates the simulation engine process. This latter is represented in input generation, engine running, and output generations. Furthermore, illustrates the results graphically.

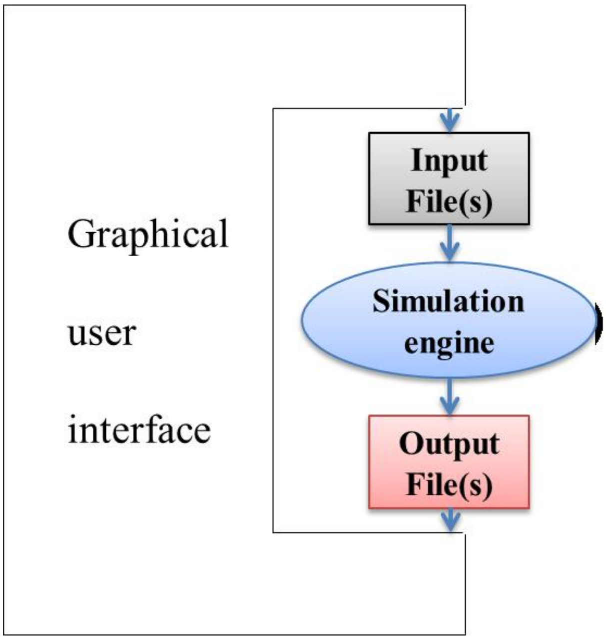



Figure 2-2.Generic Architecture of energy simulation tools adapted from (Maile, Fischer et al. 2007)

*C- Modes of heat transfer*

The first step towards calculating heating and cooling loads in a building is to quantify the amount of heat transfer in the studied building. Typically, there are three distinct modes of heat transfer namely; conduction, convection, and radiation.

*Conduction*

First, Conduction (diffusion) is the transfer of heat between substances in direct contact as a result of temperature differences where the direction of transfer is from hot to cold objects (MIT, 2017). Examples of conductive heat loss in a building includes; heat transfer through walls and windows (see Figure 2-4).

Conductive heat transfer can be defined with Fourier's law as follows;

$$q_{Cond} = U.A.\Delta T \qquad \text{Equation 2-1}$$

$q_{Cond}$ is the amount of heat transfer measured in J/S or Watts.

U, which is the coefficient of heat transfer usually measured in (W/m$^2$K), is defined as; U=k/L.

k represents the thermal conductivity of a given material usually measured in (W/mK) or (W/m C$^O$). The bigger k is the more heat will transfer and vice-versa.

L (measured in m) is the thickness of the conduction path (e.g. wall thickness). The longer the path is the less heat will transfer.

$\Delta T$ is the temperature difference between the hot and cold object surfaces. The high temperature difference is the faster heat will transfer.

A: is the cross-sectional area that is perpendicular to the heat flow (measured in m2)

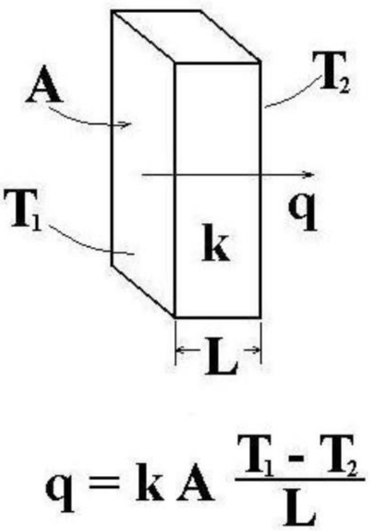



Figure 2-4.The principle of conductive heat transfer

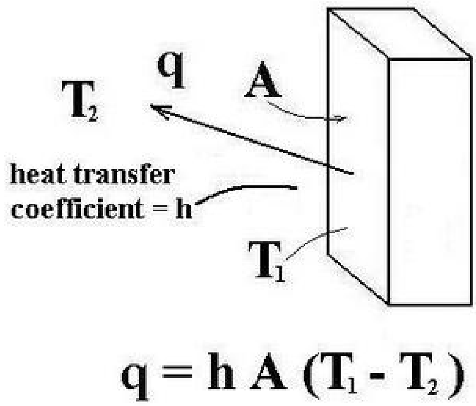



Figure 2-3. The principle of convective heat transfer

To demonstrate the impact of material properties on the amount of conductive heat transfer in a building, two common materials namely; glass and brick will be compared.

The coefficient of heat transfer (U-value) of a single glazing window pane is 5.82 (W/m2K). On the other hand, the U-value of a typical solid brick wall would be 2 (W/m2K). This means that for 1 m2 and 1C° difference between the outside and the inside temperatures, the heat transfer for the single glazing pane and the solid brick wall would be 5.82W and 2W, correspondingly. This leads to two conclusions. First, the glazing ratio in a building is crucial and does have a significant impact on the conductive heat flux. Secondly, replacing the single double pane windows by double glazing windows would reduce the heat loss through the window area by 54-64%.

For a more accurate estimation of conductive heat transfer through windows, their solar radiation is taken into account in addition to conduction (Figure 2-5). The heat gain through fenestration, which is added the conductive window heat, is calculated by incorporating a solar gain heat factor (SHGF). This will be discussed in detail in the radiation heat transfer sub-section.

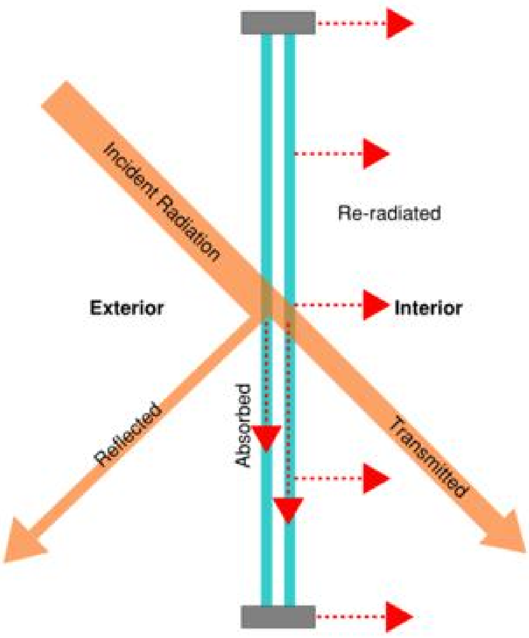



Figure 2-5. Solar gain through fenestration, adopted from (GreenSpec, 2018)

*Convection*

Convection is the transfer of heat via the motion of fluids such as the air. More specifically, a hot object surface heats the surrounding fluid which then gets carried away by the movement of this fluid (MIT, 2017). However, there exist mainly two types of convection such as natural and forced (Lu , 2011). In a dwelling, an example of natural convective heat transfer can be seen in the hot air rising from hot radiators (Figure 2-7). On the other hand, in forced convection, the fluid is forced to flow over the surface by other means such as fan, wind pressure, pump (Bahrami, 2016). Infiltration (or air leakage) through doors, windows, cracks, and latent heat (e.g. condensation from moisture in the air and heat vaporisation) is an example of forced convection in a dwelling, see Figure 2-6.

Convective heat transfer can be expressed by the Newton's law of cooling as follows;

$$q_{Conv} = h_c . A . \Delta T \qquad \text{Equation 2-2}$$

Where

$q_{Conv}$ is the flux of convective heat transfer (W or J/s)

$h_c$represents the coefficient of convective heat transfer (W/m$^2$. k). $h$ depends on the fluid type (e.g. air or hot water) and its flow properties such as velocity, viscosity, and temperature difference between the object surface and the fluid (Kreith, 2017). For example, the coefficient of convective heat transfer of air can be simplified as follows;

$$h_c = 10.45 - v + 10\ v^{1/2} \qquad \text{Equation 2-3}$$

v depicts the relative speed between the air and the object surface

$A$ is the area that is perpendicular to the convective heat flow (m$^2$)

$\Delta T$ is the difference in temperature between the convective fluid and object surface (C$^O$ or K).

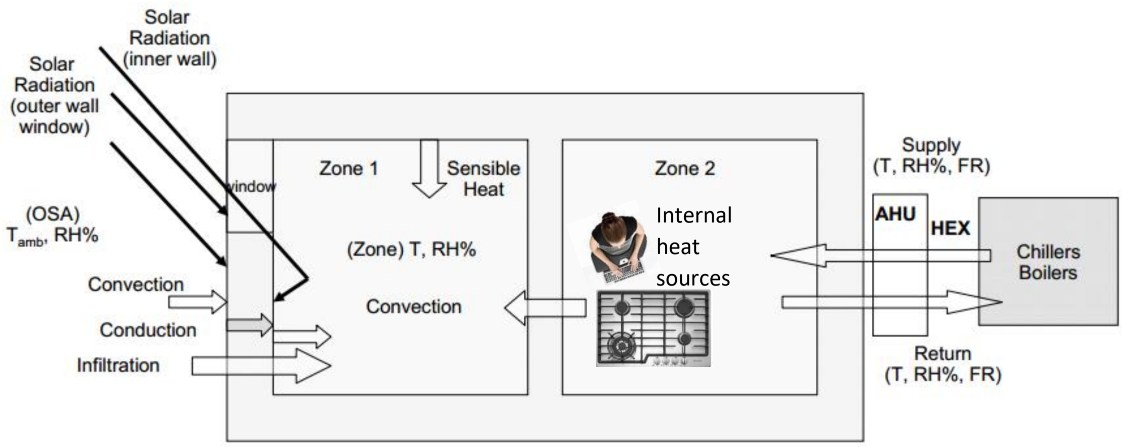



Figure 2-6. Simplified view of heat transfer in a building, adapted from

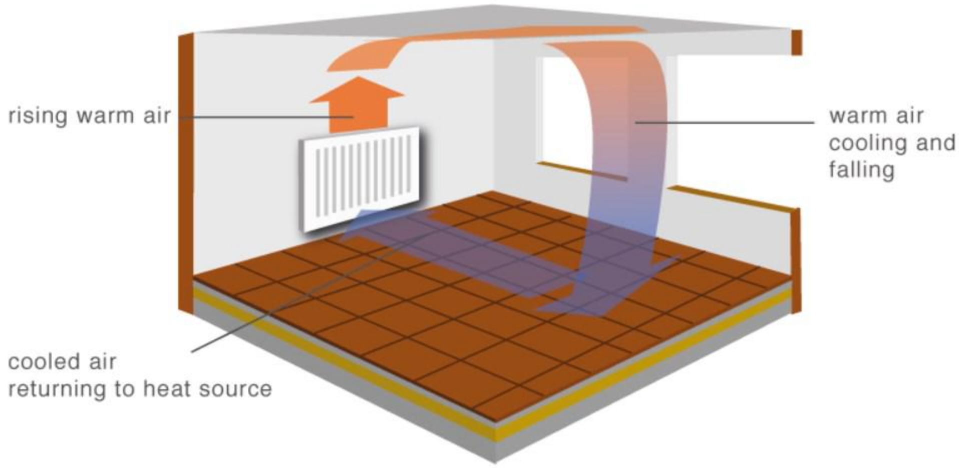



Figure 2-7. Typical example of convection in a dwelling, adopted from (GreenSpec, 2018)

### *Sensible and latent Ventilation and infiltration*

Prior to addressing the calculation procedure of sensible and latent ventilation, it is worth mentioning that sensible heat is the amount of energy required to affect the temperature of a given object. On the other hand, latent (hidden) heat is the quantity of energy that is needed to change the phase of a given object (e.g. from solid to liquid or from liquid to vapour) and without changing its temperature (Engineeringtoolbox, 2015).

First, Sensible Heat loss caused by ventilation and infiltration can be calculated as follows;

$$q_V = C_p.P.V.\Delta T$$ Equation 2-4

Where

$C_p$ represents the specific heat capacity of air (KJ/(Kg.K))

$P$ depicts the air density (kg/m$^3$)

$V$ is the air volume flow (L/s)

$\Delta T$ is the difference of tempertaure between the inside and outside

In contrast to the above, Latent Ventilation and infiltration can be expressed as follows Equation 2-5;

$$q_l = \frac{Qp(W_i - W_0)h_{fg}}{1000}$$ Equation 2-5

Where

Q is the volumetric flow of air entering the building (L/S)
$P$ depicts the air density (kg/m$^3$)
$W_i$ is the humidity ratio of the zone air g/kg
$W_0$ is the humidity ration of outdoor air g/Kg
$h_{fg}$ is the latent heat of water vaporisation at $t_i$, KJ/Kg.

*Radiation*

The third mode of heat transfer is radiation which consists of transferring heat from one body to another one but without a medium. In other words, the heat energy is transferred through electromagnetic waves instead of particles (MIT, 2017). Sun radiation is an example on this mode of heat transfer (Figure 2-8). Radiation heat can be described in relation to the "black body"(Kreith, 2017).

A block body is fictious object which completely absorbs thermal radiation waves and does not reflect light. However, most of objects in a building are not classified as a black body since they partially reflect radiation (Engineeringtoolbox, 2015). Grey body heat radiation can be described using Stefan-Boltzmann Law as follows;

$$q = \varepsilon.\sigma.T^4.A$$ Equation 2-6

Where

ε represents the emissivity coefficient of the object which ranges from 0 to 1 depending on the material and temperature of the object surface. For black body objects (fully absorb radiation), the coefficient is 1.

σ is the stefan-Boltzmann constant ($5.6703.\ 10^{-8}\ W/m^2K^4$)

$T$ represents the absolute temperature of the emitting surface in Kelvin

$A$ is the area of the of the emitting body ($m^2$)

*Heat gain from occupants*

In the light of the above, the basic net heat radiation output of unclothed human body is calculated as follows;

$$q_{Net} = \varepsilon.\sigma.A.(T^4 - T_0^4)$$

$$= (0.9)\ (5.67\text{x}10^{-8}\ W/m^2K^4)\ (1.5m^2)\ [(310K)^4 - (293K)^4]$$ Equation 2-7

$$=143\ W$$

The above equation is basic as the internal gain from occupants depends on various factors such as type of activity (e.g. Seating, working, etc…), gender, and clothing. Thus, the below equation is employed for a more accurate estimation.

$$q_{People/sensible} = N.(sensible\ heat\ gain\ per\ person\ ).CLF$$ Equation 2-8

$$q_{People/\text{Latent}} = N.(Latent\ heat\ per\ person\ )$$ Equation 2-9

Where

N represents the number of occupants

CLF is the cooling load factor tabulated by ASHRAE (1997 fundamentals, chapter 2, Table 37)

An example of latent and sensible heat gains per person based on their activity level are presented in Table 2-3.

Table 2-3. Example of sensible and latent heat gains by level of activity, retrieved from (ASHRAE, 1997)

| Activity | Total heat, Btu/h | | Sensible heat, Btu/h | Latent heat, Btu/h |
|---|---|---|---|---|
| | Adult, male | Adjusted | | |
| Seated at rest | 400 | 350 | 210 | 140 |
| Seated, very light work, writing | 480 | 420 | 230 | 190 |
| Seated, eating | 520 | 580 | 255 | 325 |
| Seated, light work, typing, | 640 | 510 | 255 | 255 |
| Standing, light work or walking | 800 | 640 | 315 | 325 |
| slowly, | 880 | 780 | 345 | 435 |
| Light bench work | 1040 | 1040 | 345 | 695 |
| Light machine work, walking 3mi/hr | 1360 | 1280 | 405 | 875 |
| Moderate dancing | 1600 | 1600 | 565 | 1035 |

*Radiation from lighting sources*

On the other hand, the amount of heat load from lights to a given room depends on the light intensity in the room, type of light (e.g. LED) and location of the light in the space (Kutz, 2015). It is calculated as shown in Equation 2-10s;

$$q_{lighting} = W.F_{ul}.F_{sa} \qquad \text{Equation 2-10}$$

$q_{lighting}$ is the heat gain from light sources (W)
W represents the total wattage of the light source measured in Watts
$F_{ul}$ is the factor of light usage which, in turn, represents the time the lights will be in operation. For example, this factor is equal to 1 in offices and hospitals.
$F_{sa}$ is the lighting special allowance factor which considers the amount of heat released from electrical ballast. For instance, the light special allowance of fluorescent lights is 1.2 (Kreith, 2017).

*Radiation through windows*

As indicated previously in the conductive heat transfer section, the solar gain of window areas is calculated by incorporating a solar heating gain factor as indicated in Equation 2-11.

$$q = A.SHGF.SC \qquad \text{Equation 2-11}$$

Where

$SHGF$ is the solar gain factor which can be obtained from the ASHRAE fundamentals 1997-chapter 29 Table 15. This factor does, in turn, depend on other factors namely;

- Latitude (refers to the angular distance of a particular site North/ South of the equator (National Geographic, 2015))
- orientation,
- time of the day and year.

$A$ is the surface of the glass ($m^2$)

$SC$ is the shading coefficient of the employed glass which can, in turn, be retrieved from the ASHRAE fundamentals handbook.

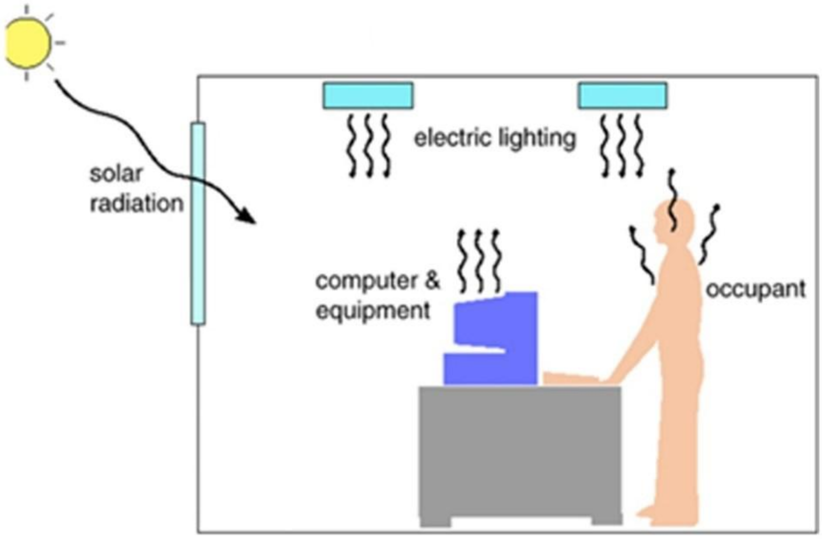



Figure 2-8.Difference radiation heat sources

*Heat gain from appliances*

Sensible and latent heat gains from electric, gas, and steam appliances can be calculated in function of their energy consumption (rated energy input) $q_{input}$ ,usage factor($F_U$), and radiation factor ($F_R$) as shown below in Equation 2-12 and Equation 2-13.

$$q_{appliances/Sensible} = q_{input} \cdot F_U \cdot F_R \quad \text{Equation 2-12}$$

$$q_{appliances/Latent} = q_{input} \cdot F_U \quad \text{Equation 2-13}$$

Examples of radiation and usage factors of some common domestic appliances such as Convection oven, are included in Table 2-4 (below).

Table 2-4.The radiation and load factors of some domestic appliances, adopted from

| Appliance | Usage Factor $F_U$ | Radiation Factor $F_R$ | Load Factor $F_L = F_U F_R$ Elec/Steam |
|---|---|---|---|
| Griddle | 0.16 | 0.45 | 0.07 |
| Fryer | 0.06 | 0.43 | 0.03 |
| Convection oven | 0.42 | 0.17 | 0.07 |
| Charbroiler | 0.83 | 0.29 | 0.24 |
| Open-top range | | | |
| without oven | 0.34 | 0.46 | 0.16 |
| Hot-top range | | | |
| without oven | 0.79 | 0.47 | 0.37 |
| with oven | 0.59 | 0.48 | 0.28 |
| Steam cooker | 0.13 | 0.30 | 0.04 |

*D-Calculation of heating and cooling loads*

This sub-section aims to discuss the thermal physics principles that are used to attain an indoor thermal comfort (e.g. temperature and humidity).

Considering that buildings are not perfectly insulated nor protected from the sun-light, achieving such comfort is challenged by climatic conditions. In simple terms, heat penetrates through the building envelop during in warm and humid climate conditions, whereas it escapes under cold and dry weather conditions (Lee et al., 2011). In order to maintain a comfortable indoor environment, HVAC (heating ventilation, and air conditioning) systems, which account for 45% of the total domestic energy usage in the UK (DECC, 2012b), are employed to compensate for the influence of outdoor climatic conditions. Nevertheless, heat gain from internal sources such as occupants, lighting and appliances, contribute significantly to the heating and cooling loads of buildings (Bhatia, 2017), Figure 2-9.

In order to calculate the heating and cooling of a building, the heat transfer and internal gains in the studied building have to be quantified. In building energy simulation tools, the calculation of cooling/ heating loads can follow three distinct approaches namely; weighting factor analysis (Transfer function method), the thermal network method, and heat balance (Kutz, 2015). However, the latter, which forms the basis of many load forecasting methods, is the most rigorous approach. Furthermore, it is more flexible (ASHRAE, 2001). Thus, it is adopted by well-known energy simulation tools such as Energy plus and ESP-r (Krarti and Ihm, 2009).

The heat balance approach is based on capturing a building envelop thermal physics using the first law of thermodynamics (energy conservation). However, the heat balance approach (like any steady-state approach) is based on the following assumptions. The first one is because the air temperature in a zone is homogeneous. Similarly, the second assumption is represented in the homogeneity of building surfaces' temperature and radiation. The third assumption is that long (LW) and shortwaves (SW) irradiations should be homogeneous. Finally, the conductive heat transfer must be one-dimensional (ASHRAE, 2001).

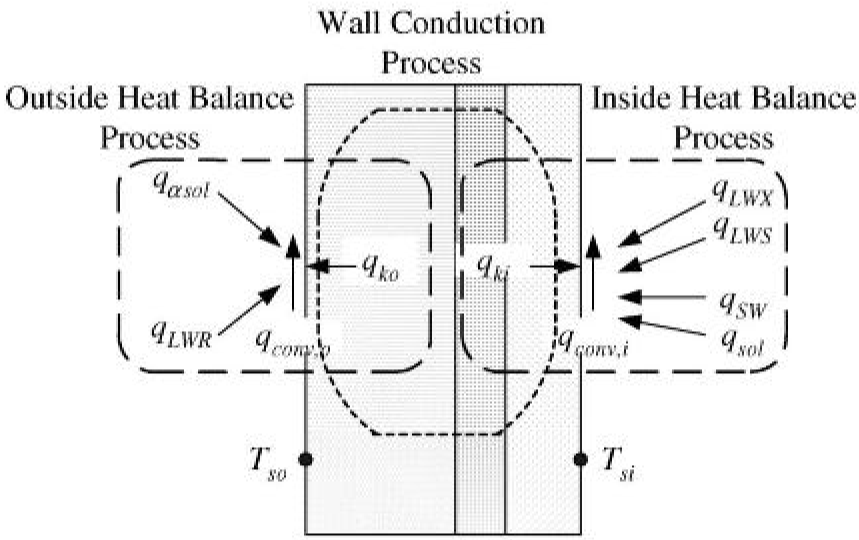



Figure 2-9. Heat balance procedures of a typical building, adopted from Krarti (2009)

Considering the above assumptions, the heat balance for every single zone in a given building is calculated following four distinct processes. These are outer building surfaces heat balance, wall conduction process, inner building surfaces heat balance, Air zone balance (Figure 2-9). Figure 2-10 (below) illustrates the association between these processes where the part enclosed in a box is repeated for each surface surrounding a given zone (e.g. ceiling, wall, etc...). Please note that a similar process is adopted for transparent surfaces except that outside face heat balance would be replaced by solar heat gains appearing in the conduction process (ASHRAE, 2001). For more information, please refer to the heat transfer modes section.

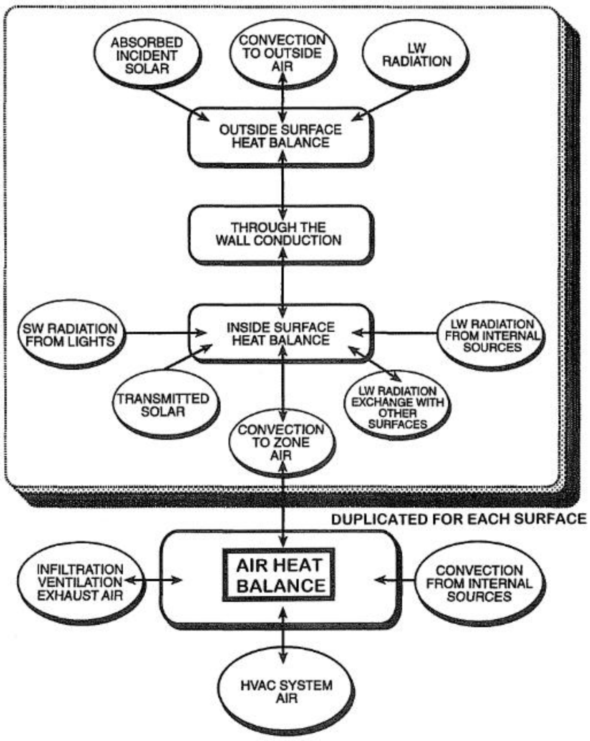



Figure 2-10. Schematic version of Heat balance procedures of a typical building, adopted from (Strand et al., 2001)

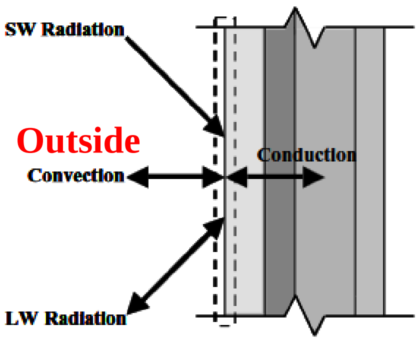



Figure 2-12. Outside surface heat balance (Strand et al., 2001)

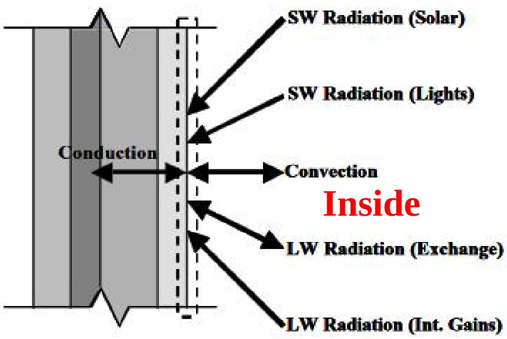



Figure 2-11. Inside surface heat balance (Strand et al., 2001)

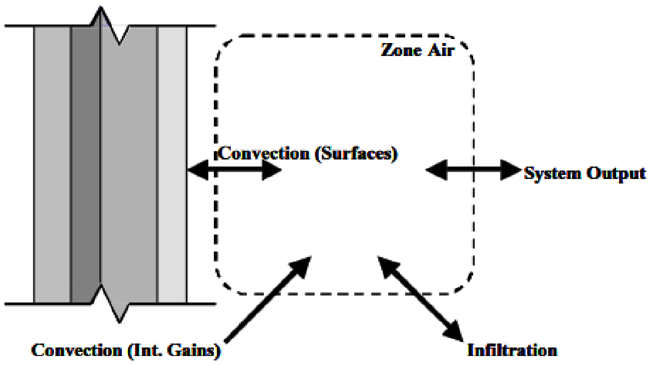



Figure 2-13. Zone air heat balance, adopted from (Strand et al., 2001)

*Outside face heat balance*

The heat balance on the outside face of a building element such as wall (Figure 2-12), is defined as follows in Equation 2-14 (Strand et al., 2001);

$$q''_{\alpha_{sol}} + q''_{LWR} + q''_{Conv,o} - q''_{ko} = 0 \qquad \text{Equation 2-14}$$

Where

$q''_{ko}$ represents the conduction flux into the wall
$q''_{\alpha_{sol}}$ depicts the absorbed and diffuse solar radiation heat flux (shortwave)
$q''_{Conv,o}$ is the convective flux exchange with outside air
$q''_{LWR}$ is the net long wavelength radiation flux exchange with the air and surroundings.

In this respect, it is worth mentioning that the source of shortwaves is usually from the sun because such wave types contain high amount of energy. On the other hand, long wave radiation are emitted from the atmosphere and objects who usually receive short-wave radiation (Molotch, 2010).

*Inside face heat balance*

The heat balance of the inside faces of a given zone (Figure 2-11), which represents the heart of the heart balance method is defined as follows in Equation 2-15 (Strand et al., 2001);

$$q''_{LWX} + q''_{SW} + q''_{LWS} + q''_{ki} + q''_{sol} + q''_{Conv} = 0 \qquad \text{Equation 2-15}$$

Where

$q''_{LWX}$ is the net longwave exchange flux between zones surfaces.
$q''_{SW}$ represents the net radiation flux to a surface originated from lights.
$q''_{LWS}$ depicts the longwave radiation flux from the zone equipment (e.g. appliances)
$q''_{ki}$ is the conduction flux through the wall
$q''_{sol}$ is the transmitted solar radiation (e.g. from a window) and which are absorbed by the surface.
$q''_{Conv}$ is the convective heat flux to zone air.

In this context, it worth mentioning that the calculation of the net longwave exchange flux between zones is usually approximated in the majority of building energy simulation tools. More precisely, the air is considered fully transparent to longwave radiation which means that it does not participate in the exchange of longwave radiation of a zone surfaces (Xu and Spitler, 2014).

One advantage of the heat balance approach is that it considers the radiant and convective heat exchange of furniture in a zone. Moreover, it considers furniture thermal mass added to the space (Kreith, 2017).

*Zone air heat balance*

In the air zone heat balance model, the thermal mass (capacitance) of the air is usually neglected (Kreith, 2017). As a result of this simplification, the air heat balance follows a quasi-steady balance (Equation 2-16)

The steady state assumption is utilised when one part of a system reacts much more quickly than the other one. This leads to the simplification of the time dependence of the faster reacting system on the slower part which diminishes the number of variables that need to be solved (Medina, 1999).

$$q_{Conv} + q_{CE} + q_{IV} + q_{sys} = 0$$ Equation 2-16

Where

$q_{Conv}$ represents convection heat transfer from the surfaces which is the sum of convective heat transfer from the inside surface heat balance.

$q_{CE}$ depicts the convective part of internal loads.

$q_{IV}$ represents the sensible loads resulting from the infiltration and ventilation air

$q_{sys}$ is the heat transfer to/ from the HVAC system.

To solve the zone temperature, Equation 2-16 must be simultaneously solved with the one of the outer and inner heat balance as shown below in Equation 2-17 ;

$$t_{aj} = \frac{\alpha + \sum_{i=1}^{N} A_i\, h_{ci} t_{si,j} + \rho c V_{infil} t_{oj} + \rho c v_{ventj} t_{vj} + q_{c,intj}}{-b + \sum_{i=1}^{N} A_i\, h_{ci} + \rho c V_{infil} + \rho c v_{ventj}}$$ Equation 2-17

Where

$t_{aj}$ represents the zone air temperature measured in $^{O}$C at time step j

N is the number of surfaces in a zone

$A_i$ is the area of ith surface in m$^2$

$h_{ci}$ is the convection coefficient for ith surface

$t_{si,j}$ is the temperature of surface ith at time step j

P is the air density in kg/m3

Cp is the specific heat of air in (j/kg. K)

V is the volumetric flow rate of air (m3/s)

$t_{oj}$ is the outdoor temperature at time step j ($^{O}$C)

$t_{vj}$ depicts the ventilated air temperature at time step j ($^{O}$C)

$q_{c,intj}$ is the sum od convective flux of internal heat gains at time j (W)

### Inputs of energy simulation tools

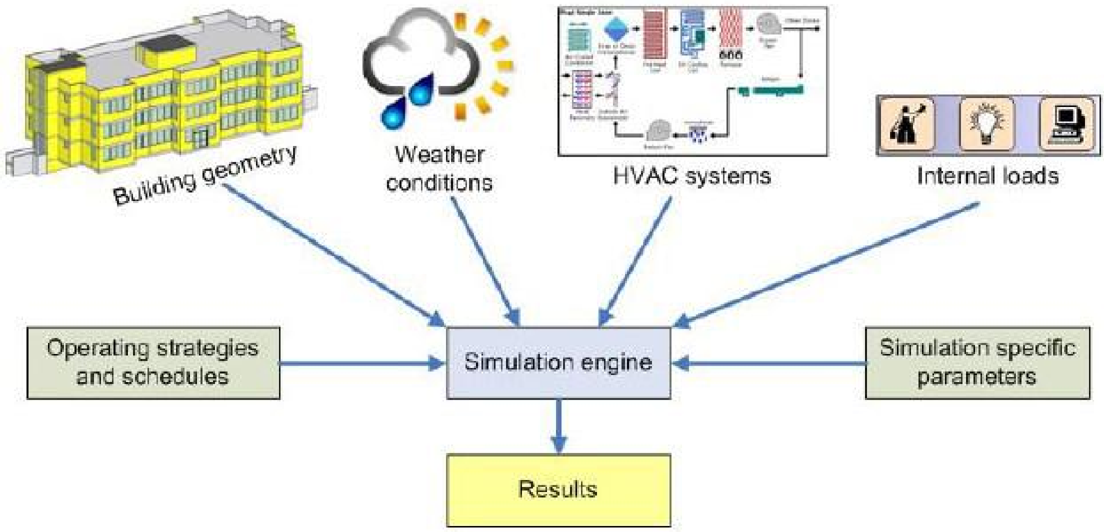



Figure 2-14.Main components and data workflow of building energy simulation tools, adopted from (Maile, Fischer et al. 2007)

Building energy simulation inputs constitute an important part in calculating the heating and cooling loads of the studied building(s). Therefore, this section aims to discuss the relevant ones in detail.

Figure 2-14 depicts the five types of input data that are required to run building energy simulations namely; 3D building geometry, internal loads, HVAC system components and controls, weather data, occupancy schedules and simulation related parameters (Maile et al., 2007).

### 3D Building geometry

It is considered as the essence of energy simulation tools as it permits the extraction of geometric parameters such as zone volume, zone boundary, layered construction data, glazing ration, etc… It is either directly simulated in the same CAD/BIM platform or exported as a **gbxml** file to be analysed, if the simulation software is standalone. However, in all cases, the simulation engine has to convert the 3D building geometry model to a thermal building model. The nature of this transformation depends on the envisaged complexity and accuracy of the simulation tools. For example, in the case of three-dimensional heat transfer analysis that is performed in CFD tools such as COMSOL Multiphysics, each building's space is decomposed into a large number of control volumes with either homogeneous or heterogeneous global mesh. However, since CFD are very complex, time-consuming, and require previous knowledge on fluid dynamics, their use is uncommon for urban planning purposes. Instead, the majority of simulation software used today employs one-dimensional heat transfer (Foucquier, Robert et al. 2013). That is mainly because one assumption of heat balance models is that conductive heat transfer is one dimensional. This approach groups spaces (room) with close thermal characteristics and control patterns (e.g. bedrooms), together, although there might be difference in temperature or pressure within the same space. As a result of those assumptions, the following issues with geometry arise;

- Freestanding walls or columns are not taken into account since it is assumed that there is no difference in temperature between their surface and the room they belong to. Thus, there is no heat transfer which may affect the validity of the simulation results.
- None of the one-dimensional heat transfer tools can represent curved surfaces as such. Instead, an approximation through surface planes is employed as shown in Figure 2-15, which certainly affects the accuracy of the results when complex geometries are involved.
- The conversion of the CAD/BIM models to a thermal one is still challenging and time-consuming since they need to be thoroughly validated beforehand for accurate results.

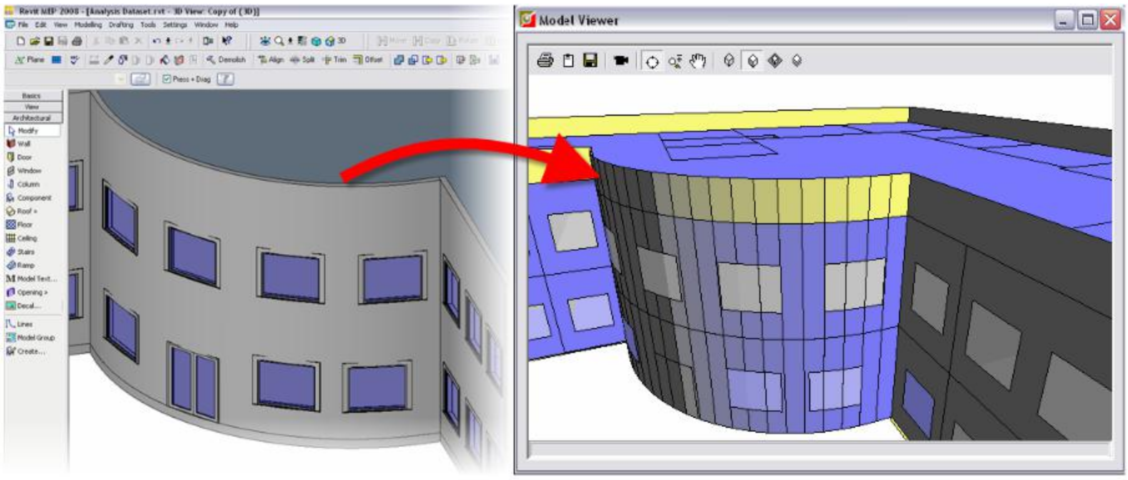



Figure 2-15.Showing the conversion of 3D model into thermal model (Maile, Fischer et al. 2007)

## Weather data

It strongly affects external loads and energy demand of buildings; consequently, have a significant impact on the choice of physical improvement measures (Fumo 2014). For those reasons, they are a fundamental component in every building energy simulation tool. A weather data file contains the typical climatic parameters of various cities around the world which have been gathered and statistically combined. However, since those characteristics are not related to specific years in most of the one-dimensional heat transfer simulation tools, the change in weather conditions over time and within same locations is not taken into account.

Indeed, Lund, (1991) has claimed that "*Generally, the typical year data are valid only for a limited geographical area and should not be considered as a climatological description of a specific town or region*". This can be a real issue for urban planners who are interested in comparing the energy performance of a building (s) in different years. Therefore, it is sometimes preferred to measure weather data directly onsite or from the nearest weather station, when commissioning some projects of high importance (Maile, Fischer et al. 2007). There are a variety of weather data files used in building energy simulations such as TMY. However, it should be noted that for long-term energy prediction, the majority of simulations companies, (e.g. energy plus), recommend two types of files namely; the typical metrological year (TMY) and weather year for energy calculation (WYEC). The comparison of different weather file formats is outside the scope of this research; please refer to the extensive study of Crawley (1998) for more information. Overall, these files, regardless of their formats, include the below parameters;

- **Location data:** name, latitude, longitude (angular distance of a site East/ West of the Greenwich meridian), elevation (altitude), time zone.
- **Site data:** ground reflectance (ground albedo is a measure of how well a surface reflects solar energy. It varies between 0 and 1), terrain type (country, suburbs, city), and wind exposure.
- **Design weather data:** outside min/max air dry-bulb temperatures, min/max wet-bulb temperatures. Dry bulb temperature refers to the ambient air temperature which is measured by a thermometer and not affected by the moisture in the air. On the other hand, wet Bulb temperature is the lowest temperature that can be attained by the evaporation of water into the air. Thus, it is always lower than the dry Bulb temperature (EngineeringToolBox, 2015).

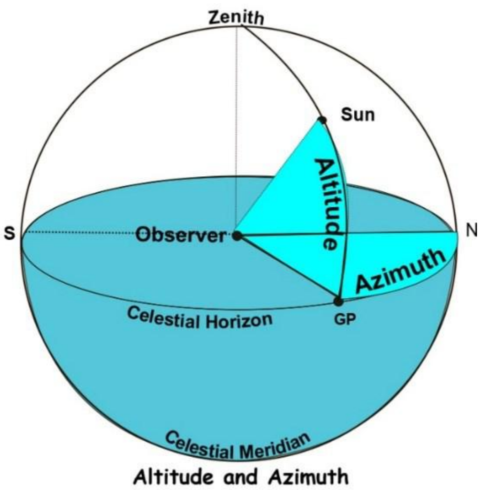



Figure 2-17.Sun altitude and azimuth, adopted from (astronavigation, 2015)

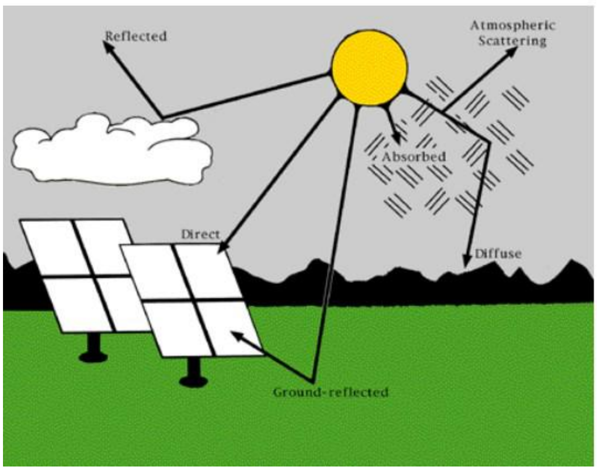



Figure 2-16.Global, direct, diffuse solar irradiance, adopted from (astronavigation, 2015)

- **Performance weather data:**

    - Solar altitude and azimuth: azimuth represents the angular distance of a projected sun (using a vertical circle) from the north or south of horizon (Figure 2-17). Sun altitude; however, refers to the vertical position of sun relative to the earth's horizon as shown in (Figure 2-17) (astronavigation, 2015). Altitude has a significant impact on the temperature levels of a particular region or site. More precisely, the higher the altitude, the colder the temperature is. However, this also depends on the latitudinal position of the site and time of the year. For example, solar altitude will at its maximum during summer (Sciencing, 2014).

    - Air temperature, wind speed and direction.

    - Cloud cover (refers to the proportion of sky covered by clouds).

    Global, direct, and diffuse horizontal solar radiation. First, direct normal irradiance represents the amount of solar radiation by a surface that is perpendicular to the sun rays. However, diffuse horizontal irradiance is the quantity of radiation whose source is not directly from the sun but from atmosphere, instead (Figure 2-16). Finally, global horizontal solar radiation refers to the amount of radiation from shortwave received by a horizontal surface to the ground (FirstGreen, 2012). These parameters play an important role in evaluating and optimising the performance of PV panels.

    - Atmospheric pressure (or barometric pressure): refers to the force exerted on a given surface by the air surrounding it.

**HVAC system components and controls**

In the past few years, the realistic representation of a given HVAC system within a predefined structure use to be a very challenging task in older tools such as DOE-2. However, this complexity has been simplified with newer tools which give users the ability as well as flexibility to modify, import, or even create new HVAC technologies such as, under floor heating. In addition to HVAC components, control strategies are an integral part of HVAC components input since they govern their behavioural pattern. Subsequently, they have a significant impact on the simulation results. Thus, surveys on domestic energy sustainability and energy audits usually take into account the type of controls in the dwelling.

#### Internal gains and occupancy schedules

They include the energy load generated by people, lights, and different appliances. This varies from one space to another depending on the nature of the performed activities and the behaviour of occupants (Guerra Santin 2011). However, there is a major problem that is presented in the modelling of those parameters. Regardless of assumptions that have to be made about occupants' behaviour, other factors like socio-economic and cultural, which directly affect this behaviour and suppose to diminish the assumptions related to it, are not taken into account.

#### Simulation related parameters

They influence the numerical behaviour of building energy simulation tools. Convergence tolerance, timeframe, and the time step are examples of simulation specific parameters.

### 2.3.1.2 The Calibration of engineering tools

In the light of the above, it is clear that the inputs related to weather and internal loads data in any building energy simulation tool are based on assumptions. Considering the envisaged level of accuracy, this might result in major discrepancies between the measured and simulated data during the operational phases. Therefore, a calibration process, which consists of "tuning" or adjusting the simulation inputs with measured data, is applied (Agami Reddy 2006). However, due to the fact that calibration is a highly under-determined issue, there is no single uniform solution or model (Scofield 2009). Agami Reddy (2006) classified different calibration approaches into the following four distinct categories as depicted in Table 2-5 below. It should be noted that the below categories can be combined as mentioned by Mustafaraj, Marini et al. (2014) *"All these methods are not exclusive and could be coupled (e.g. use of graphical and statistical analysis methods to support iterative manual calibration...)"*. For more information on approaches to the calibration of building energy simulation tools, please refer to Appendix A.1.1.

Table 2-5.shows the different types of calibration and provides some examples of each category.

| Calibration Category | Example of case studies |
|---|---|
| **Manual, iterative and pragmatic intervention** | (Kaplan, McFerran et al. 1990); (Diamond, Hunn 1981) (Hunn, Banks et al. 1992) |
| **A suite of informative graphical comparative display** | (Haberl, Bou-Saada 1998) (Bronson, Hinchey et al. 1992) (Haberl, Sparks et al. 1996) |
| **Special tests and analytical procedures** | (Manke, Hittle et al. 1996) (Liu, Claridge 1998) |
| **Analytical/mathematical methods of calibration** | (Sun, Reddy 2006) (Westphal, Lamberts 2005) |

### 2.3.1.3 Lifecycle stages relevance

Table 2-6.The applicability of some well-adopted energy simulation tools in the building lifecycle

| Simulation tools | Lifecycle usage | Example of relevant case studies | Purpose of use in the study |
|---|---|---|---|
| **EnergyPlus** | Multiple phases but powerful for operational phases | (Attia, Gratia et al. 2012) | To simulate the energy performance of different design proposals. |
| **eQUEST** | Design stages (mainly schematic/ detailed design phases) | (Yu, Yang et al. 2008) | To estimate the annual energy consumption of a typical residential building in hot summer zones in China |
| **Design builder** | Multiple stages but weak in operational phase | (Ortiz, Bonnet et al. 2009) | To estimate the environmental impact of certain physical measures on a typical dwelling in Barcelona during the design and operational stages |
| **IES-VE** | Multiple stages (design, operation, maintenance) | (Al-Tamimi, Fadzil 2011) | To investigate the effect of different shading devices on the thermal comfort of high-rise buildings in hot-climate weather |
| **BSim** | Design stages | (Marszal, Heiselberg et al. 2012) | To calculate the hourly load demand of a reference residential multi-storey zero emission building in Denmark. |

Table 2-6 represents the life-cycle usage of some widely adopted energy simulation tools and puts them into context by providing examples of previous studies. Traditionally, Engineering tools were developed to support the design phases. Thus, their usage was only confined to evaluating the energy efficiency of different design alternatives. However, over the last few years their capabilities have been expanded to cover the commissioning, operation, and maintenance of buildings (Coakley, Raftery et al. 2014; Crawley, Hand et al. 2008). The difference between both categories is that the tools aiding operational and maintenance stages require a higher level of detail to validate the performance of HVAC systems and their controls during commissioning (Maile, Fischer et al. 2007).

#### 2.3.1.4 Advantages of engineering approach

- Building energy simulation tools do not rely on historical data, albeit the latter can be used to increase their accuracy (Swan, Ugursal 2009).
- They have a great ability to model different onsite energy renewable sources such as solar panels.
- Easy to use for evaluating the energy performance of different design proposals.
- Flexible, explicit, well integrated and highly interoperable with powerful CAD/BIM systems which make their adoption widely accepted by urban planners.

#### 2.3.1.5 Limitations of engineering methods

- May require more inputs than other methods like statistical one.
- If certain data in unavailable, the performance of very accurate simulation results can be tedious, inefficient, and require experts' assistance (Zhao, Magoulès 2012).
- The modelling of occupants' behaviour is poor and inaccurate due to the "standardisations" as well as assumptions.
- One major drawback of this method is that the demographic evolution and the socioeconomic variations of the occupants are not taken into account. This argument was supported by Guler et al. (2008) in the following statement *"A*

*difficulty with this method is the inclusion of consumer behaviour and other socioeconomic variables that have a significant effect on the residential energy use".*

### 2.3.2 Artificial intelligence approaches

Although the artificial intelligence umbrella groups a wide range of approaches, SVM (*support vector machine*) and *artificial neural networks* models (ANN) are the most used for the prediction of energy consumption in the residential sector (Zhao, Magoulès 2012).

#### 2.3.2.1 Artificial neural networks approaches

*A. Structure and workflow*

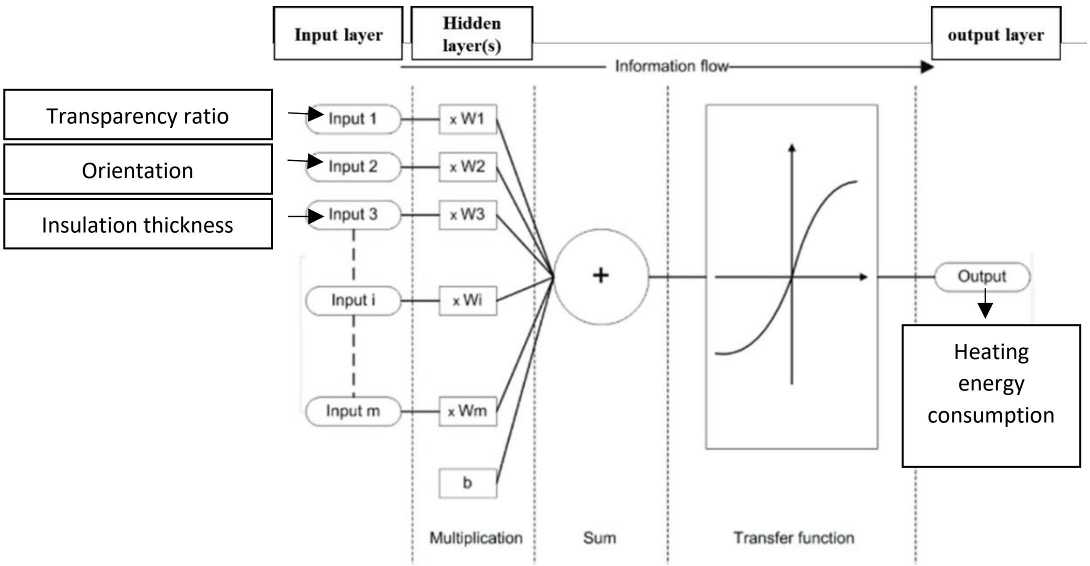



Figure 2-18.Working principle and information flow of neural network models adopted from (Krenker, Kos et al. 2011)

ANN's are simplified mathematical models that mimic biological neural network (Guler et al., 2008). They are not programmed conventionally but instead trained using historical data representing the performance of a system. As illustrated in Figure 2-18, their generic structure comprises three main components namely; input layer, one or several hidden layers, and an output layer. Every single layer encompasses several neurons in which each one is linked to other ones in the adjacent layer with various weights. Overall, information flows into the input layer, through the hidden layer(s), to the output layer. Each entering signal, which comprises a given parameter value such as, insulation thickness (cm), or glazing ratio (%), is then multiplied by its corresponding neuron weight and summed up with the bias contribution to determine the total input of hidden layer neuron (netj). An activation function will be applied to the latter to define the output of neurons, electricity consumption, for instance. However, it should be noted that this is just a generic description because the relationship between the three layers can vary widely from one ANN model to another. For example, in the *Feed-forward ANN* models, the flow of information is mono-directional with no back loops, whereas in the *Recurrent ANN* models, information is transmitted forwards and backwards as presented in Figure 2-19.

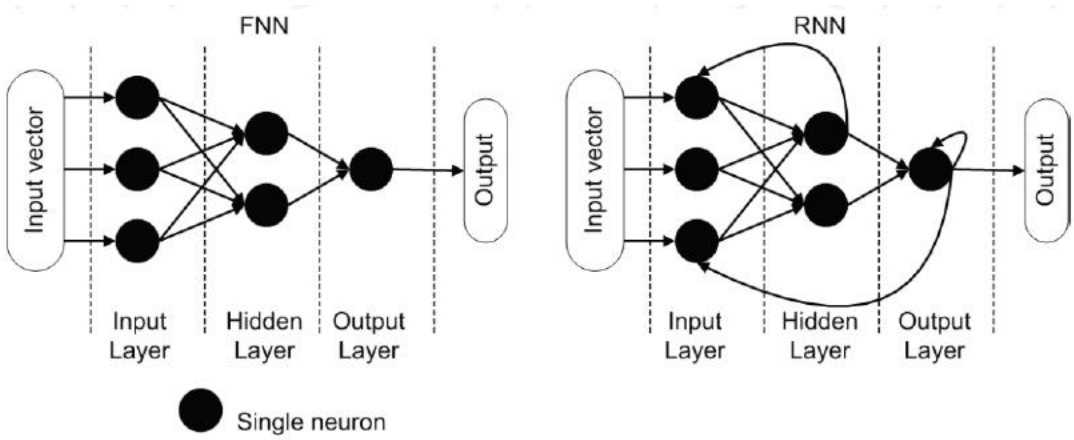



Figure 2-19.The structure of Feed-forward (left) and Recurrent artificial neural network models (Krenker, Kos et al. 2011).

*B. Applicability and Life-cycle stages relevance*

A generous amount of work has been devoted to using ANN models in the residential sector since originated in the 1990's. It has been mainly confined to utility load forecasting due to their ability to represent nonlinear processes. In general, the scale of those studies ranges from a single dwelling to few ones, although there is an exception of one work of (Aydinalp, Ismet Ugursal et al. 2002) on the Canadian national scale. Table 2-7 (below) summarises some key studies in this particular area based on the structure of the ANN model, its inputs, training process, and predicted variable(s).

For example, Kalogirou and Bojic (2000), to assist designers in the process of assessing the efficiency of passive solar design, employed four layers recurrent ANN model to predict the hourly energy consumption of a passive solar holiday home in Cyprus during summer and winter. This was achieved in function to the following variables (inputs) namely; the level of insulation, masonry thickness, nature of heat transfer coefficient (constant/ variable), and time of the day. The author concluded that model's prediction was satisfactory and quick than dynamic simulation tools. Two years later, Mihalakakou et al. (2002) developed feed forward back propagation ANN model for estimating the hourly electricity consumption of a residential building in Athens based on several climatic parameters such as air temperature and solar radiation. The model was trained with hourly energy consumption data of the building collected over five years. Although the predicted results were satisfactory, the potential of the collected multi-year data in showing the annual changes was not utilised as dates were not entered as an input.

In addition to load forecasting at the operational phases, ANN models were also employed at early design stages. For instance, Ekici and Aksoy (2009) developed a back propagation ANN model to predict the heating demand of three building designs based on orientation, insulation, walls thickness, and transparency ratio. Energy consumption data, which were calculated using the finite difference method of transient state one-dimensional heat conduction, were employed to train the model inside MATLAB ANN toolbox. Following the same process, Dombaycı (2010) predicted the hourly energy usage of a house during the design stage but the used ANN model was trained with energy consumption data calculated by degree hour method. This calculation was based on temperature record obtained from a local metrological station and covering the period from 2004-2007.

Table 2-7.some relevant examples on ANN energy prediction models in the residential sector

| Researcher(s) | Type of ANN model | Inputs | Estimated parameters | Training datasets/ data resolution |
|---|---|---|---|---|
| **(Kalogirou, Bojic 2000)** | **Four layers** recurrent ANN | -level of insulation<br>-masonry thickness<br>-heat transfer coefficient<br><br>-time of the day | The hourly energy consumption of a passive house | Trained with 1-year data |
| **(Mihalakakou, Santamouris et al. 2002** | feed forward back propagation ANN | -hourly air temperature<br>-total solar radiation | The hourly electricity consumption of a residential building in Athens | Trained with five years hourly energy consumption data |
| **(Ekici, Aksoy 2009)** | a back-propagation ANN | - orientation<br>- glazing ratio<br>- level of insulation<br>- wall thickness | Heating demand of three design proposals | - |
| **(Dombaycı 2010)** | a back propagation ANN | - orientation<br>- glazing ratio<br>- level of insulation<br>- wall thickness | Hourly energy consumption of a house during the design stage | Trained with three years energy consumption data |
| **(Olofsson, Andersson 2001)** | feed forward back propagation ANN | -Indoor/outdoor daily temperatures<br><br>-Heating energy demand<br><br>-electricity energy consumption | Heating energy demand of single family dwellings | - |

Apart from load prediction, ANN models have been also utilised in optimising or enhancing the performance of energy management systems. For example, in an attempt to save the electricity used for water heating, Wezenberg and Dewe (1995) adopted some ANN models in generating an operation schedule for a residential water heating system based the cheapest tariffs and without compromising the thermal comfort. This was performed by training the models with historical tariff rates, ambient temperature, humidity rates, and time data (hour, day, month, year, specific holidays). Other applications than energy forecasting in the residential sector such as control of power generation systems as well as solar radiation and wind speed prediction, is outside the scope of this research. Therefore, please consult the work of (Kalogirou 2006).

*C. Advantages of artificial neural network tools and approaches*

- In contrast to statistical approaches, the ANN ones have a higher prediction accuracy due to their ability to implicitly identify all nonlinear relationships between inputs and outputs (Swan, Ugursal 2009).
- The process of ANN model training can handle several algorithms (Aydinalp-Koksal, Ugursal 2008).
- In a widely cited work by Aydinalp et al. (2002), the developed ANN models showed their ability in evaluating the impact of socio-economic factors on end-energy use.

*D. Disadvantages of artificial neural network tools and approaches*

- ANN models have a limited ability to make the relationship between variables explicit. Furthermore, they cannot directly deal with uncertainties (Guler et al., 2008).

- ANN modelling is not cost effective since enormous historical data and computational resources are required (Zhao and Magoulès 2012).
- ANN models are exposed to overfitting issues in the data training process, which means that the trained ANN model might consider noise as part of the data pattern. As a result, predictions made are outside the range of the training (Tu 1996).
- They are not flexible in assessing the effect of building conservation measures
- In contrast to engineering approaches, ANN models cannot be generalised to different buildings under various conditions (e.g. Weather and occupancy). This requires a new ANN model for each building/ situation (Foucquier et al., 2013).

#### 2.3.2.2 Support vector machine (SVM) approaches

Support vector machine (SVM) is a supervised machine learning algorithm that is well-known for its robustness and accuracy. Like ANN, SVM models have to be trained with historical data representing the behaviour of a system. Due to its ability to solve nonlinear regression and classification problems, SVM is increasingly being employed both in research and industry since first implemented by (Cortes and Vapnik (1995). The principle of SVM in solving a classification problem is based on the division of a dataset into sub categories, whose properties are specified by the user. On the other hand, it is premised on describing a given dataset by a particular equation whose complexity is defined by the user, if the main concern is to tackle a regression problem to forecast future trends in the data. The prediction principle using SVM will be addressed in detail in the next section.

*A. Prediction principle using SVM regression*

The aim of using SVM for regression is to seek the optimal model generalisation to promote sparsity. Let us consider a given training dataset$[(x_0, y_0), \ldots, (x_n, y_n)]$. $x_i$, $y_i$ represent the input and output space, respectively. For example,$x_i$ denote household income values, whereas $y_i$ are energy consumption values. To solve a non-linear problem such as, energy consumption, the non-linearity between variables x and y has to be transformed using linear mapping (or transformation) through two steps. The first step consists of projecting the non-linear problem into a high dimensional space known as the feature space. After that,$f(x)$ the function fitting best the behaviour of the problem has to be determined in the feature space. The uniqueness of SVM is that it allows an uncertainty $\in$ in the regression model. This means that any error less than $\in$ in the SVM model is tolerated. The function f(x) is defined as follows ( Equation 2-18) (Cortes and Vapnik, 1995);

$$f(x) = < \omega, \varphi(x) > +b \qquad \text{Equation 2-18}$$

Where $\varphi$ is a variable in the feature space (high-dimensional space) and <, > is a scalar product. b and $\omega$ are defined in function of an optimisation problem known as the primal objective function Equation 2-19 and Equation 2-20 (Cortes and Vapnik, 1995).

$$min_{\omega,b,\xi_i\xi_i^*} \quad \frac{1}{2}||\omega||2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \qquad \text{Equation 2-19}$$

$$\text{Subject to} \quad \begin{cases} y_i - \langle\omega, \Phi(X_i)\rangle - b \leq \epsilon + \xi_i \\ \langle(\omega, \Phi(x_i)\rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_{i|}\xi_i^* \geq 0 \end{cases} \qquad \text{Equation 2-20}$$

$C$ (known as regularisation parameter), represents the trad off between the flatness of $f$ and the first larger value than $\epsilon$ defined by the user. $\xi_i$ and $\xi_i^*$ are the slack variables enabling the constraints flexibility (Figure 2-20). Finally, the second step to make the complex non-linear map a linear problem is to apply a kernel function without the need to evaluate$\Phi(X)$ as follows Equation 2-21 (Cortes and Vapnik, 1995);

$$k(x,\acute{x}) =< \varphi(x), \varphi(\acute{x}) > \quad \text{Equation 2-21}$$

There are three types of Kernel functions namely; linear, polynomial and radial. Choosing the right kernel is very important since it has an enormous impact on the learning ability and generalizability of SVM algorithm which in turn affect its prediction accuracy.

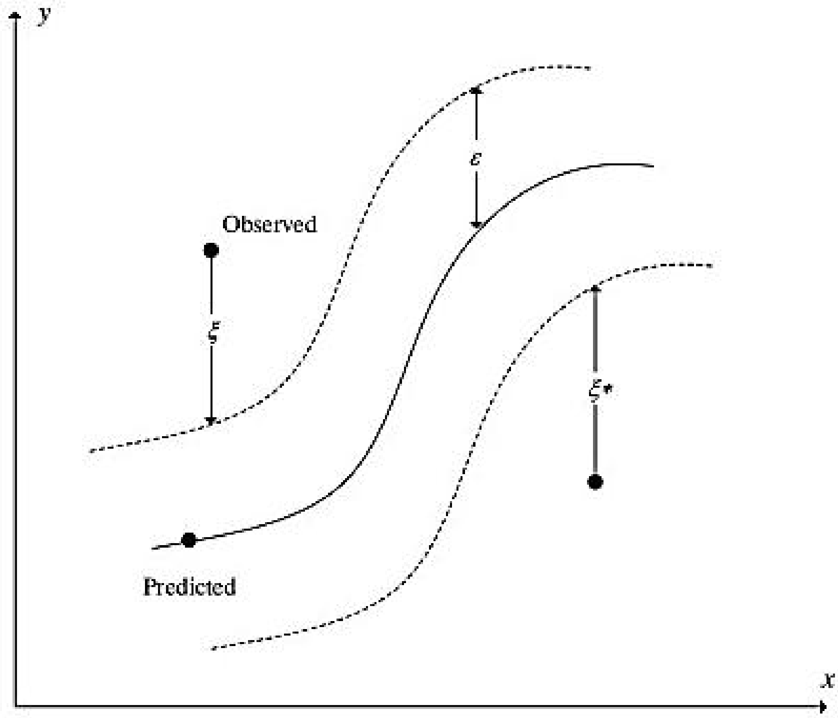



Figure 2-20. SVM regression function with slack variables, observed and predicted data points (tripod, 2015)

*B. Applicability in the building lifecycle*

Since first adopted by Dong et al. (2005) for energy forecasting until now, the use of SVM models has been mainly confined to predicting energy consumption and temperature of individual buildings (mostly commercial or administrative) during their operational stages (Ahmad et al., 2014). However, recent work done by (Son et al. (2015) used SVM regression models to predict the electricity consumption of government owned buildings in the early design stages. This was achieved by first retrieving the relevant parameters by applying a variable selection algorithm called RreliefF. Afterwards, the SVM model was trained with an existing dataset for 175 government owned buildings. The trained SVM could predict the energy consumption of public buildings during the design stages but with Mean Absolute Percent Error (MAPE) of 35% which means that the SVM prediction accuracy is reasonable according to Lewis scale (Lewis, 1982).

On the other hand, many publications have addressed the prediction of electricity consumption in non-residential buildings during the operational stages using SVM regression models. For example, Lai et al.(2008) forecasted the electricity consumption of a building over three months after the SVM model had been trained with electricity consumption data measured over one year period. The authors obtained a satisfactory match between measured and predicted data. Similarly, Li et al.(2009) compared the prediction accuracy of a back propagation neural network with an SVM model while forecasting the cooling load of an office building in Guangzhou in function of outdoor dry-bulb temperature, humidity, and solar radiation. The results have shown that the SVM model had a better prediction accuracy than the ANN model. Similar findings were reported one year later by (Li et al. (2010) after the following models namely; SVM, back propagation neural network, radial function neural network, and general regression neural networks have been employed to predict the annual electricity consumption of 9 office buildings.

In an attempt to address the lack of studies applying SVM to residential buildings, Jain et al.(2014) questioned the applicability of SVM regression models for forecasting the energy consumption of multi-family buildings. This was achieved by investigating the accuracy of the employed SVM regression models in function of different temporal (e.g. Hourly and daily) and spatial granularity of the measured electricity consumption data. Their findings suggested that SVM can be extended to cover energy forecasting of multi-family buildings as long as the measurements used to train the SVM model are performed at least at every floor and with a minimum of 1-hour interval.

*C. Advantages of support vector machine approaches*

- Comparatively to existing artificial intelligence approaches, SVM for regression have shown better accuracy.
- In contrast to artificial neural networks, SVM for regression are less prone for over-fitting issues due to their regularisation parameter $C$.

*D. Disadvantages of SVM tools and approaches*

- One of the limitations of SVM for regression is the lack of a universal method for selecting the appropriate Kernel function (Yilmaz et al., 2015).
- SVM are not cost-effective since they are computationally more expensive to train and test when using large datasets.
- SVM for regression do not rely on 3D models. Hence they are not flexible in assessing energy conservation measures, and their adoption by urban planners is complex.

### 2.3.3 Hybrid methods

Hybrid or Grey-box models, as called by others, were first introduced in the early 1990s to improve HVAC control systems efficiency (Foucquier et al., 2013). Generally speaking, they are based on the idea of combining physical, statistical, and artificial intelligence models when data samples are reasonably small, incomplete, or subject to uncertainties (Zhao and Magoulès, 2012). This leads the way to three possible strategies (Table 2-8). The first one involves the estimation of **optimal** physical parameters with the help of machine learning algorithms through the combination of one-dimensional heat transfer models with optimisation algorithms (Genetic Algorithms usually). On the other hand, the second one consists of describing the dwelling behaviour by implementing a learning model through the use of statistics. Finally, the third approach, which not referenced in building energy prediction literature according to Foucquier et al. (2013), comprises the employment of statistical models in areas where physical/ thermodynamic models are inadequate or inaccurate. For example, when the thermal properties of a given room are unknown.

Table 2-8.Possible strategies of hybrid approaches

| Possible hybrid approach | Approaches involved | Common purpose of use |
|---|---|---|
| 1 | • Engineering<br>• Artificial intelligence | Estimation of optimal physical parameters |
| 2 | • Artificial intelligence<br>• Statistical | Describing the building thermal behaviour |
| 3 | • Statistics<br>• Engineering | Estimation of optimal physical / techno-socioeconomic parameters which could have been missing or inaccurate. |

#### 2.3.3.1 Applicability and Life-cycle stages relevance

From analysing the dedicated literature, grey-box models are mainly used for the estimation of optimal physical parameters and energy consumption prediction to a lesser extent. However, despite the fact that they are not directly linked to energy prediction modelling, the determination and optimisation of such parameters to achieve low-energy consumption, certainly assist urban planners in their physical improvement strategies. This does not only apply to the design stages but also the operational ones. For example, Yuce and Rezgui (2017) combined Artificial neural networks (ANN) with a Genetic algorithm (GA) as an attempt to reduce the energy consumption in buildings without affecting the thermal comfort of elderly occupants. To achieve that, the authors first compensated the lack of historical energy consumption data on the pilot building by utilising building energy simulation tools such as Design builder. More precisely, the 3D model of the select building was simulated under different conditions (e.g. Lighting and temperature), and the inputs and outputs related to each simulation scenario were exported in a separate file. Next, the scholars identified the most influential parameters on energy consumption by using principal component analysis (PCA) and multiple regression analysis. Based on those factors, an Artificial neural networks energy prediction model was developed. Finally, a Genetic Algorithm-based optimisation tool, which uses the trained ANN model as a predictor (cost function engine), was built to generate optimised energy savings rules. The authors reported 25% of energy reduction in the pilot building located while satisfying the thermal comfort preferences of the occupants.

Similarly, Znouda et al. (2007) used Hybrid models to assist the energy planning in the design stages. The authors coupled an energy simulation engine, which was specifically developed for Mediterranean weather namely; CHEOPS, with a genetic algorithm to retrieve the optimum architectural and physical parameters (e.g. Wall thickness, level of insulation and orientation); consequently, improve energy efficiency in summer and winter. The strength of this approach lies in its simplicity as only a few parameters are involved. However, other variables related to occupancy and which have a noticeable impact on energy saving were absent in this study.

On the other hand, Tuhus-Dubrow and Krarti (2010) followed a similar procedure but with a different energy simulation engine (DOE-2). More specifically, the authors determined the most efficient dwelling form(s) for amongst U shape, L shape T shape, H shape, trapezoidal, and rectangular across five different US climatic zones. The findings suggested that the two latter shapes were the most efficient.

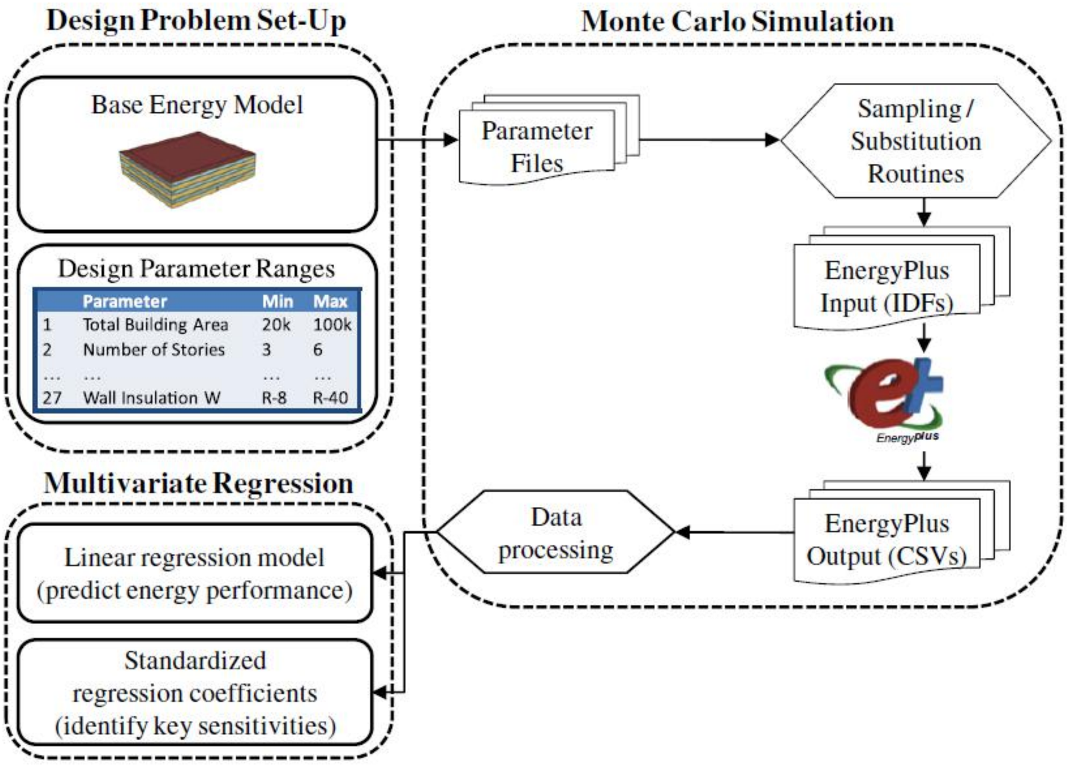



Figure 2-21.The structure of the hybrid energy prediction model developed by (Hygh, DeCarolis et al. 2012)

Apart from combining artificial intelligence algorithms with engineering methods, there also studies that merged statistical regression methods with building simulation tools. For instance, Hygh et al. (2012) proposed a multivariate regression model for predicting the energy consumption of medium size rectangular shape dwellings in 4 major climatic zones in the US. As shown in Figure 2-21, It was achieved by first, investigating the range of parameters that are frequently changed at the early design stages and likely to affect energy consumption through analysing the dedicated literature. These factors include; building area, number of stories, aspect ratio, orientation, glazing ratio, and U-values of walls and windows. Secondly, based on the possible min-max values of each parameter, Monte Carlo simulation was adopted to incorporate all possible inputs probabilities and replace default values in the referential thermal model. Thirdly, all the resulting iterations were embedded in EnergyPlus simulation engine using Perl script before computing the annual energy consumption of each scenario. Afterwards, the rich data base generated by Monte Carlo simulation, which contains the latter and the randomly selected values for each parameter, was used to develop multivariate linear regression model. Similar approach was also adopted by Asadi et al.(2014).

As for operational phases, Guo et al. (2011), for instance, applied a grey model to forecast the energy consumption of a water heat pump in some dwellings. The experiment also tested the impact of data sampling interval on the accuracy of the prediction. The authors concluded that four weeks was the best interval.

#### 2.3.3.2 Advantages of Hybrid approaches

- Grey-box models are an excellent alternative to ANN and statistical regression models when there is a limited number of parameters, or data samples are reasonably small (Foucquier et al., 2013).

- Unlike engineering methods, the hybrid ones do not require a detailed description of the building geometry and thermal properties of its envelope. Furthermore, their outcomes can be interpreted from a physical point of view (Foucquier et al., 2013).

#### 2.3.3.3 Disadvantages of Hybrid approaches

- Hybrid models require a considerable amount of computation time and resources.

- Although Grey models offer great flexibility, their implementation for urban energy planning purposes requires a high level of support and training because they combine two distinct approaches.

## 2.4 Urban scale (bottom –up) approaches

Urban energy forecasting models to assist physical improvements are mostly bottom-up models (Heiple, Sailor 2008). This implies that the energy demand or CO2 emission of the residential stock of an urban area are forecasted in function of a representative sample usually via extrapolation. This is often referred to as "Archetyping" or "Clustering" in the literature (Jones, Lannon et al. 2001). It should be noted that other bottom-up models described by Swan and Ugursal (2009) utilising appliances distribution and samples are analysed in chapter 3 due to their support to socio-economic improvements.

From examining recent literature, three main approaches namely; GIS energy forecasting, CityGML based semantic energy modelling, and Bottom-up statistical models, were distinguished. These models are mainly employed in the calculation of energy demand, identification of areas with potential energy savings, and the evaluation of certain measures' effectiveness (Saran et al., 2015). However, since their prediction accuracy depends heavily on the inputs' level of detail, all of these methods are still facing two major obstacles. The first challenge is related to the lack of data availability which leads to either making assumptions or utilising time-consuming data collection methods such as surveys to compensate this shortage. Secondly, there is an issue of computational resources which can be only solved by finding the right compromise between the outcomes' accuracy and computation time. The subsequent sections will discuss these approaches in more depth in terms their structure and prediction pipeline while putting them into context through analysing relevant examples.

### 2.4.1 2D GIS based urban energy planning and forecasting models

#### 2.4.1.1 Structure

In general, the majority of GIS based urban energy prediction models, despite their diversity, are composed of the following six components;

- **Inputs:** mostly comprise variables which describe the physical characteristics of the dwellings and standard occupational schedules data. Variables like dwelling age, dwelling typology, heated ground floor area, are the most encountered in the literature (Mavrogianni et al., 2009; Rylatt et al., 2003; Jones et al., 2001). The collection of such data is attained through different types of surveys ranging from visual inspections to energy audits, and from GIS mapping tools or existing databases (e.g. government data).

- **Assumptions:** usually apply to some variables which describe occupants' behaviour as well as the thermal performance of dwellings such as household size, patterns of presence at home, patterns of heating, level of insulation, U-values, type of HVAC systems, main fuel for heating, efficiency of HVAC systems, and water tank size (Swan and Ugursal, 2009; Jones et al., 2001).

- **Baseline models:** The estimation of energy consumption at the urban scale (in a bottom up fashion) is produced with the help of baseline models which are usually physically based ones such as BREDEM (the building research establishment domestic energy model). The inputs requirements and complexity of energy prediction varies from one baseline model to another. Reviewing baseline models is beyond the scope of this research; please refer to the work of (Kavgic et al., 2010).

- **Calculation engine:** it is the software in which baseline models or algorithms are incorporated to perform energy consumption calculations. SBEM (simplified building energy model) is an example of calculation engines.

- **Data exchange module:** it allows the transfer of data between the GIS platform and calculation engine. For instance, Mavrogianni et al. (2009) adopted a dynamic data exchange (DDE) to transfer data between the BREEM spreadsheet and GIS database.

- **GIS platform:** Although it provides certain geometric inputs, a GIS platform is mainly dedicated to visualising the energy consumption or CO2 emission of buildings using thematic or choropleth mapping techniques (Batty, 2007)

### 2.4.1.2 Conventional prediction Pipeline

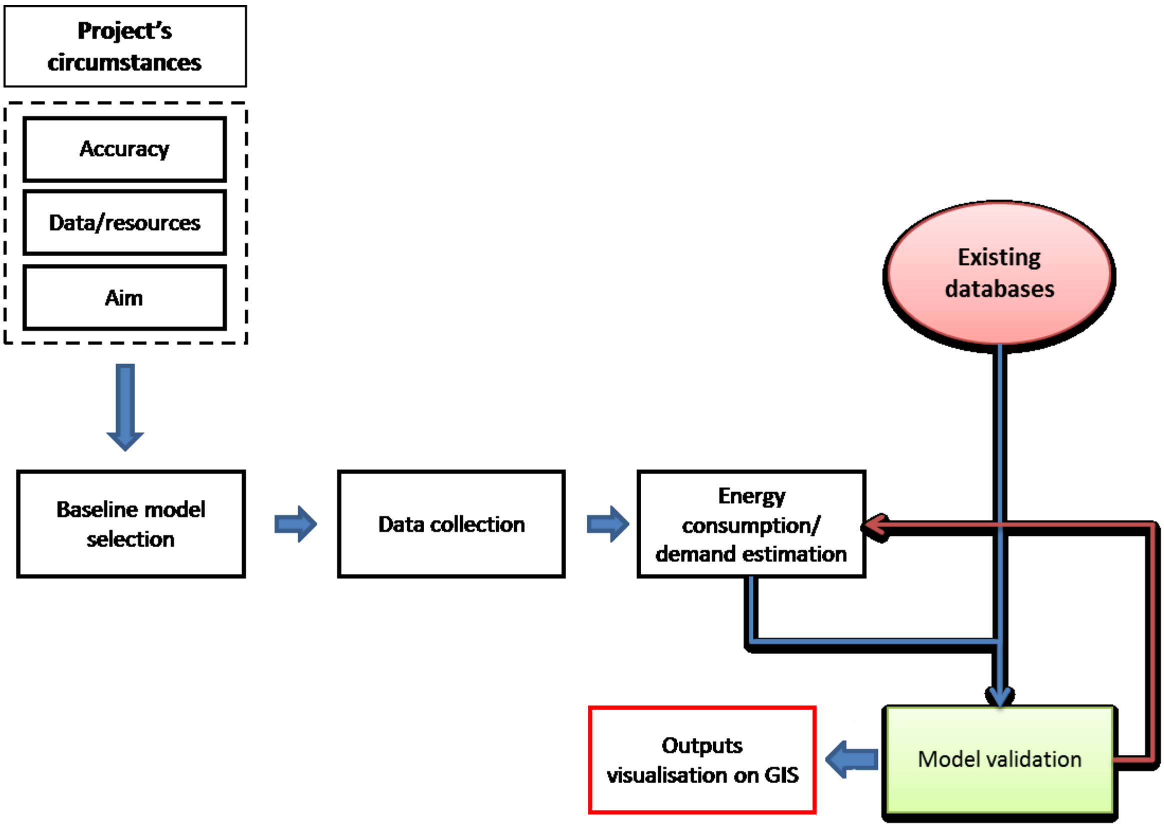



Figure 2-22.Conventional prediction process of 2D GIS based urban energy planning and prediction models

Generally speaking, the process begins with determining the type of the baseline model and its calculation engine based on the aim of prediction, availability of inputs, envisaged accuracy, time, and availability of resources Figure 2-22. Once determined, although not necessary if it is available, data will be collected using various methods as mentioned at an early stage. After that, the geometric variables extracted from GIS along with the collected data will be inserted to perform the calculation. This will be followed by a validation process that will compare the outcomes with existing databases such as regional statistical data (top-down consumption data) (Mavrogianni et al., 2009). If there were not a good agreement between both datasets, inputs would be checked and adjusted accordingly. Finally, with the help of data exchange module, GIS platform will automatically geolocate and visualise the outputs on the 2D map of the studied urban area. However, this process can vary from one study to another. Therefore, a comparative analysis of the prediction processes of two well-cited examples has been presented in appendix A.1.2.

#### 2.4.1.3 Applicability in the building lifecycle

2D GIS based tools are mainly employed in the operational phases of the building lifecycle to estimate and compare the energy consumption or CO2 emission of building areas before and after physical improvement measures (e.g. renewable resources, retrofit, etc…) have been applied. Furthermore, to determine building areas with potential energy savings. For example, Jones et al. (2001) estimated the annual energy consumption of Neath Port Talbot residential sector, UK, in function of a representative sample composed of 100 archetypes of dwellings. After that, certain physical measures included in the UK home energy conservation act (HECA) such as, water tank insulation, were applied to the archetypes whose energy consumption values were re-estimated accordingly. Finally, the archetypes were extrapolated in GIS to thematically visualise the energy consumption and CO2 emission of the whole residential sector before and after physical improvements. Similarly, Heiple and Sailor (2008) estimated the energy consumption of the building stock in Houston, the USA in function of 30 archetypes composed of 8 dwellings and 22 commercial buildings. Secondary databases, more precisely residential energy consumption survey (RECS) and GIS lot tax database, were used to retrieve geometric parameters, thermal characteristics, and age of buildings. After the total energy consumption had been calculated using a building simulation tool eQUEST, the outcomes were mapped into a GIS platform to visualise the hourly energy consumption at 100 m grid resolution.

2D GIS based approaches have also been highly associated with renewable energy planning in the literature. For instance, Aydin et al. (2013) and Janke (2010) adopted 2D GIS to identify suitable areas for solar and wind farms, as well as pumped storage hydro plants. Kucuksari et al. (2014) has recently suggested a framework integrating 2D GIS, optimisation algorithms, with simulation engines to assist the placement and dimensioning of solar panels in dense urban areas. In addition to that, 2D GIS approaches played a significant role in evaluating the benefit of expanding district heat systems in particular studies Nielsen and Möller (2013).

#### 2.4.1.4 Advantages of 2DGIS approaches

- Have the potential of being integrated into urban planners' everyday life, since GIS tools are increasingly being adopted in urban planning.
- They are good for analysing and communicating potential renewable energy areas.

#### 2.4.1.5 Limitations of 2DGIS approaches

- There is a lack of universal guidelines on the processes, nature, and granularity of parameters involved in this approach.
- Energy forecasting using 2D GIS based approaches require a considerable amount of computation time of resources since they rely on building simulation tools.

### 2.4.2 CityGML (3DGIS) urban energy planning and forecasting models

The growing interest in 3D CityGML models over the recent years, the widespread use of laser scanning, and the availability of high-quality sensing data (e.g. LIDAR) have led many researchers recently to shift the focus from conventional 2D GIS urban energy planning models towards CityGML (3DGIS) models. This change was mainly characterised by exploring its spatio-semantic capabilities to improve or assist urban energy planning decision making (Strzalka et al., 2011; Krüger and Kolbe, 2012; Gröger and Plümer, 2012). Before addressing the structure and prediction process of 3D CityGML urban energy forecasting models, it is important to provide a brief overview of CityGML, its thematic modules, and geometry representations.
CityGML is an XML object-oriented information modelling approach for representing, storing and sharing 3D city model data. It provides the standard mechanisms that govern the description of different 3D objects in relation to their geometry, semantics, and topology. To describe and organise all urban environment features, CityGML is equipped with ten core thematic modules including building, vegetation, relief, and others as illustrated in Figure 2-23. However, other features such as the ones related to energy forecasting or noise propagation, can also be added through application domain extensions (ADE) (Krüger, Kolbe 2012).

The building module, which is the backbone of CityGML, permits the geometric as well as semantic representation of buildings and their elements (Kolbe, Gröger et al. 2005). Buildings in CityGML possess the following attributes namely; *class, address, usage, function, roof type, building height, number of floors, year of construction, year of demolition, and cross-reference number*, in which the three later attributes are shared among all types of buildings and their components. The cross-reference number facilitates the update of objects' features and also the extraction of additional information from other databases.

Apart from semantic properties, the geometric description of buildings is built upon the GML 3.1.1 Model (ISO 19107) developed by Cox et al. (2002). Therefore, buildings can be represented as solid or multi-surface geometry. However, solid representation is preferable for energy related applications since the calculation of volumes is a straightforward task (Gröger, Plümer 2012). Furthermore, boundaries or surfaces shared between two solid objects can be easily identified due to the XML link concept (Kolbe, Gröger et al. 2005).

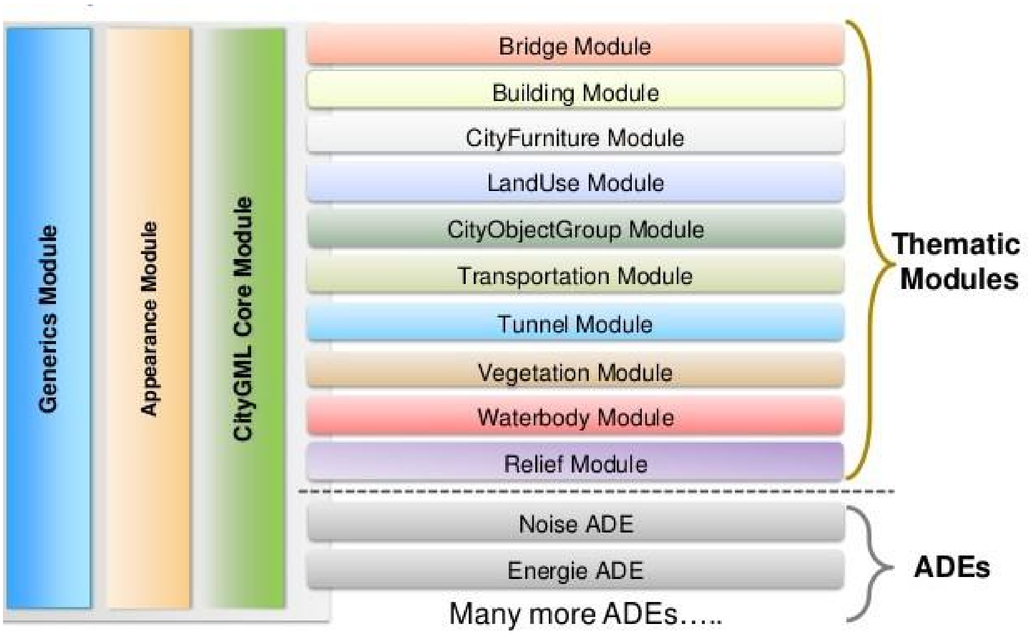



Figure 2-23.CityGML thematic core and ADE modules (Gröger, Plümer 2012)

One advantage of CityGML over other 3D city models formats lies in its flexible discrete level of detail (LOD) which is not only confined to geometric but also semantic characteristics (Chalal and Balbo, 2014). However, since the granularity of geometric as well as semantic information varies from one LOD to another, the chosen level detail(s) influence the energy prediction accuracy. For instance, a building with a gabled roof in LOD1 is represented as a block resulted from the extrusion of its footprint with a flat roof surface, whereas in LOD2, it processes a rough gabbled roof structure, which certainly has an impact on the building heated volume calculation. For more information on the CityGML levels of detail(s), please refer to appendix A.1.3. The wide adoption of BIM 3D models and 3GDIS systems in the architecture, engineering, and construction industry over last few years has led a lot of researchers to investigate the interoperability between IFC (industry foundation class) and CityGML (Hijazi et al., 2009; Mignard and Nicolle, 2014). This will be subsequently discussed in more detail in the 2.5.2.1.

#### 2.4.2.1 Different approaches to interoperability from IFC to CityGML

A lot of ink has been shed on developing approaches to facilitate interoperability between IFC and CityGML (Amirebrahimi, Rajabifard et al. 2015).

Table 2-9.Principles, strengths, and weaknesses of existing interoperability approaches between IFC and CityGML

| Approach | Approach Principle | Relevant Example | Strengths | Weaknesses |
|---|---|---|---|---|
| **Approach 1** | - Unidirectional conversion from IFC to CityGML | (Nagel and Kolbe, 2007)<br>(Isikdag and Zlatanova, 2009)<br>(Nagel et al., 2009) | -Widely Supported by commercial software packages.<br>-Benefit urban planning applications. | -Loss of information.<br>-Conversion to lower LOD (LOD1, LOD2).<br>-Focused on geometric transformation issues.<br>-Extensive post-processing is needed. |
| **Approach 2** | unifying approaches (Bi-directional conversion) | (El-Mekawy et al., 2011)<br>(Xu et al., 2014) | -Intends to bridge the interoperability gap between IFC and CityGML. | -Remains fairly at the conceptual stages. |
| **Approach 3** | Extending the CityGML model by either; | | | |
| | a- Introduction of generic city objects. | (Gröger et al., 2012) | -Ability to model and exchange new features in CityGML. | - Nomenclature issues.<br>-The validation of generic objects and attributes represent an issue for XML parsers. |
| | b- Using Application extension domain (ADE) | (Van Berlo, 2009) | -Defined by an additional XML schema to avoid nomenclature conflicts.<br>-Widely adopted for urban energy planning applications. | -Massive file size<br>-The combination of different ADE modules is not possible |

Table 2-9 (above) summarises the principles, advantages, and limitations of existing interoperability approaches between IFC and CityGML. Overall, there exist three distinct approaches to interoperability between CityGML and IFC namely; unidirectional conversion, unifying approaches (Bi-directional conversion) and approaches based on extending the CityGML models.

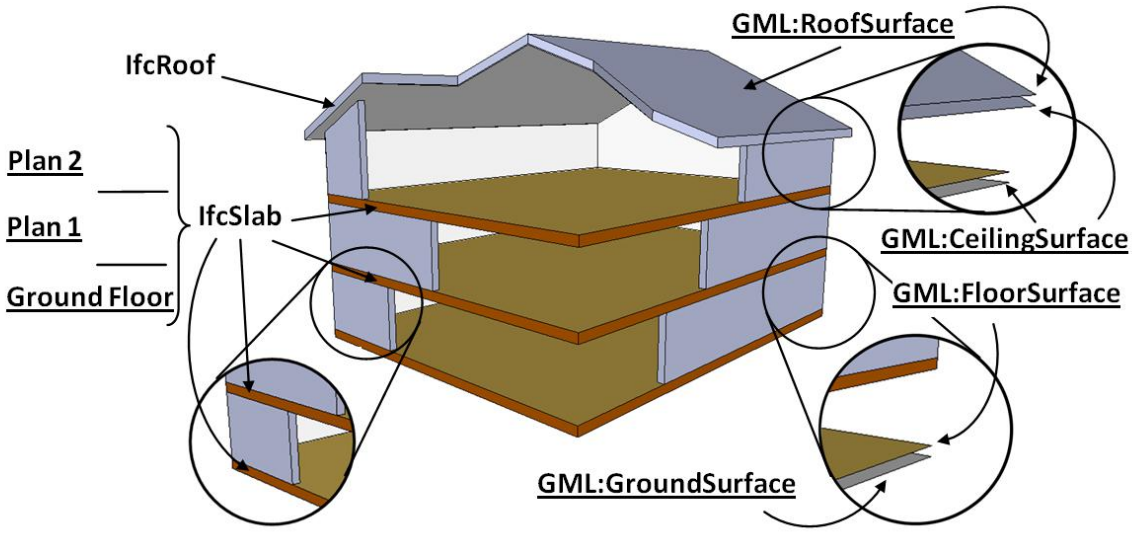



Figure 2-24.Roof and slabs in IFC and CityGML

The first one consists of exchanging information between both standards in a unidirectional fashion, meaning from IFC to CityGML or vice-versa. Unidirectional conversion from IFC to CityGML is the most addressed in the literature and widely supported by some commercial software packages such as FME (feature manipulation software) or GeoKettle. However, none of them can generate 3D building models which comply geometrically and semantically with CityGML standards Table 2-10 (Donkers et al., 2015). This is owing to the structural differences between IFC and CityGML which cause a lot of loss of information during the conversion process. Another drawback of this approach is that the geometric transformation to a higher CityGML level of detail (mainly LOD3, LOD4) is still challenging and require extensive post-processing as well as users' judgement (Delgado et al., 2013; Boyes et al., 2015). For example, in Figure 2-24, which depicts the conversion of a roof and slabs from IFC to CityGML, it is clear that a slab in IFC is represented as a solid object and defined by an IFC class (*ifcSlab*). Conversely, in CityGML, it is composed of two surfaces classified under *CeillingSurface* and *FloorSurface* or *GroundSurface*. The upper surface is considered as floor surface for the upper building level, whereas the lower one, represents a ceiling surface or ground surface for the lower building level. As for CityGML Roof, the upper surface is a roof surface, whereas the lower one is a ceiling surface for the below storey (El-Mekawy et al., 2011). For more information on unidirectional convention, please refer to appendix A.1.4.

On the other hand, some researchers like El-Mekawy et al. (2011) and Xu et al. (2014) tried to bridge the interoperability gap in unidirectional conversion approaches by integrating BIM with CityGML. This method is often referred to as unifying information model or Bi-directional conversion in the literature. However, it is still fairly at the conceptual stages and has not been implemented yet (Amirebrahimi et al., 2015). For more information on this approach, please refer to the work of Nagel et al. (2009)

The third approach, which relies on extending CityGML model, can be performed in two distinct ways namely; 1) generic objects and attributes 2) application domain extension (ADE) mechanism. The first method consists of using generic objects to model and exchange features which are not predefined by existing CityGML thematic classes (e.g. walls). However, this approach is limited due to the objects nomenclature conflicts it can create between CityGML users. Furthermore, the difficulty of validating the occurrences and layout of generic objects/ attributes by XML parsers (El-Mekawy et al., 2011). Conversely, the second method relies on application domain extension (ADE) in either introducing new properties to existing CityGML classes (e.g. household size) or defining new object types.

According to the energy ADE developed by Krüger and Kolbe (2012) for instance, this integration is performed in two distinct ways. First, primary indicators variables such as building age, number of accommodation units, and building usage, are directly retrieved from buildings' semantic properties. Secondly, complex indicators values (e.g. heated volume) are defined using complex functions. Further indicators such as assignable area, can also be obtained from elementary indicators and other complex indicators. The advantage of ADE over generic objects is that the former overcome the nomenclature limitation of generic objects as it is defined within an additional XML schema definition file with its namespace (Gröger et al., 2012). However, despite this benefit, embedding information from IFC through ADE CityGML extension can result in large file formats as claimed by El-Mekawy et al. (2011). Nevertheless, ADE is the first-choice strategies for certain applications such as urban energy planning.

Table 2-10.Evaluation of existing IFC to CityGML LOD3 converters

| | FME | BIMserver | IFCExplorer |
|---|---|---|---|
| Correct Explicit IFC Geometry | ✗ | ✓ | ✓ |
| Transformation of Geometry | ✗ | ✗ | ✗ |
| Correct Semantics | ✗ | ✗ | ✗ |
| ISO Validity of Geometries | Equal to IFC | Equal to IFC | Equal to IFC |

### 2.4.2.2 Structure of 3DCityGML urban energy forecasting models

Apart from the 3D thematic visualisation, energy ADE, and the used baseline models, most 3D CityGML urban prediction models consist of a similar structure to the 2D GIS ones. Overall, the components of these models are as follows;

- **Inputs:** comprises geometric and few semantic information which mainly describe the physical as well as thermal characteristics of the buildings' envelope. *Year of construction, usage, building height, volume, number of floors, assignable area, U-values,* are examples of common inputs in these models.

- **Assumptions:** Due to the difficulty of obtaining thermal parameters at the urban scale, the majority of encountered studies in the literature apply some assumptions to internal gains, operating schedule, U-values, set temperatures, and glazing ratios (Saran et al., 2015). For example, Nouvel et al. (2013) utilised some existing German regional libraries of typologies to retrieve standard U-values based on the dwelling size and ages.

- **Energy prediction models:** are the algorithms used in the calculation of energy consumption of homes (mainly for space heating). Besides recent studies that utilised building energy simulation tools such as IESVE, researchers in this field tend to employ steady state energy balance prediction models frequently. Examples of such models include; the quasi-state model (ISO/FDIS 13790) used by (Corrado et al., 2007; Nouvel et al., 2013) or DIN V 18599 adopted by Strzalka et al. (2011).

- **Calculation engine:** unlike 2DGIS models and very recent studies in this area, Excel-based simulation tools incorporating steady-state energy balance algorithms such as Fraunhofer-IBP or EnerCalC, are used calculate the energy consumption of dwellings. These engines do not rely on any 3D geometry to generate results.

- **Data exchange modules:** transfer inputs/outputs between calculation engine, 3D model, and external databases.

- **3D city models:** are only confined to thematically visualising the outcomes, although they can be used to extract some parameters such as volumes.

- **Energy ADE:** as described previously, it is the module that facilitates the generation and integration of energy related inputs in the 3D CityGML model (Krüger, Kolbe 2012).

#### 2.4.2.3 Conventional prediction process of 3DCityGML approaches

In contrast to studies on 2DGIS energy models, the choice of prediction models and data collection methods have not been explicitly addressed in 3DGIS studies. This is believed to be due to the influence of two key studies that endeavoured to define a standard or universal inputs for this type of models. These are the energy atlas of Berlin study conducted by Krüger and Kolbe (2012) and the model developed by Carrión et al. (2010). For these reasons, it is thought that the prediction pipeline begins with ***acquiring data*** as demonstrated in Figure 2-25.

Dwellings heights, footprints, and type are usually obtained from GIS or digital cadastral databases, whereas utility companies provide energy consumption values. Year of construction, on the other hand, is generally obtained from historical sources (e.g. Raster maps). If 3D CityGML models covering the pilot area are not available, the next step consists of generating a semantic CityGML model following two possible strategies. First, it is possible to export different 3D formats to CityGML format. However, due to the lack or loss of semantic information after the format conversion process occurred, the exported model is subject to a pre-processing phase embedding all the collected data and fixing issues related to geometry. For example, Carrión et al. (2010) employed a well-known 3D files converter namely; Feature Manipulation Engine (FME), and further pre-processing operations to convert an Oracle file format into CityGML. The second strategy is to create a CityGML semantic model from scratch using one or a combination of different modelling techniques such as laser scanning and LIDAR, depending on the envisaged LOD and prediction's accuracy (Nouvel et al., 2013). This can be time-consuming and challenging for models with LOD2 or above.

After creating a CityGML model, with the help of ADE energy, further inputs such as *assignable area*, are first obtained from the geometry of the building and existing inputs then integrated into the model. Once these parameters are embedded, dwellings with similar age and size are grouped into archetypes. After assumptions covering thermal parameters (e.g. U-values) are made based on the age and size, the calculation of energy consumption (usually heating energy) of these archetypes is performed. This is achieved with excel-base software supporting steady-state energy balance models. Afterwards, the outputs are validated against existing energy consumption values and if necessary, certain adjustments are applied to the inputs until a good match is achieved. Finally, the outputs are transferred to the 3Dcitygml model with the help of a data exchange module to be thematically visualised.

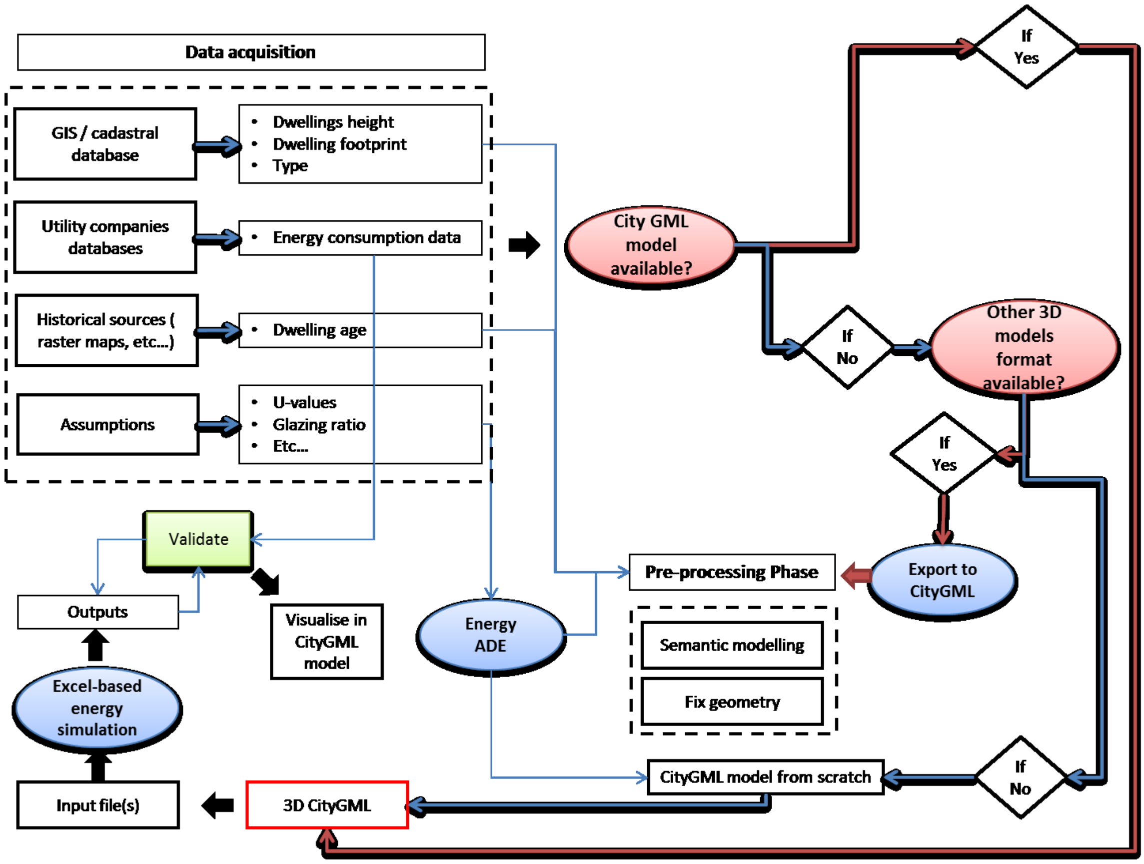



Figure 2-25.Conventional prediction process of 3DCityGML urban energy forecasting models

After analysing this generic process, it is so obvious that there is a major problem in the transition from 2DGIS to 3D CityGML models as the processes in general and visualisation patterns, in particular, are almost identical. This opens the door for the following question *"what is the point of investing more time or maybe resources in using 3D City GML for energy prediction if the 2DGIS ones can perform the same task?"*

It is believed that the answer to this question is related to the fact that the great potentials of 3D CityGML models are not fully utilised at this scale, yet. Indeed, the majority of 3D CityGML energy prediction models are modelled in LOD1 or LOD2 at most which is unsuitable for this type of application. Furthermore, as explained above the simulation processes are independent of the 3D city model, although the majority of inputs are retrieved from it. This means that many factors such as impact of shading devices, openings, and surrounding buildings, are not accounted for, which results in a non-accurate estimation. Apart from the above issues, there is non-consideration or negligence of occupancy effect on energy, which is evident in the inputs.

The following two case studies aim to support the above claims by first demonstrating the importance of 3D modelling level of detail on the outputs and secondly showing possible strategies to enhance this process, respectively.

#### 2.4.2.4 A comparative study of the suitability of different models for energy prediction

Strzalka et al. (2011) compared the energy prediction accuracy of four LOD1 3D CityGML models over two different years and following two distinct calculation algorithms as shown the above Table 2-11. Although the prediction pipeline was identical for all the compared models, the modelling techniques were different. For example, in the first model, building dimensions were acquired with a 3D laser scanner and U-values were estimated based on a regional dwelling archetype library. The second one instead, employed U-values from energy audits reports, whereas the third model utilised approximated dimensions and averaged U-values. Conversely, the fourth model, which also used estimated dimensions, took into account window orientation and measured U-values from energy audits reports. The heating energy demand of every single model was calculated twice using two different algorithms namely; a simplified transmission losses calculation method and an entire energy balance method (DIN V 18599). Overall, predictions made with the entire energy balance algorithm were the most accurate over the two years. Furthermore, the more thermal and geometric characteristics are realistic, the more accurate the prediction is, and reciprocally.

Table 2-11.Comparison between the accuracy of different CityGML modelling procedures Strzalka, Bogdahn et al. 2011)

| **Modelling procedure** | **Year** | | | |
|---|---|---|---|---|
| | 2009 | | 2010 | |
| | Deviation algorithm 1 | Deviation algorithm 2 | Deviation algorithm 1 | Deviation algorithm 2 |
| **Dimensions from laser scanning and average U-values** | -39% | -22% | -34% | -10% |
| **Dimensions from laser scanning and measured U-values** | -29% | -13% | -23% | -1% |
| **Average dimensions and average U-values** | -29% | -19% | -23% | -5% |
| **Average dimensions including real window orientation and measured U-values** | -20% | -10% | -13% | +5% |

The forth model attained the highest prediction accuracy across the two years with (-10% / +5%), whereas the third model was the least accurate with (-19% / -5%). The authors concluded that a reason behind the discrepancies between the real and calculated values of model 3 is due to the employed level of detail (LOD1). In other words, since the roof shape does not exist in LOD1, the calculation of geometry related parameters, such as, heating volume, was inaccurate for dwellings with a gabled roof which represent 80% of the area. However, it seems that thermal parameters are more influential than the geometric ones. This is evident in the comparison between model two and four in which reducing the dimensions' accuracy while keeping measured U-values in both, resulted in approximately -4.5 % deviation. This could be due to the inputs weighting of the algorithm itself.

In the light of the above, it is clear that a level of detail of at least LOD3 has to be met, and accurate modelling techniques such as 3D laser scanning have to be involved to develop precise and reliable prediction models.

#### 2.4.2.5 Applicability in the building life-cycle

CityGML based approaches have been mainly used in the operational phases of the building life-cycle. More precisely, to target building areas with retrofits potentials through diagnosing their heating demand (Nouvel et al., 2014). However, recently their applicability has been extended to cover the identification as well as assessment of renewable energy potentials across building areas (Saran et al., 2015).

For example, Carrión et al. (2010) suggested an approach to easily assess potential refurbishment in building areas. This was first achieved by deriving parameters such as *storey heights* and *building age* from the German cadastral information system as well as scanned raster maps, respectively. Other parameters such as, *building volume, building height, assignable area, the surface- to-volume ratio*, were extracted from the atlas of Berlin CityGML model either directly or indirectly using functions. Once these parameters were obtained, the heating energy consumption of the building area has been statistically calculated in excel-based simulation tool using a degree day model. Finally, the results were then compared to consumption values from existing building libraries based on the building age and typology. Similar methodology and process have been applied to areas in cities like Stuttgart, Hamburg, Ludwigsburg, and Lyon in the work of Krüger and Kolbe (2012), Nouvel et al. (2013), Bahu et al. (2013), and Strzalka et al. (2011), respectively. However, there is an exception for the work of Kaden and Kolbe (2013) who extend the capabilities of this approach to cover electricity and hot water consumption in residential areas. The estimation of electricity power demand was achieved by assigning the mean electricity values published by Vattenfall utility company to each household in the building based on their size while presuming that all of them possess similar appliances.

On the other hand, few studies have focused on exploring the potential(s) of CityGML in assisting renewable energy planning. For instance, Alam et al. (2012) proposed a method to assess the performance of PV systems through simulating roof shading analysis for direct radiation inside a CityGML 3D model. This method, which is similar to ray tracing in CAD rendering engines, comprises first triangulating all roofs' surfaces and drawing straight lines from their centroids towards the sun direction. The second step consists of checking intersections between the sun ray vector and other roof surfaces. Any triangle containing an intersection means that is shaded. After that, shaded triangles were combined to represent the overall shaded roof area.

Conversely, Saran et al. (2015) addressed the evaluation of PV systems from a completely different perspective by proposing a new approach built upon incorporating building energy simulation tools outputs into CityGML. This example is detailed in the subsequent section.

A. *Example 2: A 3DCityGML urban energy forecasting model based on building energy simulation tools*

As an attempt to bridge the interoperability gap between the building scale energy simulation and 3D CityGML urban energy prediction models, Saran et al. (2015) have recently proposed an interesting 3D CityGML energy prediction framework. It is based on the principle of incorporating the detailed outputs offered by building simulation tools such as IESVE, into CityGML to optimise the decision making regarding potential renewable energy strategies. This was accomplished by following a unidirectional process comprising six steps as illustrated in Figure 2-26 below.

- **Geometric parameters' acquisition:** to build an accurate CityGML LOD3 model, buildings' footprints and precise dimensions were obtained using IKONOS satellite maps and a total station survey, respectively.

- **Creation of 3D CAD model**: this step consisted of creating a 3D CAD model using the obtained geometric parameters and conventional CAD software which is interoperable with building energy simulation tools through gbxml format. The authors embraced Sketchup due to its 3D modelling efficiency, geolocation features, its thermal modeller plugin and high interoperability with IESVE. Each building component was assigned to a given layer complying with the structural diagram of LOD3 and facilitating the conversion to CityGML format.

- **Generation of CityGML and gbxml models**: with the help of Feature Manipulation Engine (FME) converter, the SketchUp model was exported to CityGML format. Similarly, a thermal model containing thermal zones, external/ internal walls, and windows, was created from the CAD model using gModeller-Energy Analysis Sketchup Plug-in and then exported as a gbxml file.

- **Storage of CityGML model:** to perform certain enquiries on the CityGML model, it was imported and stored in a PostGIS database.

- **Energy simulation and integration with CityGML model**: the researchers employed Suncast module in IESVE to assess the solar gain of different building parts (e.g. Roof and walls) while considering shading surfaces, orientation, and seasonal changes. Outputs, such as the percentage of the irradiated area of elements, were then integrated within the CityGML model in the PostGIS database.

- **Semantic querying and outcomes visualisation**: after the model had been enriched with more semantic information from the energy simulation tool, several enquiries could be performed. This comprised for instance, where should solar panels be installed to generate 5 MW of power per annum (Figure 2-27) With the help of query filter expression in java, CityGML displayed only the building components with matching values.

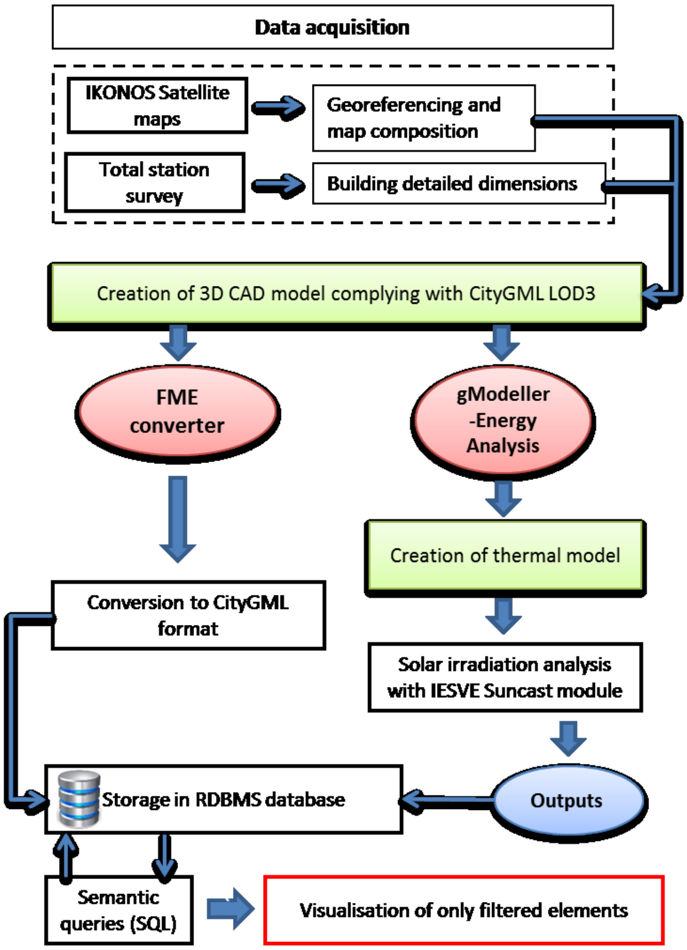



Figure 2-26.The prediction process of the 3D CityGML urban energy forecasting model adopted by (Saran, Wate et al. 2015)

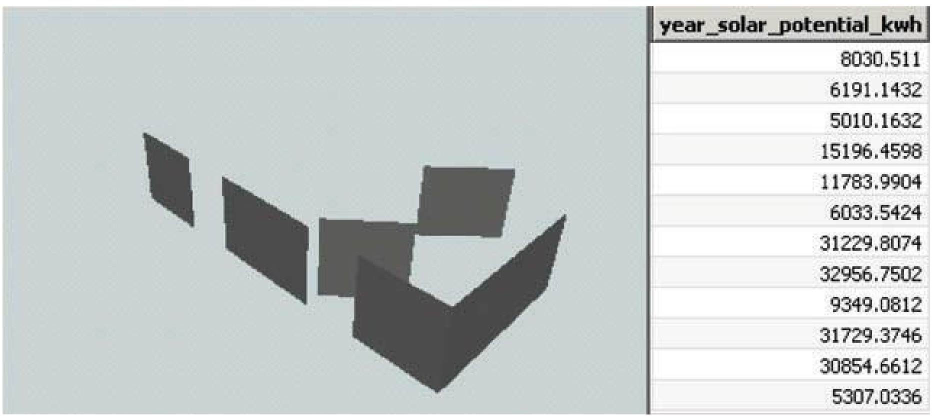



Figure 2-27. An example of a semantic query showing solar panel installation surface areas with the potential of 5MW power generation per annum

After analysing the above example, it is obvious it has demonstrated many potentials regarding the process and visualisation pattern. Indeed, it is for the first time since the adoption of CityGML in energy prediction applications, an out of the ordinary thematic visualisation pattern was introduced. Furthermore, the potential of energy simulation tools was not only utilised to provide a more comprehensive energy ADE module than the one introduced by Krüger and Kolbe (2012) but also to overcome the limitations of Excel-based simulations models. However, since tested at the building scale, it is believed that this model should be further evaluated at the wider urban context to identify as well as address specific issues related to the significant inputs/ outputs of energy simulation tools. Therefore, the undertaken research will try to build on this study by not only considering these limitations but also the potentials of conventional CAD tools (e.g. Sketchup), and advanced surveying technologies such as 3D laser scanning.

#### 2.4.2.6 Advantages of CityGML approaches

- Unlike other city 3D model formats, CityGML can model and represent objects in different level of detail(s) LOD geometrically and semantically.
- CityGML approaches have shown lot potentials for the estimation of heating energy consumption.

#### 2.4.2.7 Disadvantages of CityGML based approaches

- Their prediction accuracy relies heavily on the availability and quality of data obtained from municipalities as well as onsite measurements (Nouvel et al., 2014).
- 3D CityGML models for a particular city/region are not applicable to another since their development is mostly achieved using non-standardised data structure specified locally.

- The interoperability between CityGML LOD3/ LOD4 and other standards like IFC is still challenging and require extensive post-processing (Zhu and Mao, 2015).
- The potential of 3D CityGML models is not fully utilised for energy prediction applications. Indeed, they are mainly dedicated for visualisation purposes except providing certain geometric inputs.

### 2.4.3 Bottom up statistical methods

Since used to evaluate the impact of economic and technological factors on energy consumption, top-down statistical models would be discussed in chapter 3. Statistical bottom-up methods instead are based on correlating energy consumption/indexes with influencing factors, mostly weather parameters. This is usually achieved by applying different regression analyses, either linear or multivariate, to sufficient historical performance data (usually energy bills data), although Bayesian and Monte Carlo based approaches could also be employed (Zhao and Magoulès, 2012; Fumo and Rafe Biswas, 2015). However, historical performance data must have a high level of statistical significance to meet the accuracy requirements of energy planning (Pedersen, 2007). Indeed, this is evident in the invaluable work conducted by Cho et al.(2004). More specifically, the authors developed regression models on the basis of 1day, 1 week, 3 months' measurements that resulted in errors in the prediction of annual energy consumption of 100%, 30%, and 6%, respectively. For this reason, in addition to their ability to handle big data sets, statistical approaches are mostly employed at the urban scale. However, their applicability can be extended to cover the building scale as well, although it is uncommon.

#### 2.4.3.1 Applicability in the building life-cycle

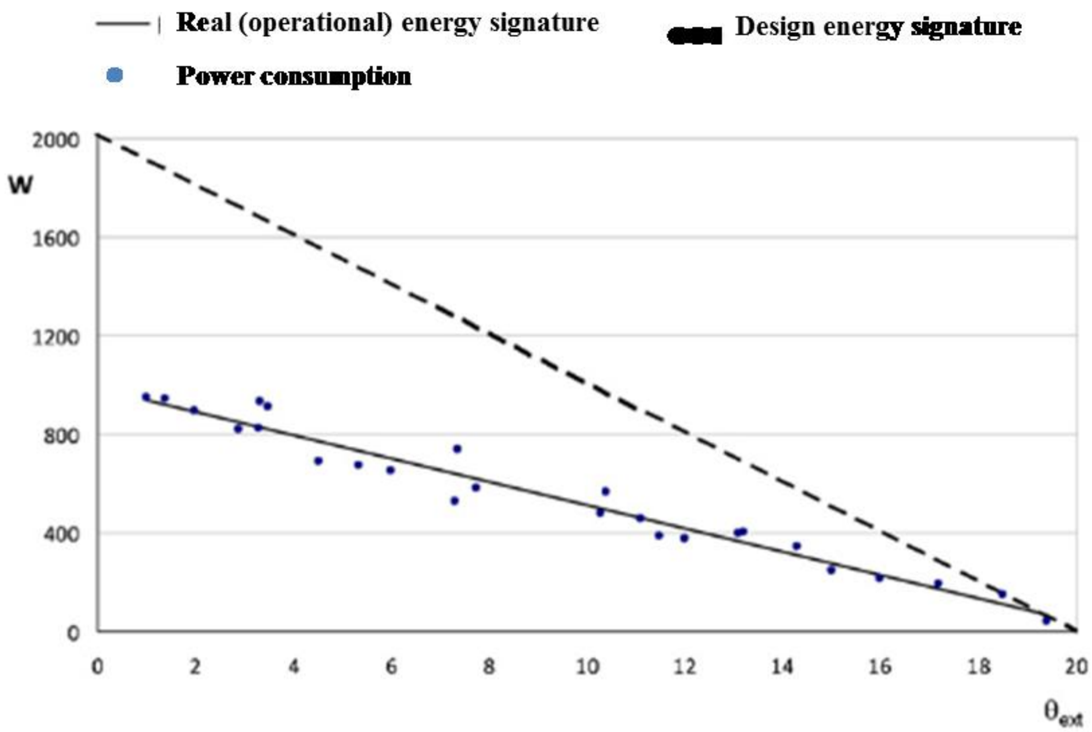



Figure 2-28.depicts a comparison between the design and real (operational) signature models by (Belussi, Danza 2012). It is clear that behaviour of building during the operational stage is different from the design one since both curves are not overlapping.

After being trained with large samples of hundreds or thousands of dwellings, statistical methods can be employed to estimate the energy consumption of entire dwellings or the thermal characteristics of their components (e.g.U-values) in function to influencing parameters (Foucquier et al., 2013). Such factors are not only prominent for the evaluation of buildings energy performance but also the assessment of energy management strategies or saving potentials during commissioning (Ghiaus, 2006). For example, energy signature, which is best fit straight line correlating energy consumption with climatic variables, was adopted by (Belussi, Danza 2012) for two main purposes. The first one is to check the conformity of certain buildings energy performance at the operational phases against the one estimated at the design phase as shown in Figure 2-28. On the other hand, the second aim consists of determining potential savings through the comparison of their energy signature models in the operational phase with the one in accordance with building regulations.

As for determining thermal characteristics, Jiménez and Heras (2005), for instance, compared the accuracy of two regression models namely; single output and a multi-output auto regression with extra inputs, in predicting the *U*- as well as *G* values of few buildings' components. The results of the latter model had a better agreement with the measured values in comparison to the former. Furthermore, the predicted values of windows in both models were more accurate than the ones of walls. Similarly, Richalet et al. (2001) proposed an assessment methodology for single family dwellings which consists of deriving the thermal characteristics through the development of building energy signatures using continuous onsite measurements. Those thermal parameters are then used to determine a normalised heating annual consumption based on standards occupant schedule and weather conditions. However, despite the fact that the authors intended to distinguish the impact of occupants' behaviour from climatic influences by evaluating occupied as well as unoccupied dwellings, the potential of on-site measurement was not utilised for this purpose. Instead, standard EU internal gains values and occupants schedules were employed for all homes. Therefore, errors in the estimation of normalised annual energy consumption were superior to 20% for some occupied dwellings. Indeed, Mejri et al. (2011) supported this claim in the following statement *"...model identification of occupied buildings is difficult because the disturbances introduced by the occupants are usually not measured."*

#### 2.4.3.2 Advantages of statistical methods

- Unlike engineering methods, the statistical ones do not depend on the 3D geometry of the building since they are not based on thermodynamic models (Foucquier et al., 2013).
- In comparison to other prediction methods, statistical models are less complex to develop (Zhao and Magoulès, 2012).

- Statistical methods have a great ability to recognise and model the variation in the households' energy behaviour (Swan and Ugursal, 2009).

#### 2.4.3.3 Limitations of statistical methods

- Since they rely primarily on historical data, their adoption is impractical in certain cases where data is unavailable or cannot be measured (e.g. lack of resources) (Pedersen, 2007).

- Statistical methods are less flexible, detailed, and accurate than other approaches due to their inability to solve non-linear problems. Therefore, their ability to evaluate the impact of physical improvements is limited (Swan and Ugursal, 2009).

## 2.5 Conclusion

Chapter 2 has extensively reviewed different energy planning and forecasting models supporting physical improvements strategies in the residential sector at both the building and urban scales. This was performed in accordance with their advantages and limitations while providing some relevant examples on their prediction pipeline as well as applicability in the building life-cycle (Table 2-12).

At the building scale, engineering models, which rely heavily on buildings' 3D models and energy simulation tools, are characterised by their flexibility, user friendliness, high integration with common CAD/BIM tools, and comprehensive outputs. However, the large number of inputs requirement and involvement of calibration procedures to generate accurate results in certain projects, represents their major drawbacks. Nevertheless, they are still first choice models for physical interventions at this scale.

Table 2-12.summary of the Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector

| Frequent Intervention scale | Approaches | Require 3D models? | Require historical data? | Easy to employ? | Prediction accuracy | Applicability in the building life-cycle |
|---|---|---|---|---|---|---|
| Building scale | Engineering | Yes | No | Easy to moderate depending on the complexity of the simulation tool. | Fair to High | Multi stages including Design, operational, and maintenance stages |
| | ANN | No | Yes | No | high | Mostly operational stages Very limited for design phases |
| | SVM | No | Yes | No | Fairly high | Mainly operational stages Very limited for design phases |
| | Hybrid | Yes, if engineering methods are involved | Yes, partially | No | Fairly high | Design and operational stages |
| Urban scale | 2D GIS | Yes | No | Moderate | Fair | Operational stages |
| | 3D CityGML | Yes | No | Moderate | Fair to High | Operational stages |
| | Bottom-up statistical | No | Yes | Yes | Fair | Operational and maintenance stages |

On the other hand, artificial intelligence energy prediction models in general, ANN and SVM particularly, do not rely on 3D models to proceed and can generate accurate estimations due to their excellent ability to handle non-linear processes. However, since trained with significant amount of historical data, they are inadequate when only small data samples are available. Furthermore, unlike engineering models, their main disadvantage is the lack of generalizability to represent the wider sample of dwellings at the urban scale. Therefore, they are not considered in the undertaken research. Conversely, Grey box (Hybrid) models, which combine the strengths and eliminate the weaknesses of the above approaches in addition to statistical ones, are well-known for their great ability to handle problems related to small samples and missing data. Thus, it could be possible to utilise this potential in case certain parameters, especially the thermal ones (e.g. U-values and glazing ratio), are unattainable.

As for the urban scale, the analysis of the 2DGIS and 3D CityGML energy prediction models has unravelled the following gaps and issues. First, the potential of 3D is not fully utilised in CityGML energy prediction models. This is evident in the similarities between the prediction pipelines of 2DGIS and CityGML models. Furthermore, in the adoption LOD1/LOD2 3D in the majority of studies, although the LOD3 is the minimum level of detail for energy prediction applications as discussed at an early stage. Not to mention the independence of 3D models from energy calculation engines. In fact, both GIS and 3D CityGML models are only employed as visualisation platforms. Secondly, there is a need to develop a new visualisation pattern matching and reflecting the capabilities of 3D models instead of relying on existing GIS thematic mapping. Finally, to the best of our knowledge, there have been no studies that have attempted to utilise game engines for the support of energy prediction application at the urban scale. Therefore, this research will try to fill this gap by exploring their potential to enhance existing urban energy prediction models.

Overall, all these models regardless of their intervention scale, nature, structure, prediction process, and ability to discern socio-economic factors or not, share two common issues. The first one is that the impact of occupancy on energy consumption is **underestimated** since the majority of parameters are about geometric and thermal characteristics of dwellings. Secondly, the variations in the households' energy related behaviour and their socioeconomic circumstances **are not accounted for** which is evident in their reliability on standard occupancy profiles. This certainly supports the statement made in chapter 1 about the non-consideration of socio-economic aspects in the majority of studies on improving the physical aspect of dwellings.

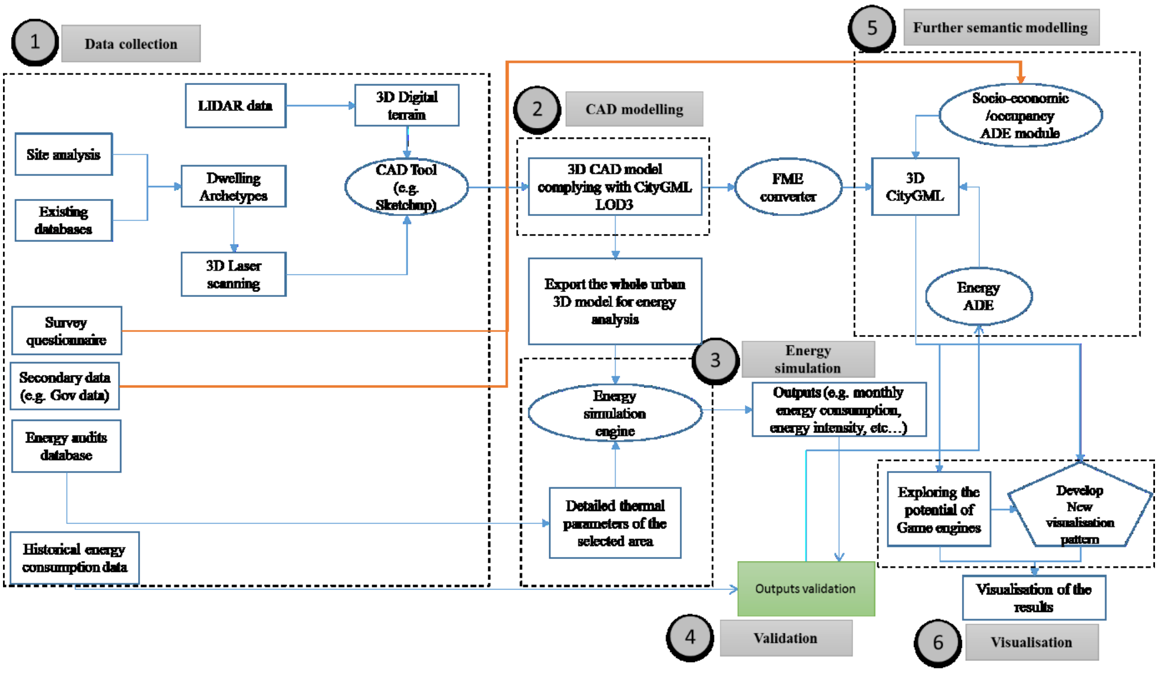



Figure 2-29.The modelling framework of the proposed urban energy planning and forecasting model

Apart from the above problems and knowledge gaps, the analysis of this chapter has also enabled the elicitation of the main physical modelling principles of the developed 3D Urban energy prediction tool, as illustrated in Figure 2-29, although this would be discussed in depth in the following chapters. The prediction pipeline of the envisaged 3D urban energy prediction tool is composed of 6 major steps namely; Data collection, 3D CAD modelling, Energy analysis, CityGML semantic modelling, validation, and visualisation of the outcomes . First, data collection methods such as Lidar data, 3D laser scanning, site analysis, and the analysis of existing data bases, will be employed to obtain an accurate 3D CAD model in the subsequent phase. On the other hand, the analysis of secondary data sources is utilised to build socio-economic /occupancy energy module in step 4. Step 2 involves using a CAD modelling software tool such as Sketchup to generate a 3D model while complying with CityGML LOD3. Afterwards, the urban 3D CAD model will be converted into an urban thermal model and exported as a gbxml file format for energy analysis. The next step involves the validation of the results with historical energy consumption data. This will be followed by, exporting the CAD model as CityGML file for further semantic energy modelling based on two distinct ADE modules; Energy and socioeconomic/occupancy. Once achieved, the outcomes will be visualised after a new visualisation patterns using the potential of game engines has been developed.

# 3

# Energy planning approaches for supporting techno-socio-economic measures

*The third chapter aims to critically evaluate the energy planning and forecasting approaches supporting the development of techno-socio-economic measures in the residential sector. This examination will be based on the frequent intervention scale, the building life-cycle relevance, strength and weaknesses of each method while pointing out to key studies from the literature.*

**Contents**

## 3.1 Introduction

From analysing the energy planning and forecasting approaches supporting physical improvement strategies in the residential sector, it is clear that physical as well as thermal characteristics such as size, orientation, and insulation thickness, have the largest impact on households' energy usage patterns. However, research has reported considerable variations in the energy consumption of dwellings with similar thermal and physical characteristics (Vringer et al., 2007). This is simply due to the influence of other factors, techno-socioeconomic and psychological, more precisely. Indeed, the literature suggests that the contribution of such aspects to the reduction of households' energy consumption accounts for 4%-30% (de Groot et al., 2006); (Guerra Santin et al., 2009); (Longhi, 2014); (Sardianou, 2005). Considering that the majority of the building sector in developed countries will benefit from physical retrofit measures before 2020, urban energy planners, as well as policy makers, are certainly exploring other alternative strategies to reduce the greenhouse gas emissions. From this premise, developing new measures to tackle techno-socioeconomic energy related issues and the households' energy behaviour is of high prominence (Frederiks et al., 2015). This argument was supported by Brounen et al. (2012) who believe that "*...However, much of the current debate regarding energy efficiency in the housing market focuses on the physical and technical determinants of energy consumption, neglecting the role of the economic behaviour of resident households.*".

This chapter does not only aim to provide a comprehensive review of the energy planning and forecasting approaches supporting the development of techno-socio-economic policies but also propose a new classification which considers new emerging models. This comprises energy consumer segmentation. To maintain clarity and consistency with the previous chapter, all these methods have been analysed according to their common intervention scale, involved parameters, prediction processes, strengths, and limitations. All those eleemnts have been supported by relevant examples from the literature. It should be noted that psychological factors are beyond the scope of this research. Please refer to the extensive review of (Abrahamse and Steg, 2009);(Sorrell, 2015);(Schot et al., 2016).

## 3.2 Existing classifications and chapter structure

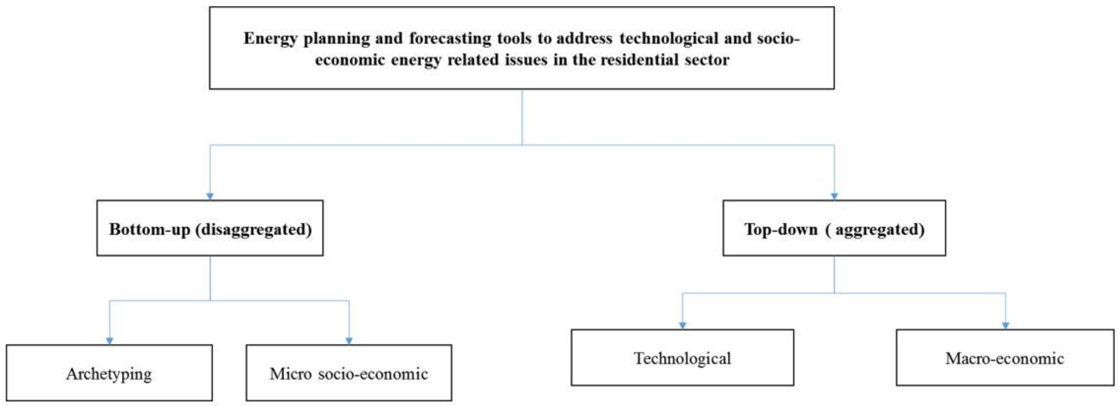



Figure 3-1.The proposed classification of energy forecasting approaches supporting techno-socio-economic measures in the residential sect

Unlike the approaches supporting physical improvement in the residential sector, whose classification is still debated among scholars as indicated in the previous chapter, there is a general agreement in the literature on the categorisation of the models supporting techno-socio-economic measures. This classification, which is hugely influenced by the international energy agency (de Sa, 1999) and the extensive review of Swan and Ugursal, (2009), acknowledges two model types namely; econometric and technological. Nevertheless, there is no a clear boundary between the two categories since their combination is common. Indeed, the extent of macro-economic or technological inputs in the model often defines its orientation. For example, if a model contains more technological parameters than econometric inputs, it is technologically driven and vice-versa.

Both categories (econometric and technological) are considered to be "Top-down" approaches since they treat the residential sector as a single consumer or as an energy sink (Böhringer and Rutherford, 2009). Thus, the employment of aggregated data, which are often historical data and widely available from different sources such as national agencies, is sufficient for long-term national or regional energy consumption/demand forecasting (Krysiak and Weigt, 2015). However, we have noticed the non-consideration of approaches which utilise cross-sectional or panel data to estimate the energy consumption of different homogeneous consumer archetypes at the disaggregated level. For example, Druckman and Jackson, (2008) extracted socioeconomic characteristics of homogeneous households (e.g. blue collar) from existing geodemographic segmentation system databases (e.g. Mosaic), to predict the average energy consumption at the neighbourhood level with their developed top-down model. Therefore, we propose to complement the invaluable effort of the above authors by reviewing existing and the suggested approaches under two main categories namely; aggregated and disaggregated as shown in Figure 3-1. First, in the aggregated models section, macro-econometric and technological models will be addressed. On the other hand, household energy segmentation approaches and micro-economic models will be analysed.

## 3.3 Aggregated model

### 3.3.1 Macro-econometric models

The prediction of residential energy consumption using a macro-econometric approach was first initiated by Houthakker (1951), although its adoption proliferated in planning since the 1970's energy crisis (Swan and Ugursal, 2009). In general, macro-econometric models are utilised to understand past relationships between energy expenditure and macro-economic predictors, including energy price, weather, gross domestic product (GDP), and population. This in turn would enable the forecasting of future energy trends of this sector (Roberts, 2008; Baker et al., 1989; Berkhout et al., 2004; Meier and Rehdanz, 2010). Subsequently, develop appropriate energy policies both at the regional and national levels or estimate the financial effect of economic shocks. Furthermore, enable international comparisons to be made. As such, these models often necessitate minor details of energy consumption processes (high aggregation level) (Min et al., 2010).

### 3.3.2 Technological models

Technological models explain energy consumption by factors related to technological changes such as appliances ownership, efficiency and their patterns of usage. The main goal of using this type of models is to help policy makers develop strategies targeting the production, distribution, and households' choice of appliances (e.g. higher-efficiency appliances). However, technological modelling still suffers from contradictive views with regards endogeneity degree of technological change. The first group of scholars argue that technological change is endogenous. This implies that its related factors interact with socio-economic parameters, mainly price (Löschel, 2002). On the other hand, the second group, who represent the majority, claim that technological change is exogenous. This means that technological improvements are non-price driven and are usually represented by autonomous energy efficiency parameters (e.g. The number of appliances) or exogenous future technological assumptions known as "Backstop" (Gillingham et al., 2008; Ermoliev et al., 2012).

The below section discusses the main drivers of aggregated energy usage through analysing the relationship between energy consumption and the main macro-economic and technological predictors in major developing countries including the UK.

### 3.3.3 Factors affecting aggregated residential energy consumption

#### A- Gross domestic product (GDP) per capita

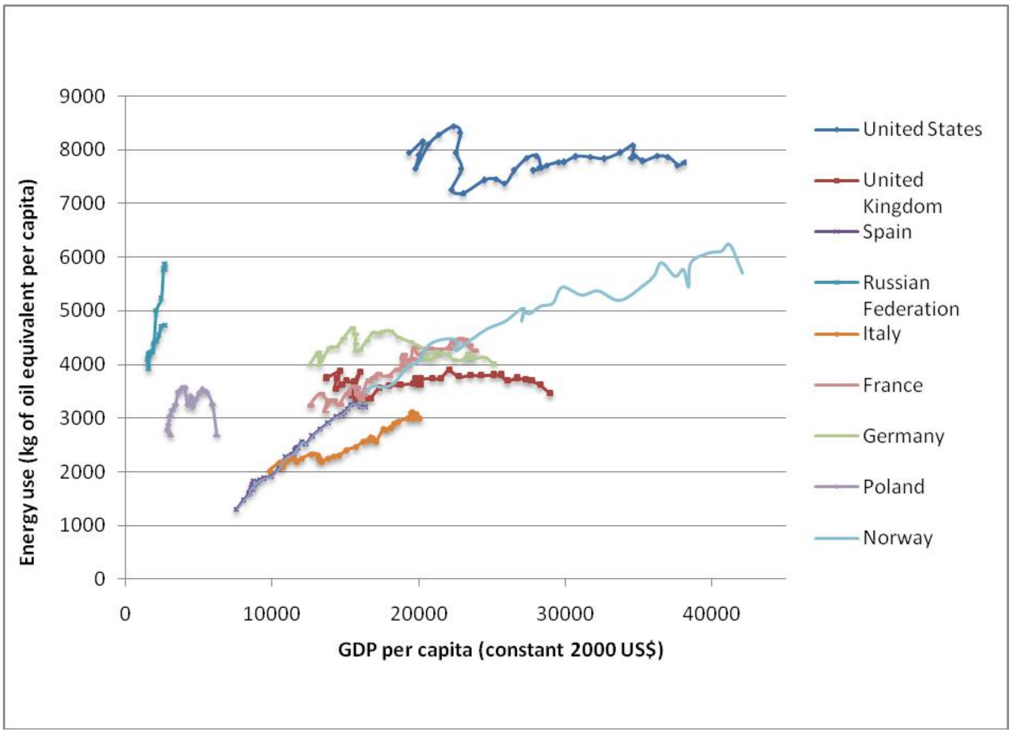



Figure 3-2.variation in energy consumption in function of GDP per capita in major developed countries between 1972 and 2008 adapted from (Reinhart and Rogoff, 2009)

Figure 3-2 is a multiple line graph which represents the variation in energy consumption in function of GDP per capita in major developed countries between 1972 and 2008. Overall, it is clear that there is a positive relationship between GDP and energy consumption in most countries. The exception applies the United States whose energy consumptions decreased over considerable periods. Conversely, the energy usage in Germany and the UK remained relatively stable over this period despite the GDP growth in both countries.

In Spain, there has been a close tie between the GDP growth and energy consumption growth with both increasing by an average of nearly 3% annually. Similarly, Italy, Norway, and France witnessed a GDP per capita growth of 1.7%, 3.5%, and 2.56 % accompanied with 0.7%, 4%, and 2.3% increase in energy usage per year, correspondingly (Platchkov and Pollitt, 2011). As for Germany and the UK, who witnessed an annual GDP growth of around 1.7% and 1.8% yearly over this period, respectively, while the energy consumption remained relatively stable. This could be attributed to a combination of offshoring and energy efficiency measures. Finally, in the case of the US, this decrease could be due to the energy price shocks that occur in the mid-1970's and 1980's. Indeed, there has been a 1% yearly increase in energy consumption accompanied with 3% increase in GDP since 1986.

In the light of the above, it is clear that gross domestic product (GDP) represents a major determinant of energy consumption.

### B- Energy price

Figure 3-3 shows the relationship between the average energy intensity (kg oil equivalent per $2000GDP) and the average energy price in some developed countries including the EU 15 countries between 1990 and 2005. In general, it is evident that countries with higher energy prices consumed less energy and vice-versa. Furthermore, across a range of countries, every 10% difference in energy price of a country in comparison to another is associated with 10% less average energy intensity. This suggests that the "price elasticity", which is a measure of countries response to the price change, is approximately -1. However, an exception to this rule applies to Russia and the former Soviet countries due to other factors which complicate this comparison and makes it non-meaningful.

For example, the USA with approximately half energy price of Japan used more than the double of energy output per $2005 GDP unit of the latter. On the other hand, France and Germany had almost half energy intensity due to their high energy prices, which are almost 1.75 times the price in the USA. However, at the national level, elasticities are typically smaller than the ones at the international scale. More precisely, elasticities at a national scale often vary between -0.2 and -0.5 (Grubb, 2014). Moreover, a 10% increase in energy price is usually associated with 3-5% energy reduction in the long-run. The persistence of higher energy prices in the long-term provokes supply chain responses since both consumers and producers will gravitate towards less energy intensive products. This will also lead innovators to introduce efficient products to the market (Kilian, 2008).

Based on the above facts, it is clear that energy price is a major macro-economic determinant of energy usage since it can change not only an economy direction but also its production in the long-term. From this perspective, in the undertaken research, all households' energy expenditure will be transformed to energy quantities to avoid any misinterpretation of the results.

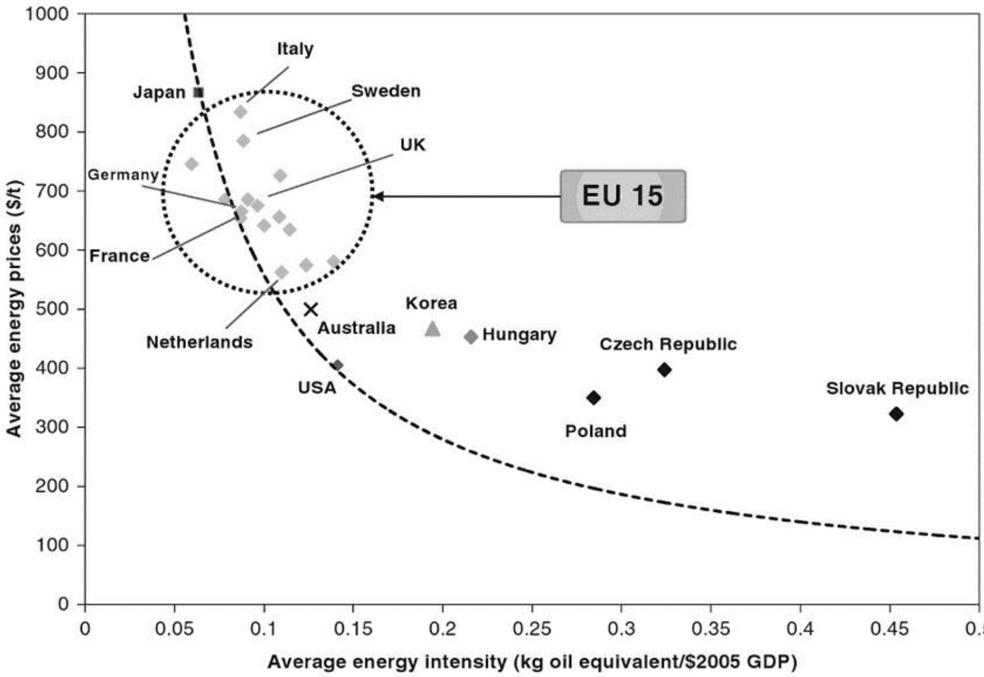



Figure 3-3. The effect of energy price on energy intensity in some developed countries between 1990 and 2005, adapted from (EEA, 2013).

## C- Population and number of household

In the literature, a positive relationship between population and residential energy consumption is usually reported (BEIS, 2013). For instance, as shown in Figure 3-4, which represents the residential energy consumption against the urban population growth in China between 1990 and 2012, the domestic energy nearly tripled as urban population doubled (EIA, 2016). However, for certain developed countries with small population growth rate, this effect is not explicit due to the implementation of energy efficiency strategies. For example, the UK domestic energy usage slightly decreased by an average of roughly 1.35% despite a 0.85% annual growth in the number of households between 2000 and 2014 according to the department of energy and climatic change report (DECC, 2015b). Similarly, the US energy information administration indicated that residential energy usage in US homes remained fairly constant since the implemented energy measures did offset the increase in the number of housing units, as depicted in Figure 3-5.

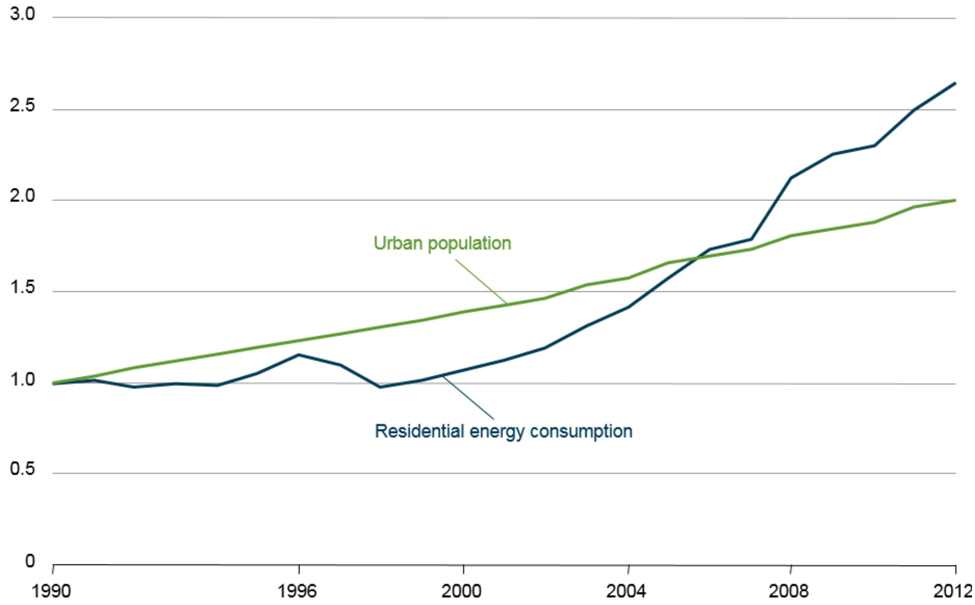



Figure 3-4. Domestic energy consumption and the growth of Chinese urban population, 1990-2012 (EIA, 2012)

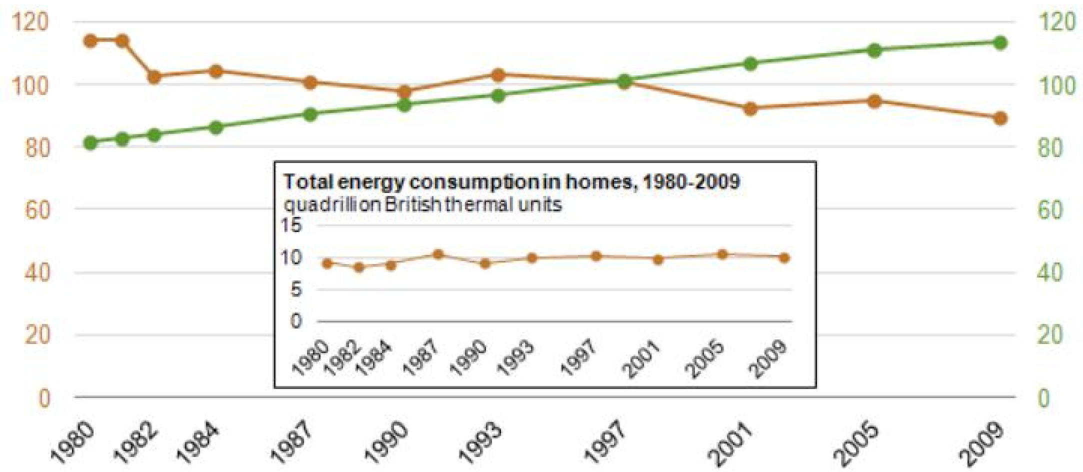



Figure 3-5.Average energy consumption per dwelling (million British thermal units) in relation to the number of housing units ( in millions) in the US between 1980 and 2009 (EIA, 2012)

## D- Appliances ownership level and efficiency

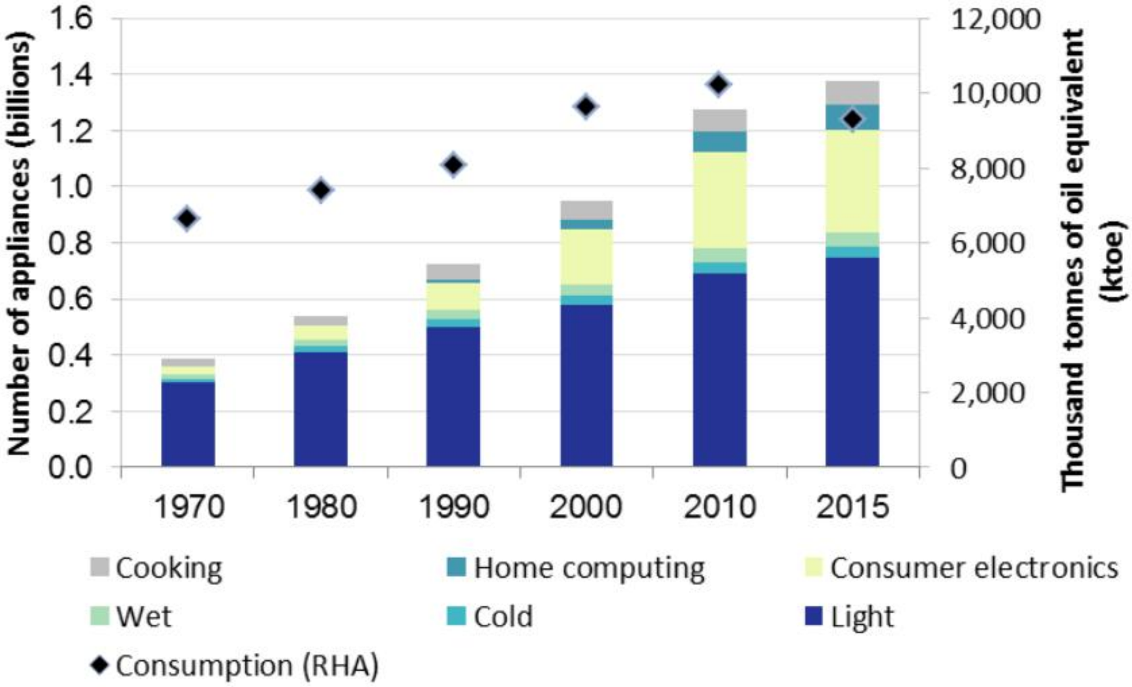



Figure 3-6.The change in the UK residential electricity consumption in relation to the increase in the total number of appliances from 1970 to 2015 (BEIS, 2016).

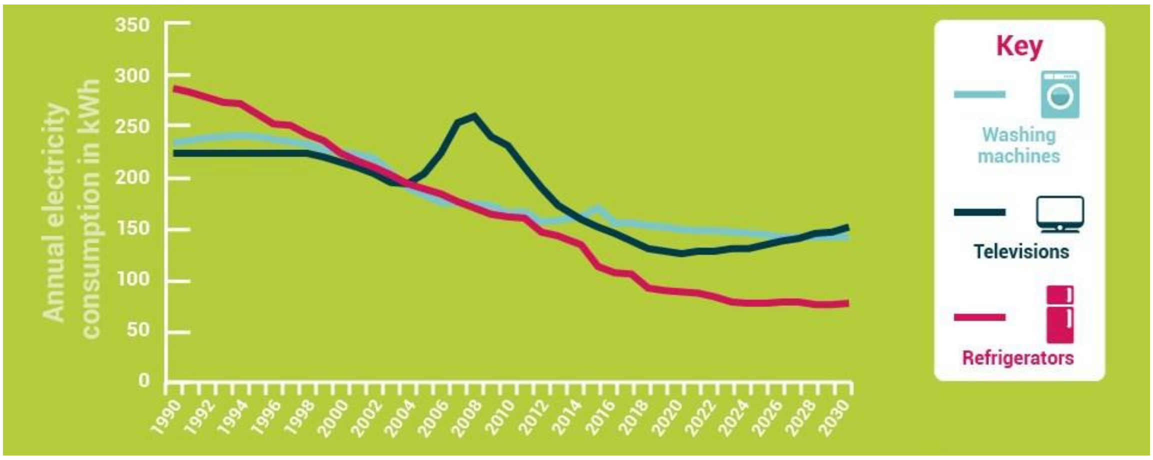



Figure 3-7. The degree of change in the annual energy consumption (KWh) of three appliances namely; washing machine, television, refrigerator, in the UK between 1990 and 2030 (DECC, 2014a).

Electrical appliances have a considerable contribution to the overall domestic energy consumption (Jones and Lomas, 2016). In particular, appliances ownership, which has been extensively addressed in the literature, is seen as a prominent factor. Several studies have acknowledged a positive relationship between the number of appliances and domestic energy use (Nielsen, 1993; Bedir et al., 2013; Wiesmann et al., 2011). For example, Nielsen, (1993) after his research on the Danish residential sector, concluded that 1 % increase in the number of appliances in Danish dwellings led to 0.35% rise in electricity usage.

Similarly, as illustrated in

Figure 3-6, there has been a rise of roughly 48% in the UK residential electricity consumption as a result of a notable 0.5 billion increase in the number of electrical appliances between 1970 and 2000. However, from 2010 to 2015, the overall electricity usage dropped by roughly 10% despite a 0.1 billion increase in the number of appliances during this period. This could be partially attributed to the improved efficiency of some appliances including light bulbs (BEIS, 2016). For example, as shown in Figure 3-7, which depicts the variation in energy consumption of certain appliances between 1990 and 2030, the average annual energy of television dropped from around 225 KWh in 1990 to approximately 150 KWh in 2016. Similarly, the annual consumption of a refrigerator and washing machine decreased by 200 kWh and 60 KWh, respectively, over the period from 1990 to 2006.

### E-Use of appliances

The frequency and duration of usage of appliances have been included in certain technological models (Saha and Stephenson, 1980; Haas and Schipper, 1998); due to their impact on the overall domestic electricity consumption. Indeed, Bedir et al. (2013) conducted a study on the determinants of the Dutch residential energy consumption and found out that the duration of using certain appliances such as HVAC and wet-appliances (e.g. washing machine), explained around 35% of the variation in electricity consumption between dwellings. Similar findings were reported by Bartiaux and Gram-Hanssen, (2005) for the Belgian and Danish residential sector.

In general, there is a positive correlation between duration and frequency of appliances use and residential energy consumption; electricity more precisely (Jones et al., 2015). For instance, Sanquist et al.(2012) found a moderate positive relationship between the use of laundry appliances and household size. However, this relationship was not found significant for certain appliances. For example, Parker, (2003) and McLoughlin et al. (2012) concluded that the frequency and usage of the main cooking appliances such as electric oven, did not have any significance on the Irish and US domestic electricity usage.

### F-Degree days

Degree days is a commonly employed index to determine the heating or cooling requirement for keeping a comfortable indoor temperature, which 15.5 $^{0}$C in the UK (Li et al., 2012). There are two types of degree days namely; Heating degree days (HDD) and Cooling degree days (CDD). HDD occurs when the average daily outdoor temperature falls below 15.5 $^{0}$C or 18$^{0}$C. On the other hand, CDD becomes applicable if the outdoor temperature rises above this standard indoor temperature. HDD is calculated by deducting the daily average temperature from standard indoor temperature, whereas CDD, is the average daily temperature minus the base indoor temperature Equation 3-1 and Equation 3-2 (Ali and Yano, 2004).

$$HDD = \sum_{nd} T^{*} - T_t \ ; 0 \qquad \text{Equation 3-1}$$

$$CDD = \sum_{nd} T_t - T^{*} \ ; 0 \qquad \text{Equation 3-2}$$

Degree days' indexes are crucial for urban energy planning since without them it is impossible to track or assess the effectivity of energy policies nor compare the energy consumption of residential sector across different regions or countries given the changing or different climatic conditions. This is because they enable the weather normalisation of energy consumption through statistically regressing past energy consumption data with HDD or CDD (Li et al., 2012).

### 3.3.4 Structure and prediction process of top-down models

The majority of top-down residential energy models utilise computer general equilibrium models (CGE) which in turn operate on the general equilibrium framework (Honkatukia, 2013). The equilibrium concept is based on the idea that supply should balance demand while taking account of different energy demand and supply characteristics. Demand is usually determined by utility function maximising of a representative household while considering a given budget constraint. On the other hand, consumption is retrieved based on the relationship between fuel type and price. Thus, the equilibrium is attained via endogenous price changes arising from utility maximising agent strategies and firms' profit maximising rationale.

In general, top-down models, including GCE, often rely on historical aggregated data which are independently or concurrently obtained from different sources (Williams, 2011). This could include gross or monthly energy values from energy providers, degree days' history from weather stations, and data from national agencies (e.g. Department of Energy and Climate Change in the UK) (Swan and Ugursal, 2009). As discussed at an earlier stage, once these data are available, statistical methods, mostly regression, are applied to predict the energy consumption or CO2 emission in function of macro-economic or/and technological parameters using one or several equations. However, there has been some recent studies such as the one of Canyurt et al. (2005), which have attempted to utilise artificial intelligence approaches, genetic algorithms more specifically. Due to their high level of aggregation, CGE models are adequate for long-term assessment of energy policies (e.g. energy taxation) and estimating the needed energy supply at national or regional scales such as energy taxation, during the operational stage of the building life-cycle. Furthermore, predicting the implications of a given economy on residential energy consumption (Bergman, 2005).

The structure of computer general equilibrium models is hierarchical where inputs and outputs are associated in a nested fashion (Krysiak and Weigt, 2015). In other words, the overall model is composed of different modules (sub-models) whose outputs feed directly into the model equation as the primary inputs. For example, Hirst et al. (1977) developed a macro-economic residential energy model which was composed of 4 modules as shown in Figure 3-8 namely; housing module, elasticity estimator module, past demographic and socio-economic condition module, and finally technological module (engineering cost).

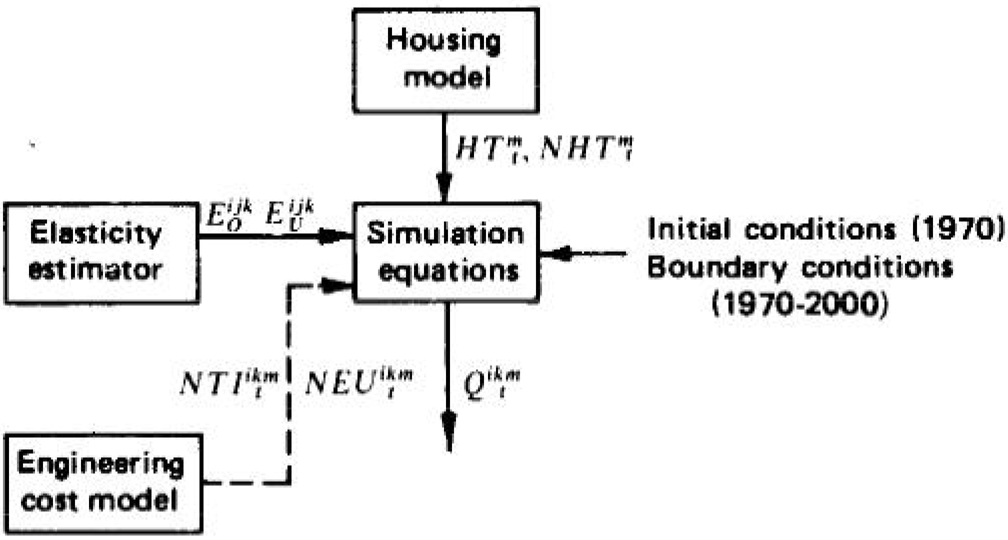



Figure 3-8.Structure of the technological model developed by Hirst et al. (1977)

### 3.3.5 Relevant examples

Bentzen and Engsted, (2001) utilised aggregate data obtained from *the Danish Energy Agency and Statistics Denmark*, to test three regression models (Equations (Equation.3-3), (Equation.3-4), (Equation.3-5) for predicting the annual energy consumption of the Danish residential sector;

$$E_{an,t} = b + c_1 E_{an.t-1} + c_2 l_{disp,t} + c_3 Pc_t \quad \text{(Equation.3-3)}$$

$$E_{an,t} = b + c_1 E_{an.t-1} + c_2 l_{disp,t} + c_3 Pc_t + c_4\, HDD_t \quad \text{(Equation.3-4)}$$

$$E_{an,t} = b + c_1 E_{an.t-1} + c_2 l_{disp,t} + c_3 Pc_t + c_4 HDD_t + c_5 Pc_{t-1} \quad \text{(Equation.3-5)}$$

Where E is the energy usage corresponding to a year t, I depicts the household disposal income, Pc is the energy price per unit, HDD is the heating degree days, b is the intercept, and c are coefficients. The researchers concluded that the long-term energy consumption in all the three models was strongly influenced by income, the energy consumption of the previous year (lagged energy consumption), and price. Furthermore, suggested that future energy prices should rise in line with income to maintain the present energy consumption figures.

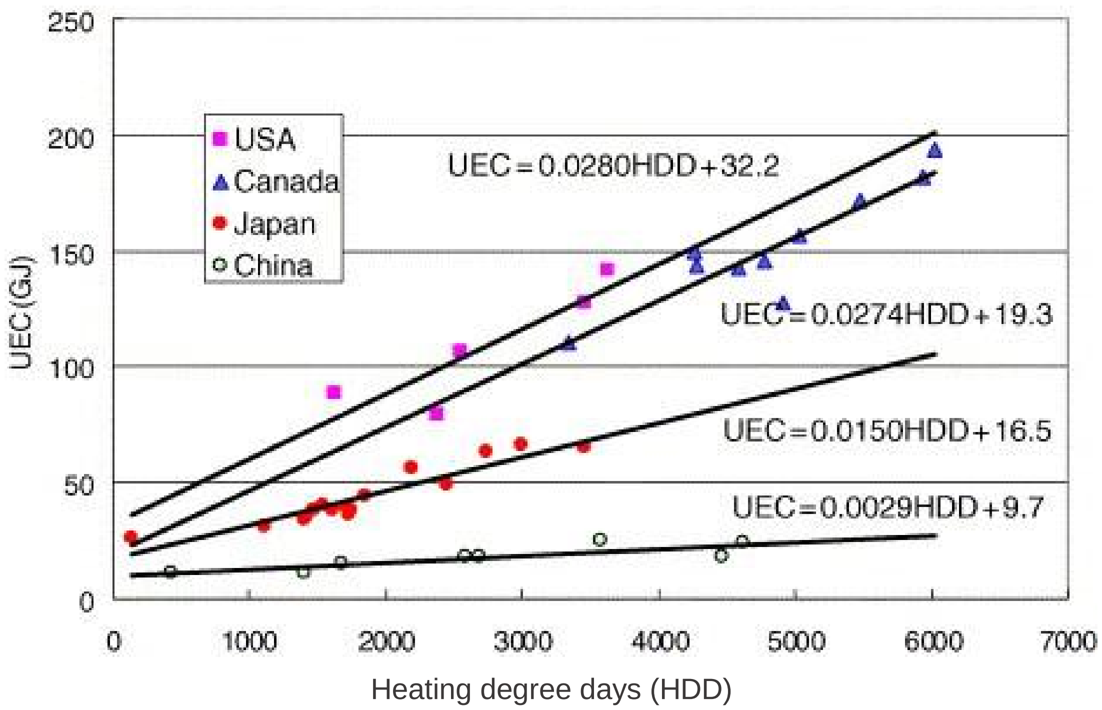



**Figure 3-9**. Comparison of annual energy consumption values of China with the ones of the USA, Canada, and Japan (Zhang, 2004)

Zhang, (2004) compared the annual energy consumption per household (UEC) values of China with the ones of the USA, Canada, and Japan (Figure 3-9). The researcher calculated the UEC values for different regions in China based on the overall residential energy consumption and the household number of each region. After that, the average UEC values of each country were normalised using heating degree days (HDD) and then compared. The findings suggested that the USA consumed roughly double the UEC of Japan and Canada, which could be attributed to the high urban density of Japan. On the other hand, China UEC represented only around ¼ of the north American UEC owing to possible low-adoption of space heating appliances.

Similarly, Leticia et al., (2012) utilised aggregated data covering 47 Spanish provinces between 2000 and 2008 to predict the domestic energy consumption of each region in relation to fuel price, income, and weather conditions. Their model, which is shown below in (Equation.3-6) , includes the following variables namely; *Eit* represents the aggregate electricity consumption while *Eit-1* is the aggregate electricity consumption pertaining to the previous year. *Yit* is the disposable income of the residential sector in euros and PEit depicts the real average price of electricity. *HSit* is household number calculated by dividing the total population POPit by the number housing units. HDD and CDD are heating degree days and cooling degree days, correspondingly. *DT* is a vector of time dummy variables, whereas εit, is the residual. Since the sample size was small, the authors compared the accuracy of three statistical regression methods namely; ordinary least squares (OLS), fixed effects, and Generalized Method of Moments (GMM). The latter regression type was found to me be the most adequate for their model since all its results were in good agreement with previous studies. These comprised a long-term price elasticity of -0.24, a disposal income elasticity of -0.3, 0.03 and 0.01 elasticities for HDD and CDD, respectively. However, despite the fact that the authors claim on the dynamic nature of their model, it is not sensitive to changes in disposable income and fuel prices. This is because the model only included the average income and electricity prices based on the year 2006, instead of the whole period.

$$\begin{aligned}\ln E_{it} = \beta p + \beta EP \ln E_{it-1} + \beta pp\ ln\ PE_{it} + \beta\gamma\ ln\ Y_{it} + \beta POP \ln POP_{it} \\ + \beta HS \ln HS_{it} + \beta GAS\ GAS_{it} + BHDD \ln HDD_{it} \\ + \beta CDD \ln CDD_{it} + BDT\ DTt + \varepsilon_{it}\end{aligned} \qquad \text{(Equation.3-6)}$$

On the other hand, Saha and Stephenson, (1980) developed a technologically oriented model to predict the energy consumption in the New Zealand residential sector between 1960-2000. This was in function of fuel type, ownership, rating, and use of major appliances namely; space heating appliances, domestic hot water equipment, and cooking, as shown in (Equation.3-7).

$$E_{an,e,f} = S.C_{e,f}.R_{e,f}.U_{e,f} \quad \text{(Equation.3-7)}$$

Where $E$ is the electricity consumption of end-users, $S$ is the number of housing units, $C$ is the appliance ownership level, R represents the rating of the appliances, and U is the use parameter. Their model had satisfactory predictions until late 1970's. This owing to the fact that their model did not account for dwelling physical improvement such as insulation levels. Young, (2008) developed a technological model, which utilises a national survey data on appliances SHEU-2003, to inspect the impact of macro-economic variables, including household income, on the retirement pattern of appliances. Their study concluded that dishwashers and freezers had the lowest and highest replacement patterns, respectively. Moreover, 40% of the Canadian households used their fridges at least for 20 years before being replaced with new ones. Although their appliances replacement patterns were different from previous studies, the authors reported that these patterns are sensitive to change in the household disposable income, average household size, and weather conditions. This suggests that the need for the development of new policies which target earlier replacement of old domestic appliances with new efficient ones.

In contrast to the above studies which encompass statistical methods, recent studies (Sözen and Arcaklioglu, 2007; Geem and Roper, 2009; Ekonomou, 2010; Kucukali and Baris, 2010); employed artificial intelligence techniques such as artificial neural network (ANN), genetic algorithm, and fuzzy logic. For example, Ozturk et al., (2004) and Canyurt et al. (2005) used a genetic algorithm (GA) to forecast the energy demand in the residential-commercial sectors in Turkey in function of macro-economic including gross domestic product (GDP), population, export, and import. Genetic algorithm, which mimics the concept of biological evolution and natural selection process, is an artificial intelligence technique for solving optimisation problems. The role of this algorithm is to modify a group of individual solutions. Thus, in each phase, it randomly picks individual solutions from the group and employs them as parents to generate children for the following generation. This Generational succession principle enables the evolution of the initial population which will in turn generate optimal solutions (MathWorks, 2016). The researchers concluded that the prediction accuracy of their genetic algorithm models were in good agreement with national energy models such as (MENR, 2016), with an average of 1% error.

Conversely, Ekonomou (2010) developed an artificial neural networks (ANN) model to forecast the Greek residential electricity consumption from 2005 to 2015 based on heating degree days (HDD), GDP, and the average electricity consumption per household. The model had good prediction accuracy when compared to recorded values with only 2% error. Based on Erdogdu (2007) argument on the limited price elasticity of the domestic electricity consumption in Turkey, Günay (2016), employed historical data from different sources such as *Turkish Electricity Transmission Company*, *Turkish Statistical Institute*, *Organization for Economic Cooperation and Development*, to attain the following objectives. First, investigate the significance of fuel price, population, unemployment rate, and inflation percentage on the gross domestic electricity consumption using multiple linear regression. Secondly, model the electricity demand of the Turkish residential sector using artificial neural networks approach by only including significant variables. The study concluded that only population, gross domestic product (GDP), inflation, and average summer temperature, were significant. Furthermore, their contribution to overall electricity demand was as follow; 44%(population), 46%(GDP), 7.9% (price inflation), and 5% (average summer temperature). Finally, the developed ANN model had a better prediction ability in estimating past electricity trends (1975-2006) and the future ones in comparison to previous studies (Akay and Atak, 2007), since its root square mean error (RMSE) was the smallest 5.7.

### 3.3.6 Advantages of Top-down approaches

- Due to the limited number of inputs involved in the model and the availability of macro-economic indicators (e.g. GDP), technological, and climate data, top-down models are easy and cost-effective to develop (Swan and Ugursal, 2009; McFarland et al., 2004).
- Top-down models allow the understanding of the implication of energy policies on the economy at the regional or national levels (Herbst et al., 2012).
- Due to their reliance on historical data, top-down approaches are adequate for analysing long-term residential energy demand.

### 3.3.7 Weaknesses of Top-down approaches

- Despite their long-term forecasting capability, top-down models falter if discontinuity is faced. This could include supply shocks or technological breakthroughs.
- By considering the gross energy use of the residential sector or the average domestic energy consumption per household in the forecasting process, they subsequently assume the homogeneity of the households. Therefore, they are not able to capture the effect of the variation in the household socio-economic characteristics, which is a major drawback.
- Although some aggregate technological factors (e.g. appliances ownership level) are included in certain top-down models, they are unable to represent technological progress. This owing to their lack of technological details on end-uses. Moreover, the indirect modelling approach of technological progress via elasticity of substitution, which is the responsiveness of the consumer of a

particular technology to the change in the price of other alternatives (Herbst et al., 2012;Karjalainen et al., 2014).

## 3.4 Disaggregated approaches

### 3.4.1 Micro socio-economic model

In contrast to macro-socioeconomic models, micro-socio-economic models are based on analysing the change in individual households' energy usage patterns in response to the variation in different socio-economic factors (e.g. income and employment mode) (Eakins, 2013). Besides their benefit of helping the development and evaluation of energy tax and welfare policies, another prominent motivation behind developing these models is to find answers to the uncertainties and contradictions about the degree of influence of different socio-economic factors on domestic energy in the literature. This argument was widely supported by Frederiks et al. (2015) who argued that "*Despite an expanding literature, we find that empirical evidence of the impact of these variables has been far from consistent and conclusive to date.*". These opposing views, will, in turn, be discussed in detail in the coming section.

In addition to the above, we have noticed that these models are highly dependent on cross-sectional and longitudinal data which are usually collected using large-scale household surveys. For example, Brounen et al., (2012) utilised a national cross-sectional data which comprises 300000 dwellings in Netherland to study socio-economic factors affecting gas and electricity consumption. On the other hand, Meier and Rehdanz (2010) used the British household panel (longitudinal) data to investigate the impact of income and household type on energy expenditure. Statistical methods such as pooled linear regression, fixed effects regression, random effects regression, and generalised method of moments (GMM), are often involved in the analysis of such type of data.

#### 3.4.1.1 The determinant of domestic energy consumption in Micro socio-economic models

Households size, household reference person age, level of education, the presence of children and teenagers, patterns of presence at home, employment mode, socio-economic status, income, household type, and tenure type, are the parameters that are often discussed in the literature. However, their significance and extent remain debated among scholars, as discussed at an earlier stage.

**Household size**

Generally speaking, the increase in household size is usually associated with a rise in the annual energy usage(Jones and Lomas, 2016; Brounen et al., 2012; Bedir et al., 2013). Jones and Lamas (2016) on their research on appliances' ownership in the UK residential sector, concluded that families with 3 or more occupants should consume more electricity than those with 1 or 2 inhabitants. Schröder (2015) claimed that doubling the household size would result in a 25% to 36% increase in their energy expenditure (Schröder et al., 2015). Similarly, Tso (2014) found that the annual household energy usage in the US is expected to rise by 488,791 kWh/year, for one unit increase in the mean household size per square root (Tso and Guan, 2014). Leahy and Lyons (2010) on their research on the Irish residential sector, suggested that one-person household consumes around 19% less electricity per week than a two-person household. Similarly, (Tiwari, 2000) acknowledged that the energy expenditure of a five-member family in India is 23% higher compared to a two-member family. Zhou and Teng (2013) concluded that an increase of one member in the household results in 8% rise in energy consumption. Conversely, Longhi (2014) found a negative relationship between per-capita energy expenditure and household size in the UK as a two person household, for instance, have 47 % less per-capita energy expenditure than a one person household.

**Age of Household reference person (HRP)**

As for the age of household reference person (HRP), the literature advises that it has a significant influence on domestic energy consumption. Furthermore, suggests that the energy consumption of HRP aged between 50-65 years is high, whereas the one of HRP aged above 65 years is low (Jones et al., 2015; BRE, 2013; Longhi, 2014). This could be attributed to the fact that mid-aged HRP have more children, whereas older HRP are more conscious about energy savings and possess fewer appliances (McLoughlin et al., 2012; Kavousian et al., 2013). Similar findings were reported by Brounen et al. (2012) and Bedir et al. (2013). However, other scholars including Abrahamse and Steg (2009) and Huebner et al. (2015) found no significance between the age of HRP and households' energy consumption.

### Level of education

The effect of education on the household energy consumption in the literature is controversial and quite unclear (Brohmann et al., 2009). However, education is usually reported to be associated with the extent of environment-friendly behaviour (McFall and Garrington, 2011).

For instance, Zhang et al. (2011) , who conducted a study on the Chinese residential sector, found out that the householder's level of education is a prominent factor that defines the change in the residents' energy usage patterns. Similarly, Zhou and Teng (2013) determined that households with an educational level higher than primary education in China had larger energy consumption figures due to their high-income levels. Conversely, Gram-Hanssen et al. (2004) claimed that the householders with primary school education consume an additional 200 KWh per annum than the ones with higher education. Similarly, Harold et al. (2015) found that householders with lower educational achievement used less gas on average. Longhi (2014) argued that households with at least one member holding a university degree consume 2% less per-capita on the overall energy consumption. Other authors Bedir et al. (2013), Mills and Schleich (2010) and Guerra Santin (2010); however, found no statistical significance between the level of education and households' energy usage patterns.

### The presence of children

The presence of children in the household has been reported to be significant by many researchers. For instance, Brounen et al. (2012) suggested that Dutch households with children consume a fifth more electricity than the ones without children. Moreover, claimed that this effect has a positive correlation with the children age since older children use certain appliances for longer hours (e.g. TV, personal computers, game consoles, etc…). Shirani et al. (2016) shared a similar view but suggested that child age has a more considerable effect than child presence. On the other hand, Bartiaux and Gram-Hanssen (2005) compared the effect of children presence aged between 0-9 years on the mean energy consumption in the Danish and Dutch residential sector. Their findings suggested a negative and positive impact on the Danish and Dutch domestic electricity usage, correspondingly.

Bedir et al. (2013), Leahy and Lyons (2010), and Longhi (2014); on the other hand, reported no significant difference between households with and without children.

### The Socio-economic status of household reference person

Socio-economic status is usually defined by a combination of occupation, education, and income (APA, 2017). However, despite this fact, its effect on domestic energy consumption remains ambiguous in the literature. For example, McLoughlin et al. (2012) found a significant effect of households' socio-economic status on the Irish residential energy consumption with higher-grade professional consuming more than the lower-grade ones. The authors referred this difference to the fact that higher grade professionals often possess larger dwellings and appliances which also suggests a possible income effect. In contrast, Leahy and Lyons (2010) found socio-economic status to be insignificant.

### Tenure type

In general, the literature advises that home owners consume more energy than the renters (Harold et al., 2015; Meier and Rehdanz, 2010). Indeed, Ndiaye and Gabriel (2011) reported that rented dwellings had the highest energy consumption due to a possible inclusion of utility bills in the rent. On the other hand, Yohanis et al. (2008) suggested that households who own their dwellings consume 25% more electricity than the ones living in rented accommodations. They argued that this effect was because a large proportion of rented accommodation in Northern Ireland was occupied by low-income tenants. Similarly, Wyatt (2013) compared the mean annual electricity consumption of British households under different tenure modes. The study reported that the lowest electricity annual mean was observed in housing associations and council homes around 3700 KWh, whereas the highest (roughly 4600 Kwh) was seen in owned dwellings.

### Income

Income has been widely reported to have a positive influence on households energy consumption by various scholars (Druckman and Jackson, 2008); (Genjo et al., 2005; Bartiaux and Gram-Hanssen, 2005; Wiesmann et al., 2011; Guerra Santin et al., 2009; Zhou and Teng, 2013; Santamouris et al., 2007; Jimenez and Yépez-García, 2016). For instance, Longhi (2014) found that 1000 £ increase in the UK monthly household income is accompanied with 1.5%-2% increase in expenditure on overall energy consumption. Chun-sheng et al. (2012) found a moderately positive relationship between income levels and total energy consumption both in urban and rural areas in China. Wyatt (2013) advised that the annual electricity consumption of households earning around £75,000 a year was almost double of the ones on low-income (£10,000). Similarly, Burholt and Windle (2006) conducted a study on low-income elderly households in Wales and concluded that the proportion of income spent on energy was higher for those earning less than 140£ per week. However, some authors including ; Tso and Yau (2007), Kavousian et al. (2013) and Carter et al. (2012), found no significant relationship between household income and domestic energy consumption.

#### 3.4.1.2 Relevant studies

Dresner and Ekins (2006) utilised the following data namely; the English survey data of 1996 as well as the family expenditure survey of 1999 and 2000, to investigate if certain compensation mechanisms (e.g. Tax and benefit system) would reduce the CO2 emission in the residential sector without harming low-income households. In other words, they explored the possibility of utilising household carbon tax to compensate low-income households, especially single-elderly who might pay more tax due to their poorly insulated dwellings and high energy bills, as an attempt to tackle fuel poverty. The authors incorporated the above data into POLIMOD, a micro-simulation model developed by Cambridge University, to evaluate the impact of actual or future tax and benefit reforms. The study concluded that the impact of income on energy expenditure and usage is negligible. Furthermore, the energy usage patterns of low-income households variated widely. Indeed, the analysis of the national energy efficiency data framework report (DECC, 2015c) suggested that households earning less than £10,000 annually consume between 2500 to KWh and 20000 KWh of gas annually. Therefore, it is not practical to enable all low-income families to get access to the benefits system to be compensated on any carbon tax that they might have paid previously.

Similarly, based on the principle that a successful CO2 emission policy should not negatively affect marginalised households, Jamasb and Meier (2011) employed a household panel data covering 17 years from 1991 to analyse the energy expenditure of low-income groups (e.g. Pensioners and lone parents). In addition to the probability of poor dwelling insulation as well as the existence of inefficient appliances, the authors claimed that these vulnerable households are the most affected by increasing energy prices. Indeed, this argument is supported, since the majority of these family types are on prepaid tariffs which are 5-10% more expensive than direct debit payment plans. Moreover, allocate a large proportion of their income to energy bills. From this perspective, Jamasb and Meier (2011) suggested that decision makers should take into account the equity aspect of envisaged increasing energy prices when developing policies to narrow the inequality as well as the "energy gap" amongst different households. This could, for instance, comprise a financial support and social tariffs.

In contrast to the above examples, Fawcett (2016) argued that the most effective way to empower vulnerable households is to re-orient the UK residential energy policy from efficiency improvement towards tackling excessive energy consumers. This is due to the fact that policies on energy efficiency cannot ensure a consistent reduction in energy consumption in the future, as claimed by the author.

On the other hand, Meier and Rehdanz (2010) utilised longitudinal (panel) data, which encompasses around 5000 households interviewed annually between 1991 and 2005, to identify the type of households who are more likely to be affected by fuel price increase and change in income. The researchers employed a random effect regression model (Equation.3-8) which predicts heating energy expenditure in function of fuel type (Fi,t), building characteristics (Bi,t), socio-economic characteristics (Si,t ), regions (Ri,t), weather conditions(Wi,t), and time (Ti,t).

$$E_{i,t} = \alpha + \beta_F F_{i,t} + \beta_B B_{i,t} + \beta_S S_{i,t} + \beta_R R_{i,t} + \beta_w W_{i,t} + \beta_T T_{i,t} + V_i + \varepsilon_{i,t,} \qquad \text{(Equation.3-8)}$$

It was found that owners reacted differently to the variation in income and energy price than renters. More specifically, renters' income and price elasticities were in general lower than those of owners. However, there was an exception for flats and terraced houses where the elasticities of owners and renters were very similar. Based on this argument, the authors suggested that this difference might be due to the difference in the dwelling type since owners tend to usually dwell in detached or semi-detached houses which are less efficient than flats. Recently, Jones and Lomas (2015) carried out a survey covering 315 dwellings in the city of Leicester to explore the socio-economic factors influencing the high electricity demand of particular consumers. The researchers concluded that household size was significant with households of 3 or more are likely to be high consumers. Moreover, households with at least one child are twice more likely to be high electricity consumers than households without children.

Similarly, a significant positive relationship was found between the number of teenagers in the household and high electricity usage. The age of household reference person (HRP) was also found significant as the ones aged above 65 years were unlikely to be high consumers, whereas the ones aged between 36 and 50 years have a high chance of using a lot of electricity. This might be because households aged above 65 often use gas as a primary heating fuel instead of electricity. However, despite the significance of these factors in addition to income, employment status and level of education were not found relevant.

### 3.4.2 Household Archetyping (segmentation) approaches

In contrast to macro-economic models, Archetyping approaches, like micro socio-economic, acknowledge the variation in energy consumption across households. Furthermore, they recognise that the implication of energy policies will vary from one consumer to another. However, the Archetyping approach has taken it one step further by trying to build segments of homogeneous consumers instead of focusing on households in isolation. This, in turn, will provide expanded opportunities to better target energy policies that address the need of different sub-groups (DECC, 2013). Furthermore, identify segments with excellent opportunities for the deployment of certain policies. Studies under this approach, employ different clustering analysis techniques, which are often statistical or artificial intelligence methods, to group households based on socio-economic, demographic, or even behavioural characteristics (Guerra Santin, 2011). However, there have been few studies, such as the one of Druckman and Jackson (2008), where geodemographic segmentation principles have been applied.

The Geodemographic segmentation umbrella encompasses a wide range of methods which classify neighbourhoods based on the assumption that households living close to each other should have similar socio-economic and demographic characteristics (Troy, 2008). One major disadvantage of the Archetyping approach is that it requires comprehensive and high-resolution data about representative households to generate reliable archetypes at the national scale which is costly and time-consuming. This was widely supported by Zhang et al. (2012) who argued that: "*...each archetypes referring to more specific and effective energy interventions, although this requires high-resolution data about local households...*"

The below sub-section will address the process of building archetypes while highlighting some examples from the literature.

#### 3.4.2.1 Process of building household archetypes and relevant examples

The conventional process (Figure 3-10) usually begins with either collecting disaggregated household data at the national scale through surveys or obtaining secondary data. Such data are often collected by governmental agencies such as the department of energy and climate change (DECC) in the UK. For example, Guerra Santin et al. (2009) utilised a survey previously conducted by OTB Delft University comprising 315 households across different districts in the Netherland. It included four main sections, building characteristics, households characteristics (e.g. income), and occupant behaviour (e.g. Ventilation frequency). On the other hand, Hughes and Moreno (2013) analysed the Household Electricity Usage Study, which comprises

250 English households interviewed between 2010 and 2011, to retrieve a set of distinctive archetypes. Once data is obtained, factor analysis, which is a statistical method used for data reduction purposes, is applied (Torres-Reyna, 2010). The aim of factor analysis is to retrieve a small group of uncorrelated variables with major effect on the dependent variable (e.g. energy usage).

This procedure is beneficial when the complex data with numerous socio-economic, demographic, and psychological variables, are involved in the Archetyping process (Blight and Coley, 2013). Hughes and Moreno (2013), for instance, embraced factor analysis to determine the most prominent factors from an overall of 29 variables related to the environmental belief and actions of the participants. They discovered that the householders' current beliefs, current actions, and future beliefs were the most influential on their energy expenditure patterns. Similarly, Guerra Santin (2011) used factor analysis to reduce the number of behaviour variables; consequently, simply the Archetyping process. The researcher concluded that only five factors were considered since their contributions to the variation in energy consumption were 15.2%, 10.3%, 8.8%, 8.1%, and 5.9%, correspondingly. After identifying the most influential factors, different clustering techniques will be applied to the database with reduced variables.

Many techniques can be employed at this stage but K-mean is the most popular due to its simplicity, cost-effectivity, and suitability for large databases. The main focus is to produce k clusters while every participant belongs to only one cluster. However, the disadvantage of this technique is that the number of clusters needs to be assumed a priori (Haben et al., 2016; Chatterton et al., 2016). The second step after defining the number of clusters involves determining one centroid for each cluster randomly. Afterwards, each data point needs to be associated to the nearest centroid. This will be followed an iterative process which consists of re-calculating new centroids and associating them with the same data points until centroids do not move anymore, as shown in Figure 3-11. Finally, clustering optimisation methods such as scree plots, or comparative assessment of square root errors of a different number of clusters, will be involved to define the optimal number of clusters (Jain, 2010). An example of studies using K-mean clustering includes the work of Hughes and Moreno (2013) where seven optimal archetypes namely; peak-time users, off-peak consumers, practical considerations, modern living, lavish lifestyle, thrifty values, profligate potentials, were defined. Another example is the study of Kubota et al. (2015) who utilised this clustering technique to group 297 Indonesian households based on two major determinant factors (identified using factor analysis) namely; income and household size. This helped the authors build a detailed energy consumption profile for each cluster.

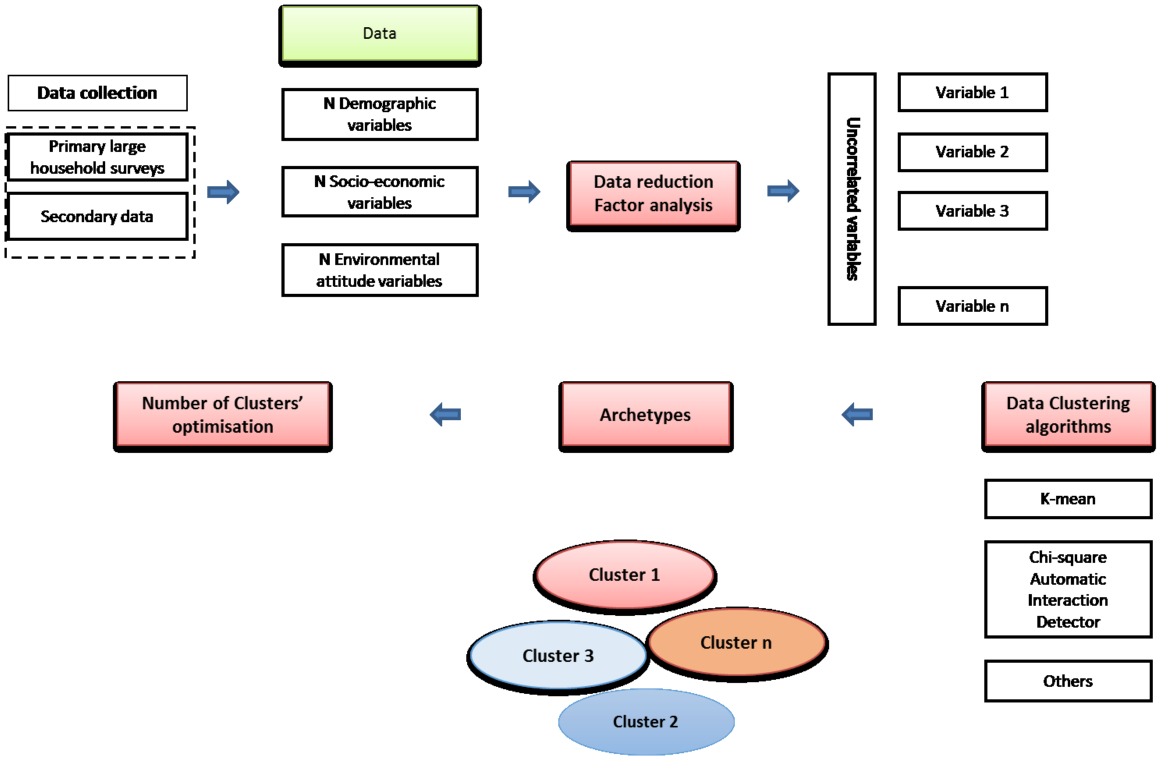



Figure 3-10. Conventional process of building households' energy consumption archetypes.

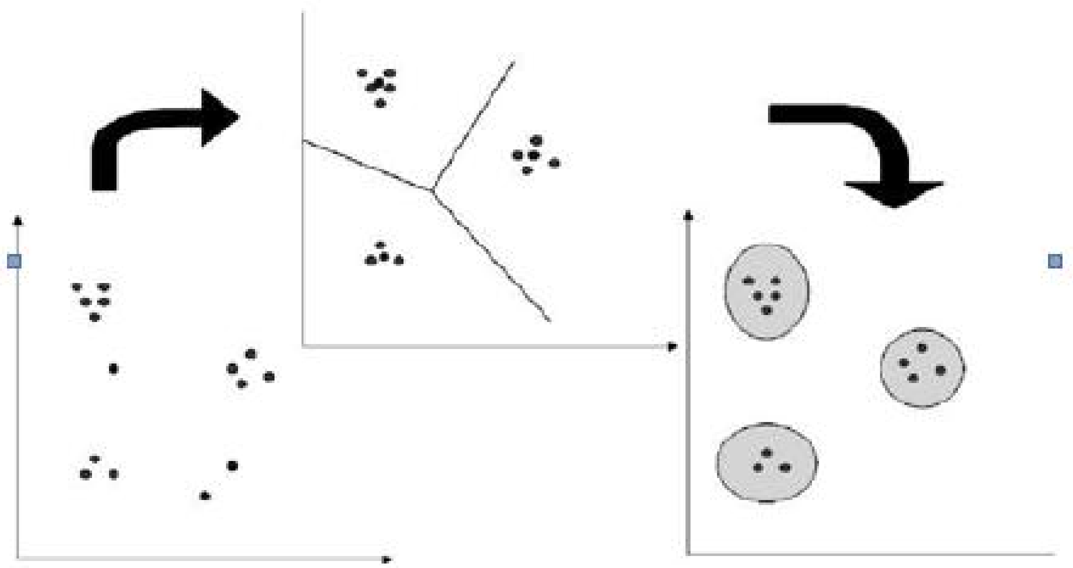


Figure 3-11.Illustration of clustering process using K-mean algorithm, adopted from (Riadi et al., 2013)

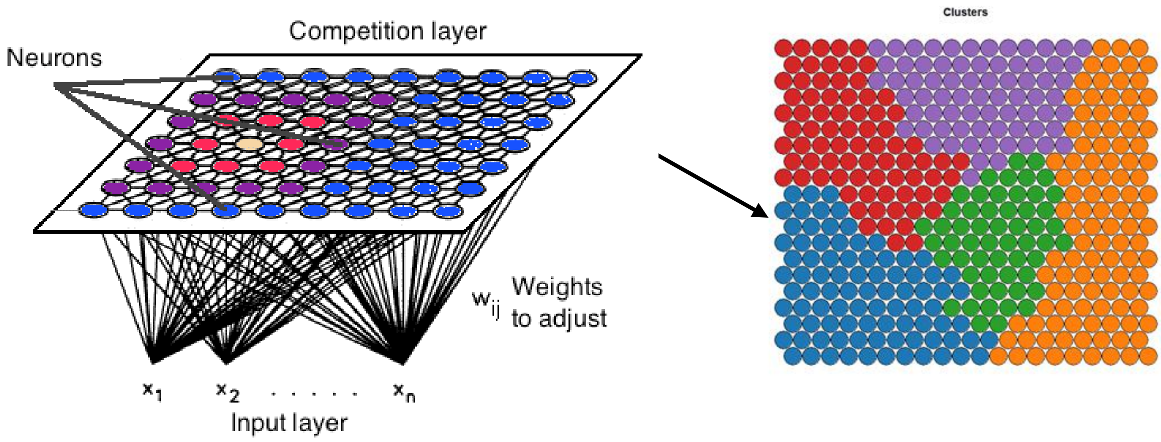



Figure 3-12.Illustration of clustering process using SOM (self-organising maps) algorithm, adopted from (IVAP3, 2015)

In contrast to the above conventional process, there has been a growing interest in using self-organising maps in recent years, especially for domestic electricity load segmentation (Martín-Merino and Román, 2006). This could be mainly attributed to the new data collection potentials offered by smart metering infrastructures. Self-organising maps or Kohonen's Self Organizing Feature Maps, as referred by others in the literature, are a type of artificial neural networks which aim to map high dimensional data into 1 or 2-dimensional space while preserving distances between data points (Figure 3-12). Therefore, they are adequate for solving clustering issues (Kohonen and Somervuo, 1998).

Unlike, *K-mean*, the number of clusters $k$ is not assumed a priori but instead computed automatically by the algorithm which in a way reduces human bias introduced in the *K-mean* initialisation stage (Villmann, 1999). From analysing the literature, most of the examples utilising self-organising maps for electricity consumer segmentation, first cluster based on electricity consumption or expenditure and explore the socio-economic characteristics of each group, afterwards. For example, McLoughlin et al. (2012) employed an extensive smart metering database of Ireland comprising around 4000 households monitored for one full day. After grouping the households into nine distinct clusters using SOM, the authors claimed that the variation in electricity consumption among certain groups is due to the difference in their socio-economic and dwelling characteristics.

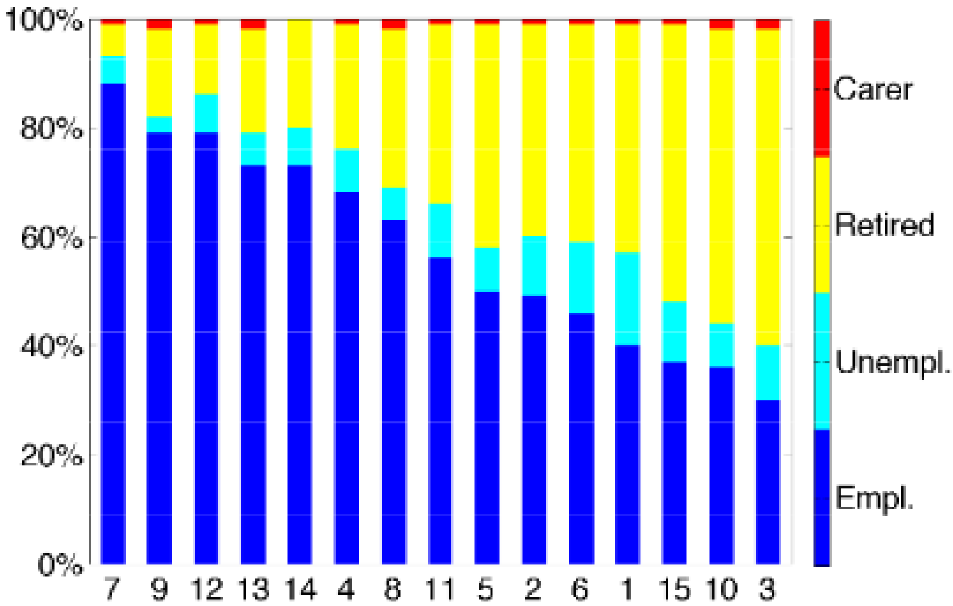



Figure 3-13.Employment mode of the household heads, adopted by Beckel et al. (2012)

Similarly, Beckel et al. (2012) applied SOM to cluster 3488 households which were monitored for 18 months using smart meters as part of the Smart Metering Electricity

Consumer Behaviour Trial of Ireland. After that, with the help of stacked bar charts (Figure 3-13), the distribution of certain socio-economic factors such as income, employment status, and number of appliances across different groups, were examined to explore their impact on electricity consumption. However, despite the potential of this approach in terms of reducing the complexity of smart metering datasets and automatically extracting energy consumer segments, studies utilising SOM are prone to some issues.

First, the significance and the extent of the examined socio-economic factors are questionable since the inspection of their impact was conducted visually. Secondly, it is believed that is not possible to classify a given household with particular socio-economic characteristics into existing groups. This is simply because the produced clusters are not homogeneous as their generation was based on energy consumption figures in the first place. In other words, a cluster can include all income bands, dwellings types, and household sizes, for instance, as illustrated in Figure 3-13. However, this limitation was overcome by recent studies which tried to utilise the potential of existing geodemographic segmentation systems in terms of grouping neighbourhoods with similar demographic, socio-economic, or even lifestyle characteristics (UKGeographics, 2016), as discussed at an earlier stage. In this regards, it is worth pointing out Druckman and Jackson (2008) study which compared the energy consumption and CO2 emission resulting from two distinct approaches namely; top-down (macro-econometric) and bottom-up (Archetyping). First, aggregate data namely; Expenditure and Food Survey (EFS), was used to retrieve the average household energy expenditure on different fuel types before being converted to energy quantities with the help of the energy price matrix developed by the Centre for Sustainable Energy (CSE). In addition to that, the average CO2 emissions were also calculated based on the emission index of each fuel. Afterwards, the overall UK residential energy consumption and CO2 emission were estimated by multiplying the average household figures by the number of UK households. This was followed by a correlation analysis of various socio-economic factors such as disposable income, with energy use to determine their effect; consequently, predict the energy use given particular households characteristics.

Conversely, the second approach involved retrieving homogeneous socio-economic household characteristics from a geodemographic system (Output Area Classification), which was in turn developed by applying different clustering techniques to the 2001 England & Wales census data including *K-mean*. The average household energy consumption belonging to each group was then calculated based on their socio-economic characteristics using the regression equation developed in the first approach. Once obtained, the overall energy usage of each group was computed by scaling up the average energy consumption by the number of the group's households. Despite the mapping capabilities of this study, which could help in identifying and targeting households with potential energy savings, it suffers from a lack of future demographic and socio-economic updates. This owing to its reliance on census data which are collected every 5-10 years. This issue was supported by many scholars including Gale and Longley (2013) who claimed that *"...census-based classifications such as, the 2001 OAC are uniformly unreliable because there is no temporal updating of input data"*.

### 3.4.3 Advantages of disaggregated approaches

- Since the effect of various socio-economic and demographic factors on the households' energy usage can be defined, bottom-up models have the ability not only to target but also to assess the effect of implemented measures on different household groups (e.g. low-income lone parent).
- In contrast to top-down approaches, disaggregated models have a better prediction accuracy.
- If combined with existing geodemographic segmentation systems, disaggregated models can be easily integrated into GIS systems to serve as an effective energy planning support system.

- Although still not matured yet, bottom-up models showed their ability to utilise the potential of smart metering data.

### 3.4.4 Limitations of disaggregated approaches

- Owing to their reliance on large-scale household surveys, the development of disaggregated models is costly, time-consuming, and usually, involves governmental agencies.
- There is a lack of universal micro-econometric models which is indeed evident in the opposing views on the significance and magnitude of different socio-economic and demographic factors.
- One major drawback of the current residential energy Archetyping approaches is their insensitivity to households' behavioural and socio-economic characteristics future changes.

## 3.5 Conclusion

This chapter has extensively reviewed the energy planning and forecasting approaches supporting the development of techno-socio-economic policies according to their conventional prediction process, strengths, weaknesses while analysing some relevant studies from the literature. Furthermore, proposed a new classification based on the frequent intervention scale of each approach into top-down (aggregate) and bottom-up (disaggregated) models.

First, as indicated in Table 3-2 and Table 3-1, aggregate models are relatively easy to develop since only a few macro-economic and technological inputs (e.g. GDP) are involved. These indicators are widely available from different sources such as government data repositories. Regardless of their type, technological or macro-econometric, these models have a great ability to assess the long-term implication of an economy on the national or regional residential energy demand and vice-versa. Furthermore, they are powerful tools to compare the effect of implemented energy policies across different regions or countries. However, they are not seen relevant to the undertaken research. This is because all households, regardless of their demographic and socio-economic characteristics, are considered to consume the same amount of energy (often an average). In this way, it would be difficult for energy planners to estimate the energy usage patterns at the building scale. Subsequently, the targeting or evaluation of the implication of energy policies directed to particular households (e.g. the ones on fuel poverty), is not possible.

In contrast to the above, disaggregated approaches have a great ability to account for the variation of the individual household energy consumption in function micro-econometric indicators such as household size, annual income, level of education, and socio-economic status. Furthermore, if an Archetyping approach based on geodemographic segmentation systems is combined with a micro-econometric model, it is possible to forecast the energy consumption at the regional or national levels via extrapolation, as discussed in the previous section. Also, such approach can be easily integrated into existing urban planning support systems such as GIS, where energy planners could employ mapping capabilities to identify areas with possible socio-economic improvements or analyse the impact of applied measures such as energy tax compensation. However, the major problem with this approach is its reliance on historical data. This could be problematic knowing that the impact of socio-economic or demographic factors (e.g. children dependency) on domestic energy is not constant and could widely vary over-time due to other external factors. Therefore, conducting regular household panel surveys is necessary to overcome this drawback. The main advantage of this data type is that it is possible to model dynamic relationships. Certain authors such as Longhi (2014) realised this potential and tried to model the change in micro-econometric indicators across three years. However, it is believed that studies pertaining to the disaggregated umbrella **have not fully** explored the potential of panel data. More specifically, both model types, either micro-econometric or energy archetypes, are **static** since they do not account for households transitions over their life-cycle. For example, from "single non-elderly" household to "Couple without children" or to "Lone parent". Undoubtedly, the concept of household life-cycle has been widely employed in other fields such as marketing, as a determinant of consumer behaviour and a prominent principle in market segmentation (Du and Kamakura, 2006). Based on that, the following questions arise;

- *How will the different British household structures evolve in the next 5-10 years?*
- *Is there a significant relationship between household transitions and energy usage patterns?*
- *If so, to what extent it does affect the households' energy consumption figures?*
- *What are the benefits that household life-cycle analysis could bring to urban energy planning?*

All these questions will be addressed in depth in chapters 6 and 8. On the other hand, the next chapter will discuss in detail the suitability of different research methodologies, methods, and instruments for answering the above research questions. Furthermore, will examine the research validity, reliability, and limitations in addition to any ethical considerations.

Table 3-1.(part 1) energy planning and forecasting approaches supporting the development of techno-socio-economic measures

| Approach | Purpose of use | Type | Data type | Frequent Inputs | Method employed |
|---|---|---|---|---|---|
| **Aggregated (top-down) models** | Evaluating the long-term implications of an economy on regional/ national energy demand | Macro-econometric<br>Technological | Historical aggregate data from national agencies, weather stations, or energy providers | -GDP<br>-Fuel price<br>-Employment<br>-number of households<br>-Appliances ownership, efficiency, and usage patterns<br>-Degree days | **Statistical methods**;<br>-Regression analysis<br>-**Artificial intelligence methods**;<br>Artificial neural networks<br>Fuzzy logic<br>Genetic algorithm |

Table 3-2.(part 2) energy planning and forecasting approaches supporting the development of techno-socio-economic measures

| Approach | Purpose of use | Type | Data type | Frequent Inputs | Method employed |
|---|---|---|---|---|---|
| **Disaggregated (Bottom-up) Models** | -Forecasting energy consumption/ demand at the building scale<br><br>-Identifying households with possible socio-economic improvements.<br><br>-Analysing the impact of applied measures (e.g. energy tax compensation) | Micro-econometric | Large scale surveys; panel data surveys, cross-sectional data, annual expenditure surveys | Households size, household age, level of education, presence of children and teenagers, patterns of presence at home, employment mode, socio-economic status, income, marital status, household type, and tenure type | **Statistical methods**;<br>-pooled linear regression<br>-fixed effects regression<br>-random effects regression<br>-generalised method of moments (GMM) |
| | | Archetyping | Large scale surveys; panel data surveys; smart metering data; geodemographic segmentation systems; annual expenditure surveys | Socio-economic, demographic, and environmental attitude variables | -**Statistical methods including**;<br>-Data reduction procedures (e.g. factor analysis)<br>- Regression analysis<br>-Clustering algorithms; K-mean;SOM |

# 4

# Research Methodology

*The fourth chapter intends to address in detail the methodological choices made in this research. Following an inverted pyramid structure, it first discusses the philosophical position of this research based on its epistemological and ontological foundations. Furthermore, explores, justifies and clarifies the adopted research methodology and methods based on the philosophical research position and the nature of research questions. Finally, the chapter expounds on the research design and practical implementations while highlighting any ethical considerations and research limitations.*

**Contents**

## 4.1 Introduction

Chapters two and three have focused on analysing energy planning and forecasting approaches supporting physical and techno-socioeconomic improvement strategies in the residential sector, correspondingly. Overall, the analysis of key studies in both chapters has demonstrated the fragmentation of existing energy forecasting tools and the absence of integrated mechanisms assisting urban planners in their energy planning decision-making.

First, the examination of different approaches in chapter two has permitted to develop a framework which outlines the main components of the envisaged 3D urban energy prediction tool. Moreover, highlights the nature of outputs of each module and the interaction between different components.

On the other hand, the study of methods assisting techno-socioeconomic measures in chapter 3 has allowed reinforcing the statements made in chapter one about the static nature of techno-socio-economic approaches. Furthermore, it has led to the questioning of the impact of households' demographic evolution on domestic energy usage patterns and its significance.

Based on that, chapter four aims to thoroughly discuss the philosophical position of the undertaken research in relation to other perspectives while setting its ontological as well as epistemological foundations. Furthermore, it expounds on the current research strategy comprising the research methodologies, methods, and instruments employed to achieve the research objectives and address the research questions.

## 4.2 Chapter structure

This chapter follows an "inverted pyramid" structure as depicted in Figure 4-1, which means that its analysis will flow from broad to narrow. Thus, section 1 aims to tackle the research paradigms by analysing its ontological and epistemological positions.

After that, section 2 discusses the methodological choices based on the philosophical research position, the nature of research questions, and the strengths as well as weaknesses of each research methodology. This will be followed by addressing the research design and practical implementation in section 3. Finally, section 4 will highlight ethical considerations and research limitations.

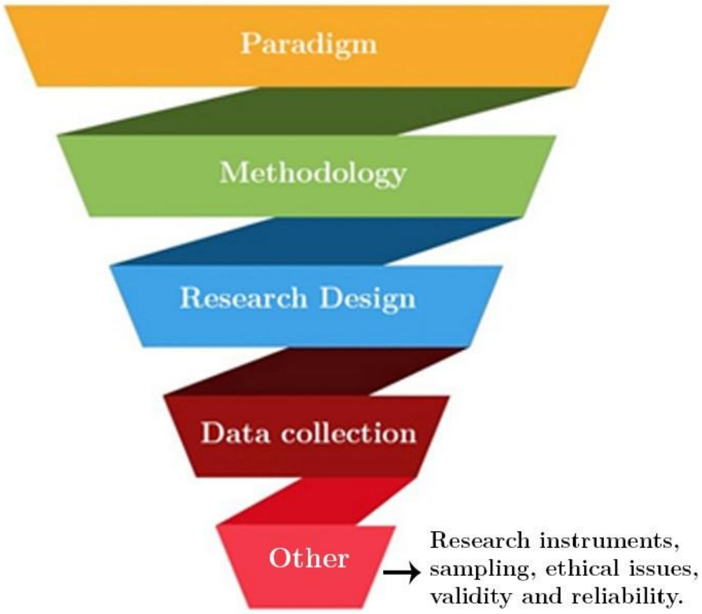



Figure 4-1.Structure of the methodology chapter

## 4.3 Philosophical position of the research

It is indispensable to discuss priorly the research philosophical positions, known as paradigms, by establishing a clear ontological position as well as taking account of the epistemology. Creswell (2013) and Slife & Williams (1995) stress that although some philosophical thoughts are implicit in research, it is necessary to identify them since they have a particular influence on its design. Similarly, Sapsford and Jupp (2006) argue that *"a philosophical stance of worldview underlies and informs a style of research"*. Ontology refers to *"the science or study of being"* where the primary focus in on the study of the nature of reality (Blaikie, 2009). In simple terms, ontology reflects a personal belief and perception of what constitutes reality and how social entities should be perceived; objectively or subjectively (Crotty, 1998). On the other hand, epistemology is *"the theory of knowledge"* which deals with the type and forms of knowledge (Cohen et al., 2013). Furthermore, it examines *"what can be known about the world and how can it be known"* (Furlong and Marsh, 2010). In other words, someone's epistemological assumptions denote how knowledge can be generated, assimilated, and communicated (Guba and Lincoln, 1994).

Table 4-1.Philosophical position of the undertaken research in relation to other positions (part1), adapted from (Creswell, 2013)

| Paradigm | Ontology | Epistemology | Considered in the research | Why? |
|---|---|---|---|---|
| **Post-positivism** | Critical realism- one reality but with fallible apprehension | Modified objectivist/ dualist; researcher position is neutral; findings might be true. | Yes | - One reality determined based on testing the following hypothesis " household transitions affect their energy usage"<br><br>-Future transition patterns are governed by some socio-economic and demographic factors.<br><br>- Knowledge is shaped by observational data. |
| **Constructivism / interpretive** | Multiple realities which are socially constructed | Subjectivism: interaction between participants and researcher(s) constitutes knowledge and values. | No | -The **primary** purpose of the research is not to develop an understanding of urban planners but instead provide a new useful tool to help them meet the CO2 emission targets. |

Table 4-2.Philosophical position of the undertaken research in relation to other positions (part2), adapted from (Creswell, 2013)

| Paradigm | Ontology | Epistemology | Considered in the research | Why? |
|---|---|---|---|---|
| **Critical theory/ Transformative** | Historical/virtual realism shaped by economic, social, ethnic, and cultural values. | Subjectivism; findings are based on values. | Yes | -Closely collaborating with urban energy planners to develop together a new urban energy planning support system.<br><br>- Extend the capabilities of the developed system in the future to include public engagement aspects; e.g. empower vulnerable consumers such as, fuel pauvre. |
| **Pragmatism** | Not devoted to any philosophical system | | No | -No rationale for the employed methodologies and methods.<br><br>- Dubious validity |

Based on the possible combination of different ontological and epistemological assumptions in addition to discipline orientations, there are four to five main philosophical paradigms that have been widely discussed in the literature. These are Positivism/PostPositivism,Constructivism,Critical/Transformative/Participatory, and Pragmatism (Lincoln et al., 2011; Creswell, 2013; Merriam and Tisdell, 2015).

Table 4-1 and Table 4-2 summarise the ontological and epistemological assumptions of the main philosophical paradigms. Traditionally, the positivism movement considers that there is always only an absolute and objective reality which can be unravelled through understanding the set of laws and causal statements that govern the world. This is achieved by applying natural science principles and methods which are empirical and based on hypothesis testing. However, the emergence of Post-positivism in the middle of the 20th century has widely challenged the positivism notion of "Naïve realism", claiming that there is no absolute truth. This is simply owing to the fallibility of observation and measurement arising from the sensory as well as intellectual limitations of human beings.

From this perspective, Post-positivism paradigm will be considered appropriate for this research since there is only one objective reality which can be unravelled through testing the hypothesis *"The future demographic evolution patterns of British household affect their energy consumption"*. Furthermore, the process of supporting or rejecting this hypothesis relies primarily on predicting the future demographic evolutionary patterns which are in turn governed by the nature and the extent of influencing factors. In the undertaken research, this knowledge is shaped by observational data, proof, and relevant considerations where information about a representative sample of the UK households will be first collected and analysed. However, despite the fact that the resulting predicted transition models can be generalised, they are expected to be prone to error as they are based on empirical enquiry.

On the other hand, constructivism or interpretive paradigm, whose origins go back to the work of Berger and Luckmann (1967), posits that social issues cannot be studied through positivism/post-positivism. Indeed, constructivists believe that the world is versatile and that are multiple realities constructed and elucidated differently. This is because individuals, who desire to comprehend the world they live in, establish subjective and intuitive meanings of their experiences considering the diversity of their backgrounds (Creswell, 2013). In addition to that, these realities are not fixed and can be changed as soon as the participant becomes more cognizant. A typical constructive research can comprise multiple, competing, and even opposing realities of the participants including the researcher(s) (Guba and Lincoln, 1994). Therefore, the role of researchers is not neutral as they recognise the influence of their backgrounds on their own interpretation. Moreover, position themselves with the participants to attain a new developed and sophisticated constructions.

The constructivism paradigm was not considered relevant to this research because the purpose of this study is not to develop an understanding and structuring of the subjective meanings of urban planners' experiences. In other words, urban energy planning approaches are well documented and understood despite their fragmentation (DECC, 2012a). Instead, the current research focuses on developing a useful tool that could assist urban planners in their sustainable urban energy planning decision making.

Critical theory/ Transformative paradigm, which goes beyond unveiling the participants' understanding and perception of their environment, evolved in Germany in the 1920's (Campbell and Bunting, 1991). In essence, critical theory inquirers claim that constructivists have neglected advocating for action agenda to empower marginalised and oppressed people (Lincoln et al., 2011). Furthermore, believe in historical realism and argue that there exist multiple realities shaped by cultural, social, economic, gender, and ethnic factors. For these reasons, the goal of a research belonging to this paradigm is not only to understand society but also change and develop a new reality (Merriam, 2014). One advantage of this paradigm is that it acknowledges the participants' expertise in their respective areas and positions the researcher among them to promote and stimulate change (Ford-Gilboe et al., 1995).

In the light of the above, it is clear that the critical theory paradigm is appropriate for the undertaken research since the majority of its principles align with certain research objectives. For example, as discussed at an early stage, one of the pillars of this research consists of closely collaborating with urban energy planners and energy experts to develop together a 3D urban energy prediction tool which considers different "decarbonisation" strategies; physical and socio-economic. However, this is not the only intention as one of the future objectives is to extend the capabilities of developed 3D urban energy prediction tool to address the value of public engagement. More precisely, there will be an attempt to encompass and empower different types of consumers, including the one resistant to change and the fuel "pauvre", since their contribution is vital to the overall sustainable development strategy.

In contrast, pragmatism paradigm, whose roots go back to the work of Charles Sanders Peirce, John Dewey, and Herbert Mead, is not devoted to any system of reality or philosophy according to Creswell and Clark (2007). In fact, it is built upon actions and outcomes instead of antecedents (Creswell, 2013). In other words, pragmatists hold the idea that the research problem is central. Moreover, think that all available approaches can be employed to derive knowledge about this issue without being loyal to any philosophical paradigm (Rossman and Wilson, 1985). Pragmatism is often referred to as the mixed-methods' fundamental philosophical framework (Teddlie and Tashakkori, 2003; Somekh and Lewin, 2005). However, other mixed-methods scholars, including Mertens (2010), oppose this argument and place themselves with the transformative paradigm. The major issue with this paradigm is that it does not discuss the process of choosing mixed methods nor justify their suitability in the research (Hall, 2013). In other words, it lacks a clear and consistent rationale on "*what works*" (Bryman, 2008). For those reasons, some researchers found some disagreement between the declared research rationale and practice (Bryman, 2007). Moreover, claimed that researchers adopting this paradigm encountered barriers in analysing and interpreting the outcomes (Bellotti, 2014). For the above reasons and other related to the dubious validity pragmatic research (Robson and McCartan, 2016), pragmatism paradigm is not taken into account in the undertaken research.

## 4.4 Methodology rationale

The undertaken research embraces a combination of quantitative and qualitative research methodologies for the following reasons. The first reason is related to the philosophical position of the undertaken research which is a mixture of post-positivism and transformative. More specifically, positivists/post-positivists predominantly embrace a quantitative research methodology (Table 4-3), whereas transformative/critical theorists, primarily adopt qualitative research methodology with action research principles. However, transformative scholars can use quantitative research methodology to add more breadth to understanding the research problem.

On the other hand, the second reason is related to the nature of research questions handled by each research methodology (Table 4-4). First, quantitative research methodology is seen adequate because it concentrates on cause and effects as well as empirical data integration. Furthermore, it is characterised by its generalisability and excellent control over samples, and it is deductive, (Balnaves and Caputi, (2001). In the context of this research, quantitative methodology enables the following actions;

- To test hypothesises to explain the association between different factors. For example, in the case of the current research, the relationship between predicted household demographic evolution and energy consumption variables will be investigated through testing this following hypothesis "*There a significant relationship between household transitions and energy usage patterns.*" (Table 4-4).

- When the developed hypothesis or research questions necessitate numerical answers, which applies to the study second research question (Table 4-4) as the focus is on finding the degree of influence of one variable over another.
- To analyse trends in different sectors such as economy and education. For instance, the patterns of demographic evolution for various household types in the future 5-10 years (Table 4-4).

Table 4-3. Existing worldviews with their respective main research methodologies

| Paradigm | Adopted in this research | Methodology (Primarily) |
|---|---|---|
| **Positivism/Post-positivism** | Yes | Quantitative |
| **Interpretivism/ Constructivist** | No | Qualitative |
| **Transformative/ Critical theory** | Yes | Qualitative with critical and action research principles. However, other approaches can be employed to complement the overall picture (Henry and Kemmis, 1985) |
| **Pragmatism** | No | Qualitative and Quantitative |

Table 4-4. The adequate methodologies to address the study research questions based on their type

| Research questions | Type of question | Nature according to (Trochim and Donnelly, 2006) | Adequate methodology |
|---|---|---|---|
| How will the different British household structures evolve in the next 5-10 years? | Open-ended | descriptive | Quantitative |
| Is there a significant relationship between household transitions and energy usage patterns? | Research question | relational | Quantitative |
| If so, to what extent it does affect the households' energy consumption figures? | Open-ended | Causal/ predictive | Quantitative |

In contrast to the above, qualitative research with action research principles seeks to interpret and document a given phenomenon from an individual's viewpoint or reference frame in an exploratory process (Creswell, 2013). Furthermore, it allows an in-depth understanding of the studied subject while enabling the sharing of information among participants to change a current situation. In the context of this research, this methodology is seen relevant because it facilitates the involvement of urban planners to collaboratively benchmark the developed 3D urban energy prediction tool that helps their energy planning decision making.

## 4.5 RATIONALE FOR RESEARCH METHODS

Table 4-5. Most frequent research methods under the post-positivism and transformative paradigms according to Merriam and Tisdell, (2015).

| Paradigm | Common research methods in the literature | Chosen research methods | Type of envisaged data |
|---|---|---|---|
| **Positivism/Post-positivism** | Experiments, surveys, observation studies, content analysis, Quasi-experiments, secondary data analysis, systematic reviews and meta-analysis, audits, | Secondary data analysis | Quantitative |
| **Transformative/ Critical theory** | In-depth interviews, focus groups, participant observation, field-notes, content analysis, case-study analysis, participatory workshops and conferences | Case-study analysis, participatory workshops | Qualitative |

Table 4-5 outlines the most common research methods employed under the positivism/ post-positivism and transformative/ critical theory paradigms. Furthermore, highlights the chosen one for this investigation and the type of the desired data.

First, the main reason behind choosing secondary data over primary data was because it is impractical to conduct a longitudinal study monitoring a large sample of households across the UK and over several years with limited time and resources. Indeed, Ghauri and Grønhaug (2005) suggested that the advantages of secondary data are their accessibility, quality, and feasibility for longitudinal since regular government or organisational studies (e.g. census, surveys) are widely available from the web.

Secondly, although case study analysis has been criticised for its bias, the lack of robustness and generalisation (Yin, 2013), it is adopted in this research for validation purposes. More specifically, by selecting a geographic area and gathering information about some residents' socio-economic, demographic, and energy usage patterns, it is possible to compare the predicted outcomes against **real** energy consumption figures. Subsequently, assess the prediction accuracy of the developed 3D urban energy prediction tool.

Thirdly, Participatory workshops (focus groups) were selected in this research since they are powerful collaborative, interactive, and evaluative tools (Chambers, 2002). Indeed, workshops offer the possibility to bring a group of local urban planners and energy experts together to share their opinions and expertise on EvoEnergy. Moreover, give them the chance to challenge the set agenda, raise new issues, and most importantly work in close collaboration with the researcher to change or improve certain aspects of the 3D urban energy prediction tool (e.g. user interface).

On the other hand, other research methods were not considered. For example, in-depth interviews were excluded since they are costly and time-consuming to recruit and conduct (Boyce and Neale, 2006). Indeed, the number of aspects addressed when evaluating the developed 3D urban energy prediction tool (e.g. interface), might result in a large volume of information, which is challenging to analyse. Apart from that, experiments were omitted because they might create artificial situations. Moreover, they do not necessarily represent real-life context due to their tight control imposed over samples. Indeed, exposing participants under such conditions might influence their real energy related behaviour. Similarly, direct observation was not considered due to the various challenges associated with installing smart meters and analysing their data at the neighbourhood scale. In fact, this requires sophisticated hardware, analytical tools, and the involvement of big data analytics experts (Depuru et al., 2011).

## 4.6 Research design

After exposing the nature of the philosophical position of this study and the employed research methodologies, it is indispensable to discuss the choice of the appropriate research design in this section. Please note that this section contains a concise justification of the adopted research design. For a more in-depth explanation on different research designs and their degree of relevance to this investigation, please consult appendix A.A.2.

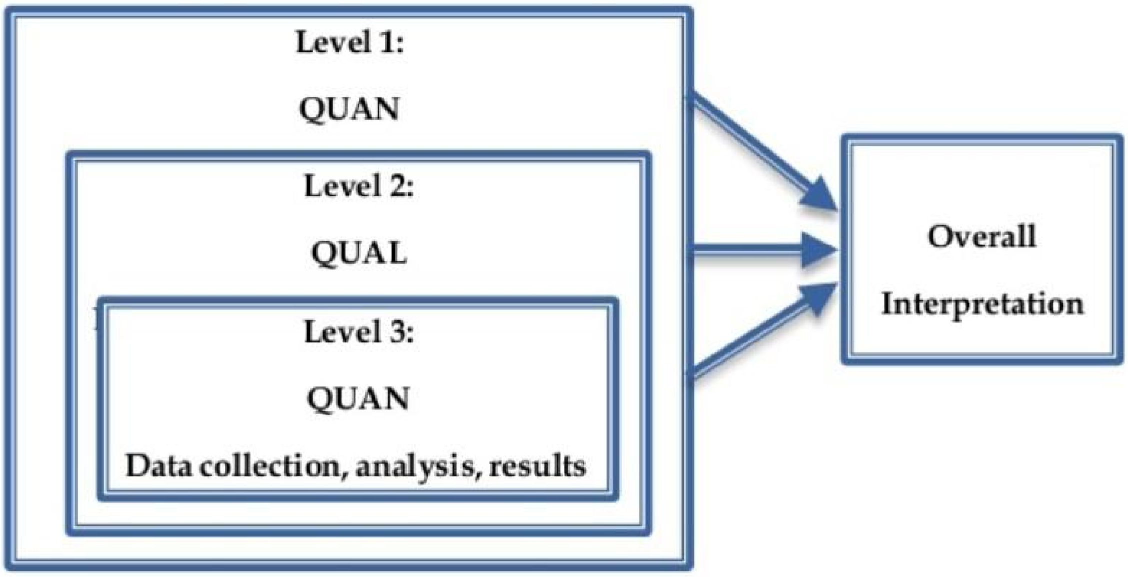



Figure 4-2. The multilevel model (Tashakkori and Teddlie, 2010)

From the premise that both quantitative and qualitative were embraced in the research, a mix-methods research design will be implemented. However, the mixed-methods umbrella is rich as it groups four types of different designs namely; exploratory sequential design, explanatory sequential design, the embedded design, and finally the triangulation design (Creswell and Clark, 2007).

In this research, a triangulated design has been adopted because it allows merging the strengths of quantitative methodology with the quantitative one. Moreover, eliminate their weaknesses. This research design is adequate when the main aim is to validate or further develop quantitative outcomes with qualitative data (Patton, 1990). Another advantage of the triangulation design is its flexibility since it is possible to collect and analyse each data type independently. For those reasons, it is regarded suitable for the undertaken research. However, the triangulated design has many variations such as the convergence model, transformation model, the validating quantitative data model, and multi-levels model.

The multi-levels model was selected because it divides the research into complementary phases. In each stage, a particular methodology and methods for data collection and analysis are employed. Once an initial analysis is performed, outcomes from each level will be combined as one unit to be interpreted (Figure 4-2), which provides a lot of flexibility and control (Teddlie and Tashakkori, 2003). Based on that and considering that this research is comprised of complementary objectives and different types of research questions, it is more practical to divide the whole practical implementation process into different phases.

## 4.7 Practical implementation

Following multi-level triangulation design model principles, this research has been carefully structured around 5 major phases.

### 4.7.1 Phase 0

**Goal and purpose:** The main intention in this stage, which allineates to objectives **one** and **two** of this research, is to examine the structure, strengths, weaknesses, prediction pipelines, and life-cycle applicability of existing energy planning approaches. This will, in turn, help unveil gaps in the pertinent literature. Moreover, facilitate the development of a framework which defines the structure of the envisaged 3D urban energy prediction tool, its development process, the nature of relationship between its different modules, and outlines its prediction process.

**Research methodology:** Mixed-methods methodology (mostly qualitative)

**Research method:** Content analysis

**Sources of content and analysis of data:** this entails the analysis of recent book chapters, conference proceedings in addition to journal and review articles on energy planning and forecasting approaches. First, approaches supporting physical improvement strategies such as retrofit were analysed in Chapter 2, whereas the ones supporting techno-socio-economic measures were discussed in Chapter 3. The examination of each approach in both chapters was in accordance with its structure, its prediction pipeline, its relevance in the building life-cycle, strengths, and weaknesses.

### 4.7.2 Phase 1

**Goal and purpose:** the main ambition in this stage, which aligns with the **third** objective of this research, is to study past demographic evolution patterns of British households. Furthermore, investigate the demographic and socio-economic factors impacting them. This, in turn, will enable the prediction of future household transition patterns over the next 10-15 years.

**Research methodology:** Quantitative.

**Research methods:** An analysis of official secondary data, the British household panel data survey (BHPS), more precisely.

**Research instruments:** Stata, which is a multi-purpose statistical software package with a wide range of capabilities including statistical analysis, data management, and programming, has been adopted in this study. The main motivation behind this choice is due to its user-friendliness and excellent handling of survey data analysis including panel data. Furthermore, its support to external packages developed by other users. SPSS has also been employed along with Stata because of its great data manipulation capabilities such as cases filtering, reshaping data, recording, transforming, and computing new variables.

### Characteristics of the British household panel data survey (BHPS)

The ability to analyse complex dynamic relationships represents the major advantage of panel data (longitudinal data). Moreover, the main motivation behind its adoption in this study. Although the UK longitudinal data survey (UKHLS) is the most recent household panel data survey and has a larger sample size (40,000 households), the British household panel data survey was preferred because it monitors households over a longer period (18 years).

The British household panel data survey (BHPS) was conducted by the Institute for Social and Economic Research (ISER), at Essex University from the period between 1991 and 2009. It comprises more than 5000 UK households aged at least 16 years, who were randomly selected from a Postcode Address File. Furthermore, interviewed each year over this period on different socio-economic as well as demographic aspects. Overall, this encompasses household composition, dwelling conditions, residential mobility, education, employment, socio-economic values, health conditions, social relationships, consumer behaviour, and expenditure. In particular, energy consumption related questions include; the existence of central heating, type of main heating fuel, type of payment plan (e.g. pre-payed, direct-debit, etc.), monthly and annually energy expenditure for gas and electricity. Furthermore, the existence of renewables and different appliances (e.g. dish washer) in the dwelling (ISER, 2016).

The BHPS database was downloaded from the UK data service repository as a zip file containing the 18 waves (UKDataService, 2016). Wave 6 had been omitted since it does not contain any energy expenditure variables before the rest of waves were merged into one database using the households' unique ID and with the help of SPSS. Since this study is part of our research project on the city of Nottingham, which is characterised by a considerable proportion of Single households aged 64 or below (roughly 25%), we decided to focus solely on the transition patterns of this household type. For those reasons, only the householders who were single aged 64 or below in wave 1, were considered. This resulted in an initial sample size of 599 households making a total 6700 across 17 waves after dropping the ones with all bills included. An extensive analysis of the demographic and socio-economic characteristics of the 6700 BHPS households is included in appendixA.3.1.

The parentage of missing data in the BHPS dataset was very small representing approximately 1.58% of all values, 17.35% of the cases, 61.54% of the variables (Figure 4-3). However, since the percentage of missing data in the overall database (1.4%) and particularly in the majority of variables is less than 5%, this proportion can be tolerated according to (Bennett, 2001). Nevertheless, the missing values were predicted using multiple imputations method initiated by Rubin, (2004).

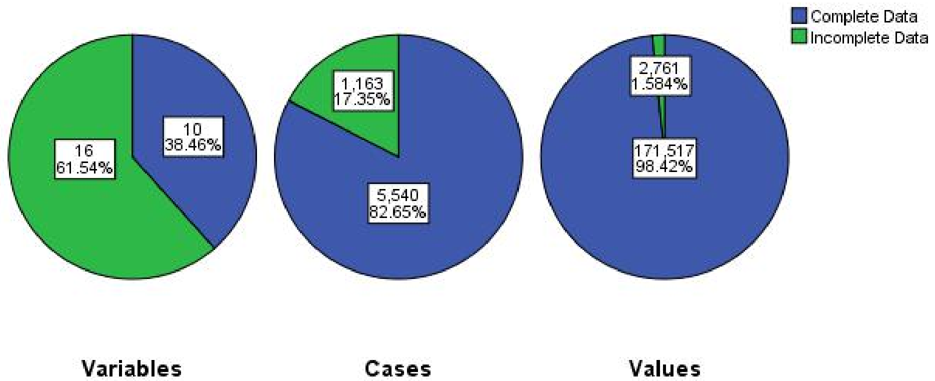



Figure 4-3. Overall summary of missing data in the BHPS database

## Data screening and transformation

The following data screening procedures were performed before data analysis. First, knowing that domestic energy prices in the UK variated significantly from 1991 to 2008 as shown in Figure 4-4, we decided to isolate the price effect on the households' energy consumption. This was achieved by converting energy expenditure to quantities measured in KWh using official inflation index namely; RPI (retail price index) for domestic gas and electricity (DECC, 2015d). Afterwards, all income and energy consumption related variables such as annual households' income, monthly rent, annual electricity and gas consumptions, were normalised by mainly applying square root transformation, except logarithm 10 for the annual electricity variable. After that, outliers were checked for in all continuous variables using Hoaglin et al. (1986) "outlier labelling rule" and deleted. However, all the removed extreme values were predicted using multiple imputations.

In addition to the above, simplification and transformation were applied to the household type variable, which originally contained nine categories. These are; single non-elderly, single elderly, couple no children, couple with dependent children, couple without dependent children, lone parent with dependent children, lone parent without dependent children, and +2 unrelated households. This was first achieved by collapsing the number of categories to five while isolating children dependency as a separate variable. This was followed by transforming the simplified categorical variable into four dummy variables to allow performing logistic regression in the second phase of this study.

In addition to the above, non-logical annual energy expenditure values such as 20 pounds per year, were encountered. To solve this problem, any electricity energy consumption value below the minimum threshold (400KWh) suggested by BRE, (2013) was removed and imputed. Similarly, it was noticed that the majority of non-logical gas expenditure values were given by households who do not possess central gas heating but use gas instead for cooking purposes. Therefore, based on the UK energy demand breakdown chart (Figure 4-5) published by DECC (2012) which suggested that cooking represents 4% of the total energy usage, any value less than 4% (550 KWh) of the average household gas energy consumption (13800 kwh) will be not considered.

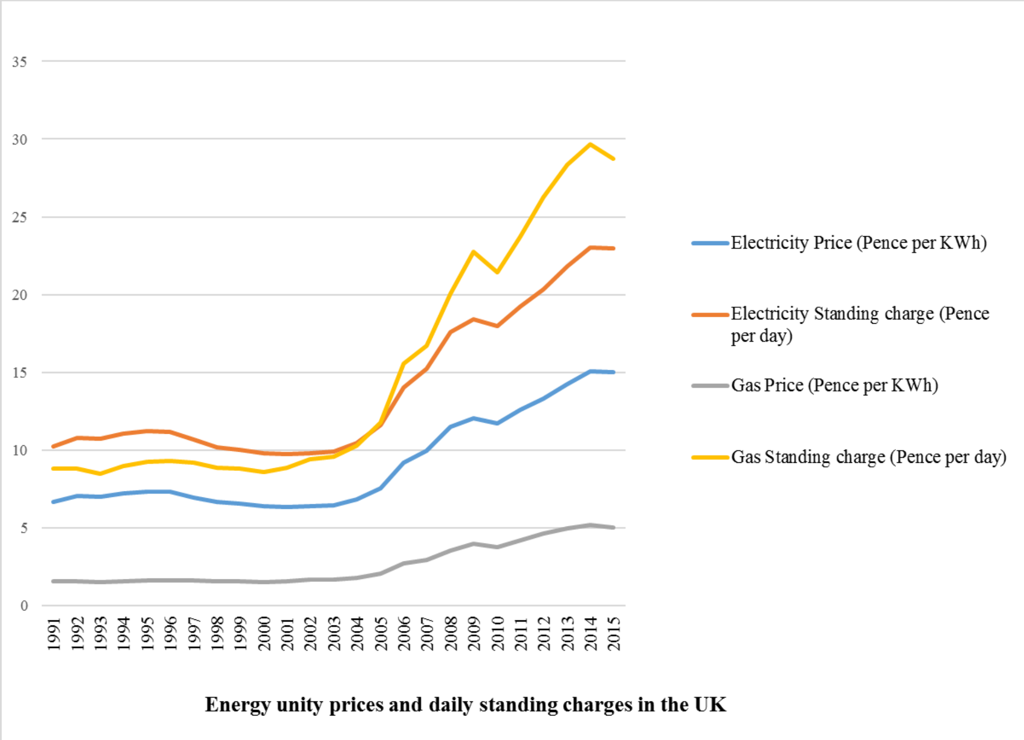



Figure 4-4. Energy unity prices and daily standing charges in pence in the UK from (1991-2015)

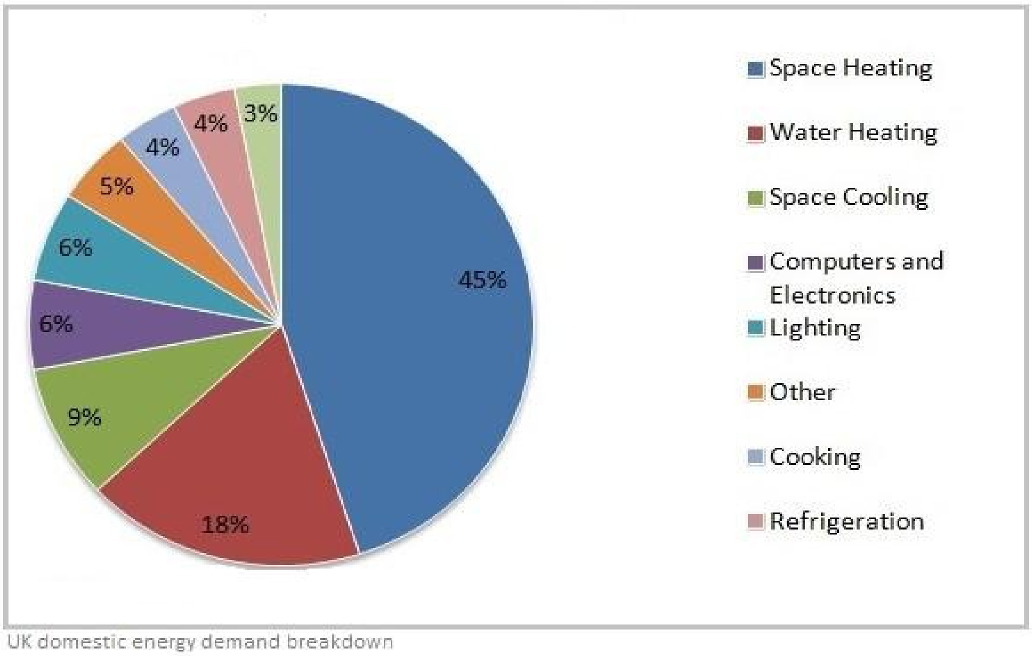



Figure 4-5.UK domestic energy demand breakdown (DECC, 2012b).

## Analysis of the British household panel data survey (BHPS)

The analysis of the BHPS dataset comprises two main stages namely; exploratory data analysis and inferential statistics. First, exploratory data analysis stage involves utilising mostly graphical tools including box-plots, stacked bar charts, and line graphs, not only to maximise insight into the key features of the BHPS dataset but also explore existing transition patterns over the whole period from 1991 to 2009.

On the other hand, the inferential statistic stage, which encompasses statistical techniques (e.g. logistic regression) facilitating the generalisation of the research outcomes on a wider population (Lowry, 2014), consists of mainly identifying the socio-economic as well as factors impacting those existing transition. Subsequently, predict future patterns of demographic transitions covering 10-15 years. Based on that and owing to the binary nature of predicted household type variables, two types of estimation models will be employed, fixed and random-effects logistic regression.

**A-Fixed-effects models**

In any panel data model, unobserved heterogeneity across different individuals (e.g. single non-elderly) is not only central to its analysis but also a major determinant of the nature of the employed estimation model (Greene, 2012; Hsiao, 1985). First, fixed effects models in principle assess how the variation in the dependent variable within each group (e.g. couples with children) is affected by the change in the independent variables inside each group. For those reasons, unobserved heterogeneity is controlled for when variables are constant over time, and so their impact. Moreover, when there is a correlation between constant variables and the independent variable (Wooldridge, 2002). In other words, the effect of time invariant variables such as gender and birth region, is not estimated in fixed-effects models, which signficantly reduces the risk of bias caused by their introduction in the model and allows more consistency (Williams, 2013). Fixed-effects logistic model is described in (Equation 4-1) as follows;

$$log\left(\frac{P_{it}}{1-P_{it}}\right) = \mu_t + \beta x_{it} + yz_i + \alpha_i$$ (Equation 4-1)

$P_{it}$ is the probability that $y_{it} = 1$, $z_i$ represents the vector of variables describing the householders. However, it does not vary across time. On the other hand, $x_{it}$ is the vector of variables that vary across individuals and time. $\beta$ and $y$ are vectors of coefficients while $\mu_t$ is the intercept which can change over time. $\alpha_i$ depicts the differences between householders who are stable over time and those otherwise, based on $z_i$. After applying some algebra on (Equation 4-1) as well as assuming the independence of $y_{i1}$ and $y_{i2}$ given any householder $i$ and its $\alpha_i$ value, $z_i$ and $\alpha_i$ will be excluded from the equation as shown in (Equation 4-2). This will lead to the elimination of householders who do not vary with the response variable and to the creation of difference scores for all time-varying independent variables (Allison, 2006).

$$log\left(\frac{\Pr(y_{i1}=0, y_{i2}=1)}{\Pr(y_{i1}=1, y_{i2}=0)}\right) = (\mu_2 - \mu_1) + \beta(x_{i2} - x_{i1})$$ (Equation 4-2)

### B-Random-effects models

In contrast to fixed effects models, which allow the correlation of unobserved individual effects with the included variables, the random-effects ones consider the unobserved heterogeneity uncorrelated with the observed variables (Greene, 2012). This will bring the following benefits. First, it will permit the estimation of the effect of time-invariant variables such race. Secondly, it will result in a usually smaller standard error of the estimates than in fixed-effects models (Wooldridge, 2010). However, omitting pertinent time-invariant variables from the model might result in biased coefficients (Allison, 2006).

A random effects logistic regression model is obtained (Equation 4-3) through the inclusion of a random intercept $\zeta j \sim N(0, \psi)$ which is assumed to be independent of the covariates and equally distributed among all householders j.

$$\text{logit}\{\Pr(yij = 1|xij, \zeta j)\} = \beta 1 + \beta 2x2j + \beta 3x3ij + \beta 4x2jx3ij + \zeta j \qquad \text{(Equation 4-3)}$$

It should be noted that the responses yij for any householder j and at any given time, are independently Bernoulli distributed (Thomas et al., 1998), which could be formally written while defining;

$\pi ij \equiv \Pr(yij \,|xij, \zeta j)$ as shown in (Equation 4-4);

$$\text{logit}(\pi ij) = \beta 1 + \beta 2x2j + \beta 3x3ij + \beta 4x2jx3ij + \zeta j \qquad \text{(Equation 4-4)}$$
$$yij \,|\pi ij \sim \text{Binomial}\,(1, \pi ij)$$

### C-Fixed or random effects model?

On the one hand, it could be initially argued that fixed-effects models are more appropriate for this investigation, considering that the employed BHPS database does contain time-invariant variables and that the characteristics of each household could vary from one year to another. On the other hand, this assumption might be violated if the predicted variables vary widely across individuals, which could result in biased coefficients and large standard errors. Based on that, it is clear that modelling unobserved heterogeneity is challenging and not a straightforward task (Longhi, 2014). Therefore, an objective determination of the nature of the model (fixed or random) namely; Hausman specification test (Hausman, 1978), will be adopted to avoid any bias, assumptions, or misinterpretation.

The Hausman test of specification compares fixed and random effects estimates based on a null hypothesis suggesting that there is no correlation between independent variables and individual specific effects (Hausman, 1978). Furthermore, on the principle that both models are consistent estimators of the predicted variable since the difference between their covariance is not statistically different from zero. This suggests that random effect-models are appropriate if the null hypothesis is not violated. Otherwise, fixed-effects models are relevant.

### 4.7.3 Phase 2

**Goal and purpose:** this stage, which aligns with objectives **four and five** of this research, seeks first to investigate the impact of households' transition patterns on their energy usage for gas and electricity. Subsequently, future energy consumption patterns will be predicted while considering the variation in the households' socio-economic as well as dwelling characteristics. For feasibility purposes, only electricity patterns will be predicted.

**Research methodology:** Quantitative

**Research methods:** quantitative secondary data analysis

**Research instruments:** Stata statistical software package will be employed to examine the effect of predicted transition variables on domestic energy usage patterns due to its ability to handle panel data analysis, as discussed at an earlier stage.

**Data collection method and analysis:** After predicting the evolutionary patterns of households in phase 3, statistical techniques, correlation analysis more precisely, will be first applied to BHPS panel dataset to test the hypothesis of whether transitions patterns have a significant impact on the energy consumption of households. Furthermore, to investigate this relationship's strength, if it exists. Since the predicted household transition variables are dichotomous (binary) and the annual energy consumption variables are continuous, point-biserial correlation is employed. Although it is considered a special case of Pearson product-moment correlation, point-biserial correlation can be handled using Pearson correlation in any statistical software package including Stata and SPSS (Kornbrot, 2005). The point biserial correlation coefficient $r_{pb}$ is calculated as follows ((Equation 4-5);

$$r_{pb} = \frac{\overline{Y_1} - \overline{Y_0}}{\overline{S_y}} \sqrt{\frac{N_1 N_0}{N(N-1)}} \qquad \text{(Equation 4-5)}$$

Y0 and Y1 are means of observations that are coded 0 and 1, respectively. On the other hands, N0 N1 are the number of observation coded 0 and 1, correspondingly. N is the total number of observation, whereas, Sy is the standard deviation of all the observation.

Like most correlation coefficients such as Spearman's rho, the coefficient of point-biserial correlation ranges between -1 and +1, in which -1 suggests a perfect negative association. On the other hand, 0 indicates no association while +1 represents an absolute positive relationship (Varma, 2006).

On the contrary, if the investigated association between households' transitions and their energy consumption is found significant, future energy patterns will be estimated in function of the predicted transition patterns, and variation in the households' socio-economic and dwelling characteristics. Based on that, two types of multiple linear regression models will be employed namely; fixed-effects and random-effects. For more information on the main differences between both estimation models, please refer to the previous phase.

In a fixed-effects model, which is shown in (Equation 4-6), unobserved heterogeneity (individual effect) $z_i$ is correlated with the vector of independent variables $x_{it}$ of individual i over a period t. $\alpha_i$ embodies the group specific constant term in the model (Greene, 2008).

$$y_{it} = X_{it}'\beta + \alpha_i + \varepsilon_{it} \qquad \text{(Equation 4-6)}$$

Conversely, in a random-effects model, unobserved heterogeneity is assumed to be uncorrelated with independent variables $x_{it}$ as shown in the following (Equation 4-7(Equation 4-8). This will, in turn, introduce $u_i$which is a random element specific to each group.

$$y_{it} = X_{it}'\beta + E[z_i'\alpha] + \{z_i'\alpha - E[z_i'\alpha]\} + \varepsilon_{it} \qquad \text{(Equation 4-7)}$$

$$= X_{it}'\beta + \alpha + u_i + \varepsilon_{it} \qquad \text{(Equation 4-8)}$$

Finally, Hausman test of specification, which is described in detail in the previous phase, is embraced to decide between both types of models.

### 4.7.4 Phase 3

**Goal and purpose:** The main aim of this phase, which aligns with the **sixth** and **seventh** objectives of the undertaken research, is to choose a geographical area within the city of Nottingham, study its socio-demographic characteristics. Furthermore, build a 3D city semantic model using the relevant standard and technologies with the possibility to support data integration across different urban domains such as CityGML. This will be crucial for the implementation and validation of the research findings.

**Research methodology:** mixed-methods methodology but mostly quantitative.

**Research methods:** case-study analysis and secondary data analysis of government official data including census data.

**Involved research instruments:** 3D modelling tools, open-source gaming engines, and C# scripting.

**Pilot area selection criteria:** the following criteria are considered when selecting the pilot area. First, due to the limited resources present in this doctoral research, geographic proximity is an indispensable criterion to take into account. According to Yin (2013), this will enable a more prolonged relationship between the investigator and the pilot case, which, in turn, allows the researcher to study a given phenomenon from different angles. Based on that, the envisaged neighbourhood should be within a walkable distance from Nottingham Trent University to enable the researcher easily access the site and perform different data collection procedures, if needed. Secondly, data availability is a prominent criterion that facilitates the pilot case selection process according to Gerring (2008). Indeed, in this research, accessibility to the pilot area residents' general information (e.g. name and address), demographic, socio-economic characteristics and their aggregate energy consumption history, is crucial for this study in general, and task 5 in particular.

**Pilot area's selection process and data collection methods:** Based on the above selection criteria, the subsequent steps have been performed. First, an open-source online mapping tool (Figure 4-6) namely; (FreeMapTools, 2017), will be employed to allocate all neighbourhoods within a 1-mile radius of Nottingham Trent University. Subsequently, generate a list of postcodes based on google maps database. After that, the availability of official secondary data sources, recent government census data, more precisely, on all the suggested neighbourhoods will be checked. Based on this, areas with no or not enough data will be excluded from the list. Thirdly, in collaboration with Nottingham energy partnership, the area with most data on the households' demographic, socio-economic, and dwelling characteristics, in addition to energy consumption patterns, will be selected.

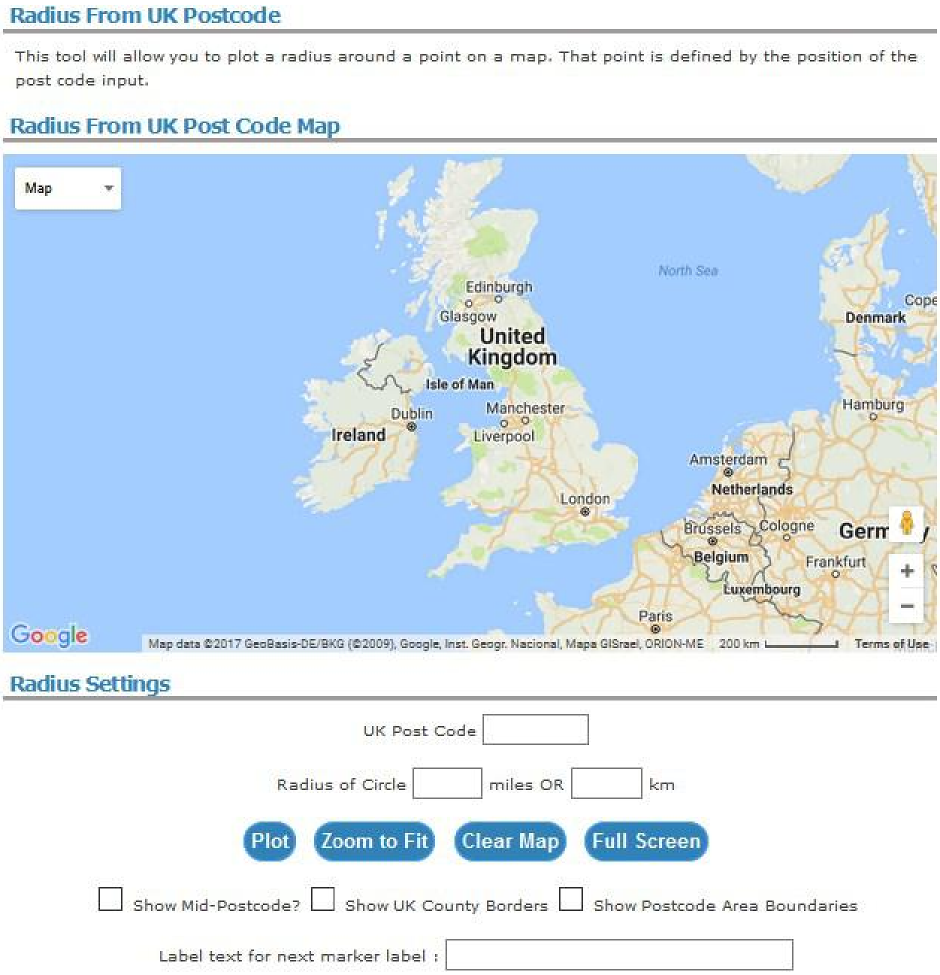


Figure 4-6. The user interface of "Radius from UK postcode" tool (FreeMapTools, 2017)

**Analysis of the pilot area socio-demographic characteristics:** Once an area is selected, a quantitative secondary analysis of UK census data is used to analyse its demographic, socio-economic, and dwelling characteristics. This involves mainly descriptive statistical procedures such as frequency analysis, measures of central tendencies (e.g. mean, median), measures of dispersion (e.g. standard deviation), and cross-tabulation. Moreover, graphical tools such as bar charts and line graphs to compare the socio-demographic profile of the pilot area with the one of Nottingham city. This would be in turn so beneficial since it enables exploring the degree of representativeness of the selected area for future work covering the city of Nottingham.

**Development of the pilot area 3D semantic model***:* the analysis of chapter 2 has unravelled the key features of the envisaged urban energy model. In particular, it has suggested that CityGML, with a minimum level of detail (s) of LOD3, is the standard for urban energy modelling. Based on that, it is evident that the 3D modelling process should articulate on the digital tools and technologies that support this standard. However, it should be noted that BIM software packages, including Revit, will be excluded from this analysis owing to their interoperability issues with CityGML LOD3 models, as discussed in chapter 2. In addition to CityGML LOD3 criterion, the developed 3D model should also meet the subsequent criteria;

- **Interoperability with major open-source game engines**: Since open-source game engines such as Unity3D or Unreal engine, will be the main platform for the developed urban energy modelling, the employed 3D modelling software packages have to fully interoperable with their standards without the loss of information and the involvement of geometry repair tools.

- **Geo-location support**: One of the requirements when choosing a 3D modelling software is geo-location and mapping capabilities. More specifically, it should contain a GIS coordinate system, so that the modelled dwellings are correctly positioned according to the global coordinate system. This, in turn, enables further semantic querying such as post-code queries. On the other hand, mapping capabilities (e.g. Topographic maps generation) permits a faster, realistic, and accurate 3D modelling.

- **Portability**: Since a future goal of this research is to address the aspect of public engagement in urban energy planning, the developed 3D urban energy prediction tool needs to run on different platforms including smart phones, tablets, and the web. Therefore, the file size is a crucial determinant of the involved 3D modelling technologies.

**Level of complexity:** this comprises the number of steps involved in the possible 3D production pipelines and the level of difficulty of each of the encompassed tools and technologies. Furthermore, the level of support (e.g. availability of online resources, training, and forums).

### 4.7.5 Phase 4

**Goal and purpose:** the main intention at this stage, which aligns with objectives **6** and **7**, is to build a 3D urban energy prediction tool (EvoEnergy) based on the 3D semantic model of the pilot area and the statistical models developed in phases 1 and 2. Moreover, to benchmark EvoEnergy in collaboration with energy experts and urban planners.

**Research methodology:** mixed-methods

**Research methods:** This involves 3D modelling and scripting. Moreover, action-research through conducting a participatory workshop.

**Urban 3D semantic energy modelling:** after accomplishing the pilot area's 3D CityGML model and statistical models, the semantic energy modelling occurs. This mainly encompasses integrating the developed mathematical models into the CityGML model while wrapping up different functionalities and components in an intuitive and user-friendly interface. This will be attained by using games engines and some scripting. The former serves as the main platform for navigating through the model, accessing different household information, performing semantic querying and predictions. Moreover, visualising the estimated energy usage patterns at different scales (e.g. building scale). On the other hand, scripting will be employed to set some rules that govern the behaviour of the different 3D urban energy prediction tool's components including; socio-economic and physical in addition to its user-interface.

**Participatory workshop:** Once the 3D energy semantic model has been accomplished, a participatory workshop, involving the researcher and urban planners, will be held. The main objective of this workshop is not only to obtain experts' opinions on the developed semantic model but also enable them to challenge and help improve its functionalities. For instance, this could include; proposing new features and re-organising different elements of the user-interface. Based on that, the workshop has been carefully planned around three mains stages namely; planning, conducting the workshop, and finally post-workshop.

**A-Planning the workshop**: According to Elliott et al. (2005), the first step towards planning an effective workshop is to ensure that a minimum of 3 staff members are involved. Thus, the researcher and the supervisory teams will work collaboratively during this stage. After securing staff members, the next step consists of finding 4-8 potential urban energy planners and energy experts based in Nottingham to be recruited. Thanks to Nottingham energy partnership for providing contact details of some potential participants. Next, the workshop's logistical matters are planned. First, owing to the specific nature of the discussed topic, the workshop will take no longer than 2 hours. Secondly, owing to its convenient location for the participants, the workshop will be held at Nottingham Trent University (city campus). Moreover, catering will be arranged while considering special dietary requirements such as vegetarian.

After arranging logistics, all questions and activities supporting an active participation in the workshop are planned, piloted, and improved, accordingly. The workshop is divided into three main parts namely; introductory part, question and answers, and participatory evaluation.

**Introductory part:** entails delivering a 5-10 minutes PowerPoint presentation which aims to provide some background information and a brief overview of the topic. The presentation covers the following points;

- The overall project's aim and relevant objectives.
- The strengths and weaknesses of existing urban energy planning approaches.
- The contribution made by the researcher to existing body of knowledge.
- How will the developed 3D energy semantic model support energy planning decision-making?

**Questions and answers part**: After the main researcher has delivered the short presentation, participants will be asked 3-5 open-ended questions. This will cover the below areas;

- Their preferences on the type of existing urban energy planning system.
- The type of information needed for effective residential energy planning decision-making, based on their experience.
- Their opinions on the importance of household transitions in urban energy planning decision-making.

The questions follow a hierarchical order based on their complexity in which the general ones are placed first and the specific ones towards the end. This will allow getting a sense of the participants' level of knowledge. Moreover, gaining a deeper insight into their perception/ misconceptions about residential energy forecasting (Elliott et al., 2005). This part will take approximately between 30 to 50 mins including time for group discussion.

**Participatory evaluation**: Once the participants' answered all the questions, the researcher will briefly, in 10 mins, present findings of the impact of household transition on domestic energy consumption. However, this analysis will be simplified using diagrams and tables. This will be followed by a demonstration of the developed 3D urban energy prediction tool. In this instance, participants have the option to follow and practise on the provided laptops. Alternatively, they practise at the end of the demonstration and ask for additional help, if necessary. After that, the researcher will open a power point slide containing the layout of the developed forecasting model and kindly ask the participants to evaluate it in relation to energy planning decision-making. This will be achieved through performing the following actions;

- Propose new features: by drawing a solid filled rectangle to show its size and location within the interface. Each created feature should include a title and one-line comment to explain its purpose.

- Re-arrange existing components: participants may re-allocate the location of different components in order to improve the process of urban energy planning decision-making.

- Suggest the improvement of visualisation as well as graphical aspects of the user-interface components. Moreover, propose other visualisation patterns or graphical representations by using text boxes with some comments.

Once the workshop session has been planned, a pilot study comprising the researcher, supervisory team, and two urban design PhD students, will occur. Any encountered challenges and feedback obtained from these participants will be taken into account to improve the workshop planning. Once the necessary changes have been made, a pack containing the participants' information sheet as well as consent form, will be emailed to the participants who are then required to confirm their attendance by replying. The information sheet explains in detail the purpose of the study, the nature of participation, what the participation entails, how information is recorded, the benefit of the study, and the right of participants (e.g. withdraw).

**B-Conducting the workshop:** Upon participants' arrival, the researcher and the supervisory team will greet them. Moreover, make (s) a small talk which is not directly related to the addressed topic. According to Stewart and Shamdasani (2014), this will allow the moderator quickly evaluate the communication styles of contributors. Next, the moderator starts recording the session and introducing himself to the participants while highlighting that the workshop is an opportunity to express their opinions and improve existing energy forecasting tools. Moreover, explains the importance of results in the current research and outlines some ground rules such as sessions' outline, maintaining anonymity and confidentiality. After that, the moderator follows the plan made in **phase A** while being assisted by the supervisory team in managing discussions. For instance, a flip chart can be employed to write down some dominant participants' opinions/ answers to help stimulate ideas from less dominant contributors. At the end of the workshop, the researcher summarises the main viewpoints and confirms its accuracy with the participants. Furthermore, answers any questions from the participants. Finally, contributors will be thanked for their invaluable contribution and informed what will happen in the coming stages.

**C-Post-workshop:** According to Stewart and Shamdasani (2014) the process of analysis should start during the workshop. In other words, the researcher should pay attention to inconsistent or vague comments and seek for possible interpretations. This could be easily achieved by paraphrasing or summarising key questions to obtain confirmation. Immediately after the workshop, a diagram of seating arrangements will be drawn. This will serve as a referential point to help remember all the participants' names when transcripts are being prepared (Morgan, 1996). Next, the participants' written answers and PowerPoint files will be analysed. Finally, a report written in a question-by-question format will be prepared. Moreover, discussed with the supervisory team and revised, accordingly.

## 4.8 Validity and reliability of the research methods

A. **The British household panel data survey (BHPS):** the suitability, superiority, and reliability of BHPS data for social science analysis have been well justified in the literature (Lambert, 2006; Rose and Pevalin, 2000; Scott et al., 2001). Indeed, Lambert (2006) claims that BHPS was initially designed to provide support for British research and open new doors for them to analyse in more depth different processes across the field of social science. This is evident in numerous studies utilising BHPS data to address previously irresolvable issues or introduce innovate research questions. This statement was supported by Laurie (2010) who claims that "*The BHPS has been a huge success over the past eighteen years and is highly values as a longitudinal data resource for the UK*".

B. **Participatory workshop:** focus groups and participatory workshops are usually prone to reliability issues due to their small sample sizes and specific context. This prevents their outcomes from being generalised on a wider population (Morgan, 1996). However, since the role of participatory workshops in the study is only to obtain initial invaluable feedback from experienced urban energy planners and to generate new ideas through the group dynamics, generalisability is not central. Nevertheless, the workshop(s) questions will be carefully formulated to be relatively specific and a pilot workshop will be conducted to avoid any arsing reliability issues.

   On the other hand, workshops and focus groups are well- known for their high content validity since they allow the analysis of participants' opinions, behaviour, and attitude (Babbie, 1992; Morgan, 1996; Freitas et al., 1998). Therefore, the results obtained from this workshop(s) are indisputably valid despite their lack of generalisability, as addressed previously.

## 4.9 Ethical issues

Although the undertaken research might unintentionally and unavoidably involve vulnerable household heads namely; people aged above 65 years and pregnant women, there are no ethical issues arising because the nature of the topic is not sensitive and do not cause any harm or discomfort to the participants. Furthermore, there is NO intimate or close examination of the participants nor intrusion of their privacy. Nevertheless, the following measures and precautions will be taken to prevent any arising issues. First, participants will be not visited at home at any time or under any circumstances. In fact, the researcher or the supervisory team will only contact them via mail/ email or phone at their convenience. Secondly, although the researcher and supervisory team are aware that the question related to the existence or type of benefit might be intrusive for a respondent, it is impossible to exclude them since they are indispensable to address fuel poverty. Apart from that, NO other sensitive questions will be asked at any point. Thirdly, to ensure the compliance with data protection act 1998, informed consent and information sheet briefly explaining the purpose of the project, the voluntary nature of the contribution, its extent, and benefits gained from taking part in the study, will be first sent to the participants to seek their permission. In addition to their right to withdraw at any time, all participants will be treated equally and with respect. Urban planners, who will attend the workshop(s) taking place at Nottingham Trent University, will be contacted to arrange a convenient time. Moreover, asked whether they have specific access needs (e.g. Wheelchair and hearing impairment) to receive the required assistance. Finally, participants' confidentiality and anonymity will be maintained at any stage. Any audio or video content will be stored in a secured folder which will be only accessed by the researcher or the supervisory team.

## 4.10 Conclusion

This chapter has addressed in detail the methodological choices made in this research to adequately answer the research questions. First, the philosophical research position was discussed in relation to existing paradigms through defining the study ontological position. Moreover, examining the epistemological assumptions of the research process. Based on that, two philosophical paradigms namely; post-positivism and critical theory/ transformative were considered relevant. Post-positivism was chosen since empirical inquiries shape part of the research generated knowledge. More precisely, the impact of household transitions on residential energy consumption patterns can be only measured using statistical methods. On the other hand, critical theory/transformative was adopted since it perfectly aligns with one of the study objectives, which is closely working with urban energy planners and energy experts to benchmark EvoEnergy. Furthermore, to possibly engage different consumer segments in the process of urban energy planning and empower vulnerable households (e.g. households on fuel poverty) in the future. Based on that, a mixed-methods methodology was embraced. In particular, a multi-level model triangulation design was preferred over other designs because of the complementary nature of the research objectives and its great flexibility. There is a total of 5 stages of implementation in which secondary data analysis, 3D modelling, and participatory workshops are the main utilised research methods.

The next chapter (5) aims to analyse in depth the demographic and socio-economic factors affecting single non-elderly household transitions to different household types in the next 10 years.

# 5

# Socio-economic and demographic factors affecting household transitions

*The fifth chapter aims to explore the demographic and socio-economic factors that have an impact on single non-elderly household transitions to various family structures in the next 10 years.*

## Contents

## 5.1 Introduction

The previous chapter has discussed and justified important methodological choices made in the study based on its philosophical stance and the nature of research questions. This, in turn, enabled the evaluation of different research designs and the development of an overall research strategy to ensure addressing the research questions adequately.

On the other hand, this chapter analyses the demographic evolution patterns of British households. Moreover, highlights the socio-economic, demographic, and dwelling related factors contributing to those transitions. In particular, there will be a focus on the transition patterns of single non-elderly households given their large proportion in the selected pilot area and the city of Nottingham with 32% and 25%, correspondingly. Based on that, the British household panel data survey (BHPS) has been utilised to predict future transition patterns.

## 5.2 Chapter structure

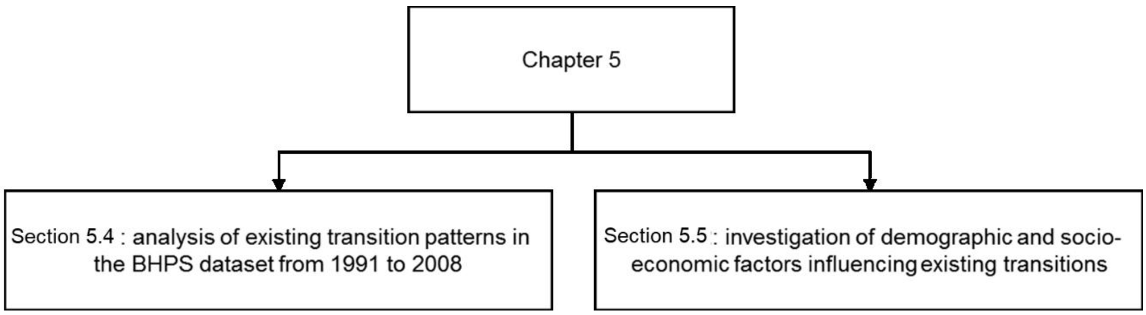



Figure 5-1.Structure of chapter 5

This chapter has been divided into two parts (Figure 5-1). The first section explores and examines the transitions patterns that occurred in the BHPS dataset from 1991 to 2008 of households who were single non-elderly in 1991. On the other hand, the second section investigates the demographic and socio-economic factors that have influenced existing transitions to predict future 40 transition models covering a 10-

year frame. It should be noted that each model represents the transition to a particular household type (e.g. lone parent) in one-year time. However, for simplification purposes, only the models of years 1, 5, and 10 are reported in this chapter. To consult the remaining models, please refer to appendix A.3.3.

## 5.3 The British household panel data survey classification of households

In the British household panel data survey (BHPS), households are classified into;

- single non-elderly
- single elderly
- couple no children
- couple with dependent children
- couple without dependent children
- lone parent with dependent children
- lone parent without dependent children
- +2 Unrelated adults
- Other households

According to the British household panel data survey user manual published by Taylor and Change (1995), the above classification is influenced by three factors namely; **household size**, **presence of couple in the household**, and **existence of dependent children**. However, since categorical variables with many groups can inflate the dimension of the analysed dataset as indicated by Shmueli et al. (2016), the number of categories in the household type variable have been collapsed into 6 categories while isolating children dependency as a separate variable in the database (for more information please consult appendixA.3.1). The new resulting 6 categories are as follows;

- Single non-elderly
- Single elderly
- Couple without children

- Couple with children
- Lone parent
- +2 Unrelated adults
- Other households

In order to follow the analysis of different household transition models in this chapter without any confusion, it was thought that it would be extremely helpful to define each household type according to the British household panel data survey manual.

**A single non-elderly**: refers a household that is occupied by one person only. This could either a male or female and must be under the pensionable age, which is 65.

**A single elderly:** It is a household type that is occupied by one person only (male or female) and whose age must be over the pensionable age (65).

**A couple without children:** It is a household type which encompasses 2 people of different or similar gender and who do not possess any children. The type of relationship that governs this household can take various forms such as legal marriage, cohabitation, and civil partnership (ONS, 2014a).

**A couple with children:** It is a household that is composed of 2 people of different or similar gender and who must have at least one child. In the children are under the age of 16 or 18 (if they are in full-time education), they would be classified as dependent. On the other hand, if the children are aged over 16 or 18 (if on full-time education), they are independent to their parents. Like couple without children households, parent's relationship can take various forms such as marriage or cohabitation (Whiting, 2010; ONS, 2014a; ONS, 2014b).

**A Lone parent:** It is a household type occupied by a person (male or female) and who has one or more than one child but no partner or husband/wife living with them (Whiting, 2010; ONS, 2014a).

**+2 Unrelated adults:** it refers to a household type that does not comprise children or couples. Instead, it consists of more than two unrelated adults living together. Typically, friends, students, flatmates, and co-workers (Whiting, 2010; ONS, 2014a).

## 5.4 Single non-elderly households transition patterns from 1991 to 2008

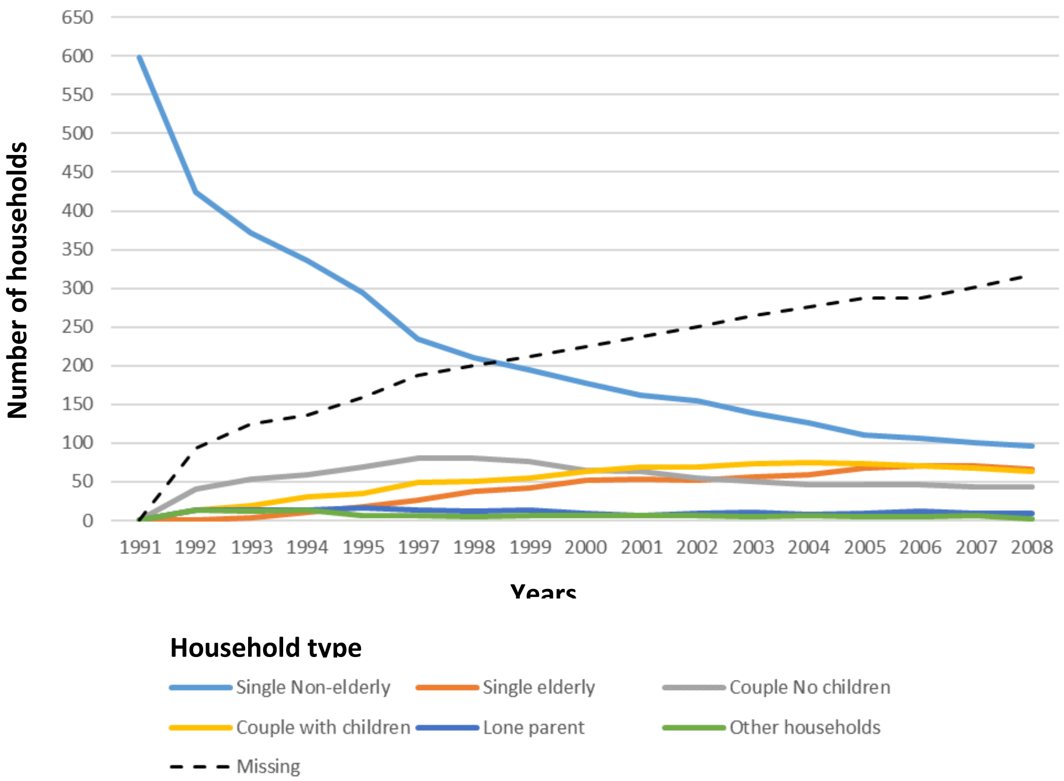



Figure 5-2.Change in the number of different households in the BHPS dataset between 1991 and 2008.

Figure 5-2 is a multiple line graph that represents the overall change in the number of various household types in the BHPS dataset over the period 1991-2008. In general, it is evident that there was a decrease in the number of single non-elderly households during this period. Furthermore, the number of couples with children and elderly households increased. On the other hand, the proportion of other and lone parent households remained relatively stable. These facts suggest the existence of hidden transition patterns in the analysed dataset. Thus, we utilised the stack bar chart

(Figure 5-3) to help closely investigate the occurring inner transition patterns of single non-elderly households over 5, 10, and 15 years.

From 1991-1997, there was a 60% drop in the initial number of single non-elderly households of 599. On the other hand, the number of couples without children, couples with children, and single elderly households increased from 0 in 1991 to reach 84, 51, and 26 in 1997, respectively. This could be explained by the fact that each year, around 20% of single non-elderly households moved to other household types, of which 6% were couples without children, 1.8% single elderly, and 4% shared among couples with children, other, and lone parent households. Considering the 20% yearly transition, the decrease in the proportion of single non-elderly households would have been greater, if there were no transition from the other family types to single non-elderly types. For instance, it was found that around 6.5% of couples without children transitioned to single non-elderly households.

From 1997 to 2002, the number of single non-elderly and couple without children households decreased by 35% and 32%, respectively. Conversely, there were 100% and 40% increases in the proportion of single elderly and couple with children households, respectively. The decline in the proportion of single non-elderly households was not as significant as in the previous period, as there was a 7.5% decline in their annual transition rate. In addition to that, the rise in the number of single elderly households is due to the increase in yearly transition rate to this family type, from 1.75% to 3.48%. Similarly, the surge in the proportion of couple with children households was affected by the lack of transition of a considerable proportion of this household type (roughly 93%) during this period despite the decrease in the transition rates from other household types to this family type.

Finally, between 2002 and 2008, there were no major changes to report from the precedent period except 20% and 74% declines in the percentage of couples without children and single non-elderly households, respectively. Thus, the proportion of couples with children increased between 2002 and 2004, before stabilising.

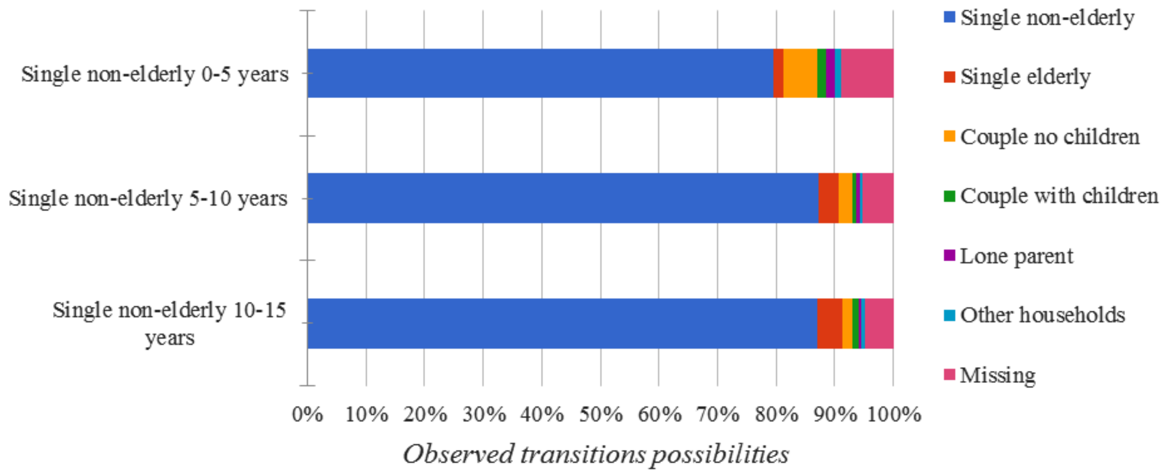



Figure 5-3. The observed transition possibilities of single non-elderly households to different family types over 5, 10, and 15 years.

## 5.5 Socio-economic and demographic factors affecting household transitions

### 5.5.1 Transition to single non-elderly households

#### General trend

Table 5-1 and Table 5-2 represent the logistic regression models of transition to a single non-elderly household in the next year, 5 and 10 years, respectively. In general, the analysis of the ten transition models suggested that single non-elderly are so unlikely to remain within the same household structure after five years, as expected. Furthermore, income variables had a negligible effect on this type of transition. Finally, the significance and magnitude of the majority of variables were not consistent over the ten models. For example, the dwelling size variable was significant in years 1, 5, and 8 but insignificant in remaining years. However, certain factors were consistent over the period from 1-5 years or 6-10 years. For instance, factors namely; cohabiting couples, household size, and householder age were compatible across the years 1 to 5. Moreover, their magnitude was relatively stable. For instance, the odds ratio of the variable householder age decreased by only 20% from year 1 to year 5. Similarly, householder age, single non-elderly, log10 annual household income, the square root of benefit income, were consistently significant during the period from years 6 to 10. The models of years 1, 5, and 10 will be analysed in the subsequent section.

Table 5-1.Logistic regression transition models to single non-elderly households in the next year, five and ten years (part1)

| | | (year 1)<br>Single non-elderly | (year 5)<br>Single non-elderly | (year 10)<br>Single non-elderly |
|---|---|---|---|---|
| **Marital status** | Cohabiting couples | 5.173***<br>(1.614) | 4.68**<br>(2.680) | 0.261*<br>(0.187) |
| | Widowed | 3.168<br>(3.117) | 2.887<br>(2.929) | - |
| | Divorced | - | 2.718<br>(1.779) | - |
| **Age and household type** | Householder Age | 0.745***<br>(0.0206) | 0.56***<br>(0.0221) | 0.363***<br>(0.0272) |
| | Single non-elderly | 4.516***<br>(1.835) | 0.149***<br>(0.0584) | 0.421<br>(0.312) |
| | Lone parent | - | 0.0686*<br>(0.0714) | - |
| | Household size | 0.547*<br>(0.167) | - | - |
| **Accommodation related variables** | Terraced house | 2.455*<br>(0.873) | - | - |
| | Purpose built flat or maisonette | - | 0.164***<br>(0.0897) | 0.272<br>(0.210) |
| | Converted flat or maisonette | - | 0.0171***<br>(0.0152) | - |
| | Semi-detached house or Bungalow | - | 0.45* | - |
| | Dwelling size | 0.842*<br>(0.0836) | 0.833*<br>(0.102) | |
| | Owned outright | 1.458<br>(0.519) | - | - |

Table 5-2.Logistic regression transition models to single non-elderly households in the next year, five and ten years (part2)

| | | (year 1) Single non-elderly | (year 5) Single non-elderly | (year 10) Single non-elderly |
|---|---|---|---|---|
| Socio-economic class and income variables | Higher-grade professionals | 0.569 (0.181) | - | 2.461 (1.227) |
| | Lower-grade professionals | - | - | 2.609* (1.144) |
| | Routine non-manual employees | | 0.256** (0.134) | - |
| | Foreman or technicians | 0.379 (0.191) | - | - |
| | Personal service workers | - | 0.0903** (0.0768) | - |
| | Log 10 total income | 0.497 (0.208) | 0.59 (0.28) | 0.124** (0.0847) |
| | Receiving benefit | - | 0.733* (0.258) | - |
| | Square root household annual benefit income | - | - | 1.015* (0.00630) |
| | McFadden's R2 | 0.553 | 0.457 | 0.568 |
| | Type of employed models | Fixed-effects | Fixed-effects | Fixed-effects |
| | Sample size | 1755 | 1382 | 1100 |

Note: standard errors in parentheses *Significance at the 95% level **Significance at the 99% level

## Transition in the next Year

The analysis of the model outlining the transition to single non-elderly in the next year has suggested that the independent variables explained 55.3% of the variation in the dependent factor. In addition to that, the subsequent results are reported. First, for cohabiting couples and single non-elderly households, the odds of remaining single non-elderly is higher by a factor of 5.173 and 4.516, correspondingly. The high odds ratio for cohabiting couples could be attributed to the fact that at least 18% of couples in the UK are prone to separation, according to a recent study on around 21000 UK cohabiting couples (BBC, 2016a). In addition to that, Benson (2013) reported a 5.3% annual break-up rate amongst couples based on their analysis of understanding society panel data. This, in turn, reinforces the undertaken research findings with regards cohabiting couples' transition patterns. On the other hand, the large odd ratio for single non-elderly households was expected since the proportion of the ones making transitions (20% yearly) is far lower than of those remaining after one year, as discussed previously. Apart from that, the odds ratio for being single non-elderly was higher by 2.455 for households living in terraced houses. Again, this could be due to several factors such as income and personal preferences. For instance, a recent English housing survey report suggested that households in poverty, of which 20% are single non-elderly (The poverty site, 2016), are more likely to live in terraced houses. Indeed, managers and professionals, who usually have high levels of income, tend to dwell in detached or semi-detached homes (ONS, 2015c).

Householder age, household size and dwelling size; on the other hand, had a significant negative effect on this transition. This was expected from analysing the characteristics of the initial sample. More specifically, for a 1-year increase in the age of the householder, the odds ratio of being single non-elderly is lower by 25.5%. This implies that a considerable proportion of the initial population is more likely to move to different household structures after one year, especially after knowing that the majority of the sample population (58%) is aged between 26 and 45. Similarly, for an additional member in the household, the odd of becoming single non-elderly in a year time is lower by 45.3%. A possible interpretation of this figure could be related to the fact that female householders with children are very likely to be lone parents with dependent children after divorce or separation (Finch, 2002). Finally, the odds of becoming single non-elderly was lower by 15.8% for an additional bedroom in the dwelling, which suggests a possible relationship between household size and dwelling size.

In contrast to the above, different socio-economic status, housing tenure, and income variables were not significant. However, the odds ratios of certain variables including; foreman technicians and higher-grade professionals, were significant at the 10% level. Thus, not reported in this study.

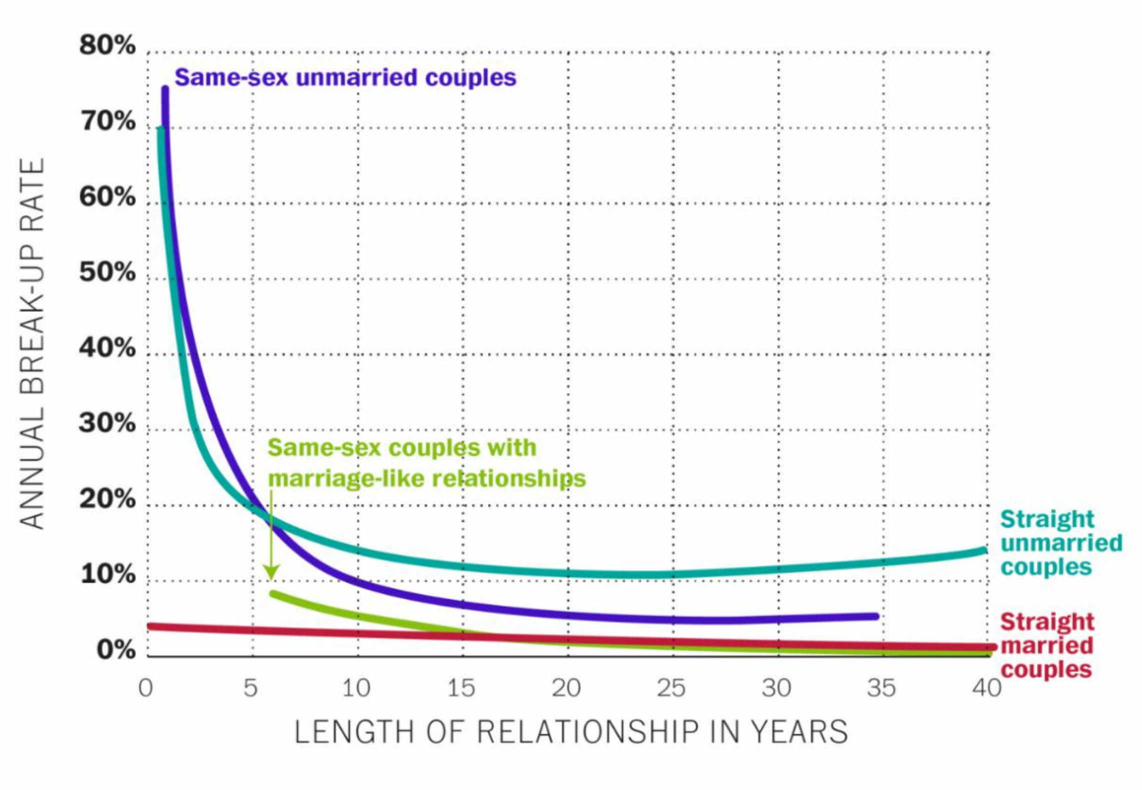



Figure 5-4. The percentage of annual break up rates between couples in function of the length of relationship in years (Rosenfeld, 2014)

## Transition after 5 years

The analysis of the logistic regression model representing single non-elderly transitions after five years has advised that the independent factors are responsible for 45.7% of the variation in the dependent variable. In addition to that, the examination of this model has unravelled the subsequent points.

First, unlike the previous models, only the "cohabiting couples" variable was significantly positive. However, its odds ratio was 2.37 higher than in year 4 transition model. Based on Rosenfeld (2014) findings on the relationship between break ups rate and time (Figure 5-4), the increase in the odds ratio of cohabiting couples becoming single non-elderly could be due to a rise in the percentage of new cohabiting couples from the year 1995 to 1997.

On the other hand, several variables had a significant negative effect on this type of transition. In addition to household age and "single non-elderly" variables, this encompasses other factors such as lone-parent households, dwelling size, purpose built flats, converted flats, and semi-detached dwellings. Moreover, the following socio-economic classes namely; routine non-manual and personal service workers in addition to households receiving benefit. The odds ratios for purpose built and converted flats were 0.164 and 0.017, respectively. These findings were in line with the centre for sustainability energy report which suggested that 48% of UK couples without children dwelled in purpose built and converted flats in 2011. The odds of becoming single non-elderly in five years was lower by 74.4% for routine non-manual classes. Similarly, for personal service workers, the odds of being single non-elderly in 5 years was lower by 91%. A possible interpretation of this figure is related to the fact that more than 40% of households with personal service occupations are aged between 25 and 44, whereas only 19% are between 16 and 25 years (Begum, 2004). This means that a large proportion of this class is more likely to be in different household types other than single non-elderly (e.g. couple with children). Finally, households who were on benefit are less likely to be single non-elderly in 5 years by an odds ratio of 0.733.

In contrast to the above, income and the remaining socio-economic status, marital status, dwelling type, and dwelling size variables were not significant at the 5 and 1% levels.

### Transition after 10 years

The examination of year ten single-non-elderly transition model has shown that the variables included in the model are responsible for 56.8% of the variation in the dependent factor. Furthermore, the model analysis also suggested a similar effect of the variable householder age but with a superior magnitude as the odds ratio is 0.363. Furthermore, it has advised a contradictory effect of the variable cohabiting couples since the odds of being single-non-elderly in the next 10 years was found to be lower by 73.9%.

Apart from that, the odds ratio for moving to a single non-elderly household in 10 years was higher by a factor of 2.609 for lower-grade professionals. This could be owing to a possible divorce occurrence amongst this socio-economic class of which 60% of them tend to get married (Goodman and Greaves, 2010). Indeed, despite the family stability of households with professionals and managerial occupations, a recent UK official report claimed that the average marriage length in the UK ranges between 8.9 and 12.2 years (ONS, 2013e). For those reasons, it is believed that this odds ratio is logical and consistent with existing governmental reports. Similarly, for a one-unit increase in the square root of the household annual benefit income, the odds of being single non-elderly in 10 years is higher by 1.5%. Conversely, the variable household annual labour income was found to have a minor negative significant effect. More precisely, for one-unit increase in the square root of this variable results, the odds ratio for becoming single non-elderly is 0.124. Finally, variables pertaining to the householder socio-economic class, employment mode, and dwelling type, were not to found be statistically significant.

### 5.5.2 Transition to couple without children households

#### General trend

Table 5-3, Table 5-4, Table 5-5, and Table 5-6 illustrate the logistic regression models pertaining to the transitions to couples without children over one, five and ten years. Overall, from analysing these models, it is evident that the type and number of significant variables were not consistent.

From year 1 to year 5, there have been some changes related to the proportion of positively and negatively significant variables. Moreover, the change in the odds ratio of certain variables. These can be reported in two stages. In stage 1 which covers the models of years 1-4, the number of positively significant factors was relatively smaller than the one of the negatively significant variables. Apart from that, the odds ratio of the variable "On pension" decreased by an average of 1.67 annually. Conversely, the odds ratios of the variables namely; householder age and the square root of annual benefit income remained small and relatively stable over those models. Finally, the factors namely; not the main householder, other households, and foreman technicians, were not significant across the four years.

On the contrary, stage 2 (model of year 5) witnessed significant changes in the direction of the association between certain variables and the dependent factors. For example, the odds ratio of the variable "couple with children" was 0.0527 in year 1, whereas in year 5, it was 4.230. This stage was also characterised by the significance of new variables namely; living in purpose-built flats, dwelling in converted flats, and living in semi-detached houses.

From year 6 to 10, there were changes in the magnitude and the direction of the association between the dependent and independent variables. These can be, in turn, reported in three distinct stages. In the first stage, which covers year 6, there was an increase in the number of positively significant variables. As a result, there is was a balance with the negatively significant ones. Apart from that, the direction of relationship between the variable "never married" and the dependent factor changed from being negative to positive. This phase also witnessed the significance of a new variable namely; lower-grade professionals. However, other variables such as routine non-manual and on pension were found insignificant. Finally, the magnitude and direction of some factors such as high-grade professionals and foreman technicians were consistent with the previous models. As for phase 2, which covers the models of years 7-8, the odds ratio of certain variables fluctuated. For instance, the odds ratio of the variable "couple with children" increased by 3.23 in year 7 and then decreased by 1.78 in the following model. Apart from that, the variables namely; higher-grade professionals and semi-unskilled manual workers remained consistent in comparison to the previous phase. Conversely, phase 3, which encompasses the models of years 9-10, has known the insignificance of marital status, dwelling type, and tenure mode variables.

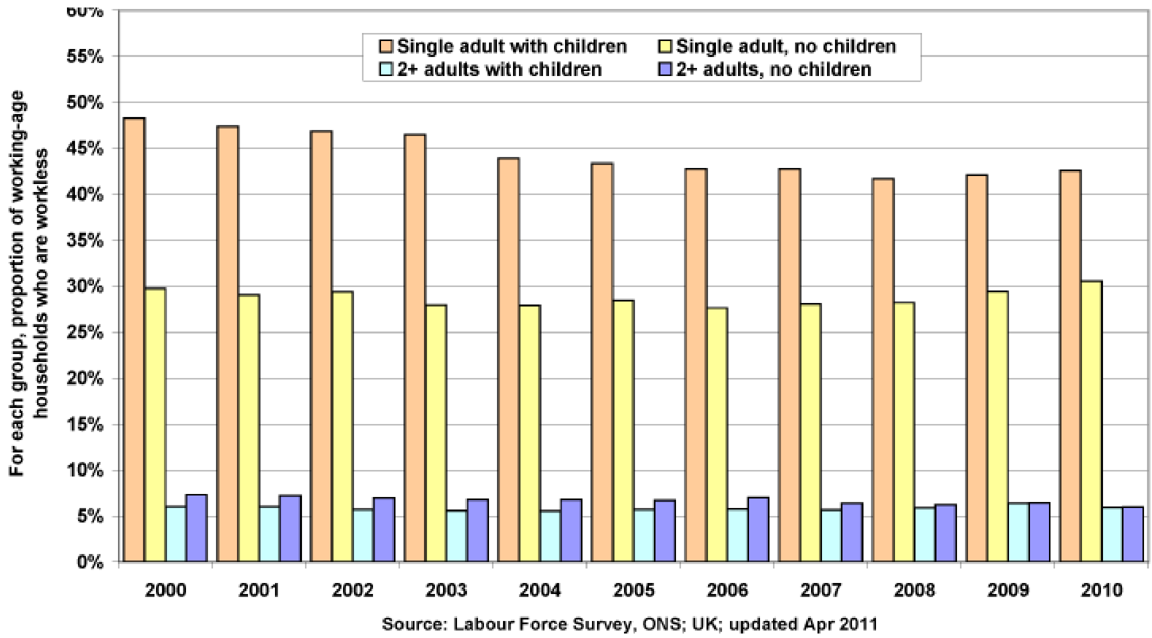



Figure 5-5. The proportion of workless adults across different UK household types (ONS, 2015d)

Table 5-3. Logistic regression transition models to couple without children households in the next year, five and ten years (part1)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Couple No | Couple No | Couple No |
| Household type | Couple with children | 0.0527** | 4.230** | 2.891 |
| | | (0.0158) | (1.646) | (2.264) |
| | Lone parent | - | - | - |
| | | | | |
| | Other households | - | - | - |
| Marital status and age | | | | |
| | Divorced | 0.0389** | - | - |
| | | (0.0169) | | |
| | Widowed | 0.0183** | - | - |
| | | (0.0164) | | |
| | Separated | 0.0690** | 6.343* | 0.4158 |
| | | (0.0319) | (4.667) | (0.3825) |
| | Never Married | 0.0359** | 1.986 | 0.900 |
| | | (0.0118) | (0.977) | (0.8798) |
| | Not main householder | 1.792* | - | - |
| | | (0.458) | | |
| | Age of the householder | 1.106** | 1.1** | 1.551** |
| | | (0.0274) | (0.0285) | (0.0767) |
| **McFadden's R2** | | 0.476 | 0.22 | 0.150 |
| **Type of employed models** | | Fixed effects | Fixed effects | Fixed effects |
| **Sample size** | | 1846 | 1251 | 662 |

Note: Couple No= couple without children / standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table 5-4.Logistic regression transition models to couple without children households in the next year, five and ten years (part2)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Couple No | Couple No | Couple No |
| Dwelling size and type | Dwelling owned with mortgage | 0.484** | 0.517 | 0.551 |
| | | (0.131) | (0.179) | (0.239) |
| | Renting from local authorities | 0.335 | 0.129* | - |
| | | (0.239) | (0.118) | |
| | Renting from housing associations | 0.171 | - | - |
| | | (0.190) | | |
| | Renting from private landlords | - | 0.317* | - |
| | | | (0.197) | |
| | Dwelling size | 1.060 | - | - |
| | | (0.0916) | | |
| | Purpose built flat | - | 7.456** | 0.7568 |
| | | | (4.214) | (0.7241) |
| | Converted flat | - | 8.086** | 1.6932 |
| | | | (5.987) | (2.1283) |
| | Semi-detached house or bungalow | - | 1.856* | - |
| | | | (0.573) | |

Note: Couple No= couple without children / standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table 5-5.Logistic regression transition models to couple without children households in the next year, five and ten years (part3)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Couple No | Couple No | Couple No |
| Socio-economic class variables | Higher-grade professionals | 1.126 | 0.455* | 0.304** |
| | | (0.384) | (0.164) | (0.135) |
| | Lower-grade professionals | 1.528 | 0.560 | 0.269** |
| | | (0.470) | (0.187) | (0.119) |
| | Routine Non-manual | 1.576 | 0.431* | 0.491 |
| | | (0.617) | (0.198) | (0.286) |
| | Foremen technician | 2.400 | 0.384** | - |
| | | (1.173) | (0.184) | |
| | Small proprietors with employees | - | - | 0.105 |
| | | | | (0.126) |
| | Skilled manual workers | - | 2.207 | - |
| | | | (1.273) | |
| | Semi-skilled manual workers | - | 0.227* | 0.450 |
| | | | (0.132) | (0.331) |

Note: Couple No= couple without children / standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table 5-6.Logistic regression transition models to couple without children households in the next year, five and ten years (part4)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Couple No | Couple No | Couple No |
| Income related variables | On Pension | 7.922** | 4.254** | 1.7955 |
| | | (3.710) | (2.292) | (1.6206) |
| | On Benefit | 0.506* | - | 1.362 |
| | | (0.141) | | (0.5998) |
| | Square root of total household annual income | 1.022* | 1.005 | - |
| | | (0.00961) | (0.00329) | |
| | Square root of household annual benefit income | 0.978** | 0.988** | 0.997* |
| | | (0.00596) | (0.00407) | (0.00456) |
| | Square root of household labour income | 0.985 | - | - |
| | | (0.00779) | | |
| | Working Full-time | - | 0.459* | 1.4626 |
| | | | (0.151) | (0.6469) |

Note: Couple No= couple without children / standard errors in parentheses

*Significance at the 95% level **Significance at the 99% level

## Transition after 1 year

Overall, the analysis of the model representing the transition to couples without children in the next year has advised that the independent factors cause the dependent variable to variate by 47.6%. Apart from that, the analysis of this model suggested the following points.

First, it is evident that the majority of variables had a negative impact on this transition. In particular, for households with a never married marital status, the odds of moving to couples no children was lower by 96.41%. This result was excepted since only 8.1% of single non-elderly in the BHPS dataset moved to couple without children in 1992. Similarly, the odds were lower by 96.11%, 98.17%, and 93.1% for divorced, widowed, and separated households, correspondingly. These findings align with a recent study from the insurance company Aviva (AVIVA, 2014), which claimed that 36% of the divorced and separated respondents preferred to stay single in the future. This argument was influenced by the high hidden costs of divorce or separation. Moreover, the time required to settle which can be up to 2 years. Apart from that, the reason behind widowers preferring to stay single no-elderly could be associated with the ageing of this population over the period from 1981 to 2011 according to a recent official UK report (ONS, 2014c). In addition to that, as expected, for couples with children, the odds ratio for becoming couples without children after a year was lower by 94.73%. The variable dwelling owned with mortgage also had a significant negative impact as its odds ratio was lower by a factor of 0.484. A possible interpretation of those findings is based on the fact that 53% of couples without children are more likely to own their properties outright, whereas only 34% of them prefer to own their dwellings with a mortgage. Finally, for households receiving benefits, the odds ratio for being couples without children in one year is lower by 49.4%. However, the impact of square root annual benefit income on this type of transition was negligible since the odds ratio for one additional unit of this variable was lower by 2.2%. The low odds ratio for "receiving benefit" variable is directly related to the fact that couples without children are amongst the least likely to be workless, as illustrated in Figure 5-5 (above). Thus, they are less entitled to benefits than those with children according to the institute for financial studies (IFS, 2015).

In contrast to the above, variables namely; not the main householder, householder age, on pension, and log10 total household annual income had a positive effect on this type of transition. More precisely, the odds ratio was 10.6% higher for one additional year in the householder age. This could be associated with the age of the initial population of which 58% are aged between 26 and 45. Indeed, recent UK official reports suggest that the average age of couples with children at the first child birth is 31.6 years (ONS, 2015a). Similarly, for household members other than the main householder, the odds ratio for becoming couples without children is higher by a factor of 1.792. After a closer inspection of this unexpected result in the BHPS dataset, it was found that over 18 years, 46% of household members other than the main householder became couples without children. Furthermore, around 70% of those are females. Apart from that, households on pension are more likely to be couples without children in the next year. A possible interpretation of this outcome could be linked to the fact that 48 % of the total UK pensioners are couples without children according to the family resource survey (ONS, 2012). Finally, the log 10 of gross annual household income was significant, but its effect was weak as the odds ratio for being couples without children was higher by 2.2% for a one-unit increase in this variable.

Finally, all variables on the dwelling type and householders' socio-economic status were not found significant.

### Transition after 5 years

The examination of the logistic regression model of transition to couples no children after five years has shown that the independent factors are responsible for 22% of the variation in the dependent variable. In addition to that, the analysis of this transition model has advised fundamental changes in the direction of the association between certain independent variables and the dependent factor.

First, the odds ratios of the variables couples with children and separated households were 4.230 and 6.343, respectively. These results are reasonable because there was an increase in the odds ratios of these variables over the four previous models. Indeed, a recent report on divorces in England and Wales claimed that the average length of separation before the divorce or reuniting is between 2 and 3 Years (ONS, 2013b). Similar findings were also reported in the US by Bramlett and Mosher (2012). This, in turn, suggests that the time needed to embark on a new relationship, especially after divorce, might be slightly beyond three years, which is in line with the model's odds ratio for separated households. Apart from the above variables, the odds ratios for being couples without children in 5 years were 7.456 and 8.086, for households living in purpose built and converted flats, correspondingly. This could be attributed to the fact that over 50% of the UK cohabiting couples prefer to live together at least 5 years before the first child birth, hence prefer to remain in flats (Goodman and Greaves, 2010). Similarly, the odds ratio was also higher by a factor of 1.856 for households dwelling in semi-detached houses. This finding is considered valid since these dwelling types are very popular to middle-income households, who are mostly couple without and with children (ONS, 2015b; WhatPrice, 2017). The age of householder did not only have a significant positive effect, but also its odds ratio (1.1) was consistent with previous models. After a closer inspection of this result using an age groups variable instead, it was found that for householders aged 26 and 36, the odds ratio for becoming couples no children was higher by a factor of 1.745. Indeed, this aligns with the Births by Parents' Characteristics in England and Wales report which suggests that the average age of couples with children at the first birth was 31.6 years (ONS, 2015a).

In contrast to the above, the odds ratios were lower by 54.5%, 56.9%, 61.6%, and 77.2%, for the variables namely; higher-grade professional, routine non-manual, foreman/ technicians, and semi-unskilled manual workers, respectively. This could be because roughly 45 % of households from these socio-economic classes prefer to have at least 1-2 children given their financial stability (Whiting, 2010). In addition to that, for households working full-time, the odds ratio for being couples without children in 5 years was lower by 54.1%. This, in turn, could be interpreted in two distinct ways. The first argument advises that households with full-time jobs have more income stability (JRF, 2005). As a result of that, they are more willing and prepared to have children. Conversely, the second possible interpretation might be related to the fact that some future parents prefer to work part-time to avoid paying the prohibitive annual childcare cost of £11,000 on average according to NCT, (2014). Indeed, around 55% of women with dependent children are in part-time employment (ONS, 2014b). As expected, the odds ratios of variables related to tenure mode namely; renting from local authorities and renting from private landlords, were 0.129 and 0.317, correspondingly. These findings were in good agreement with the UK recent official report on home ownership and renting (ONS, 2013d). Finally, the square root of total annual benefit income was significant but had a minor effect, whereas the log10 of household annual gross income was insignificant.

Other insignificant variables include; other households, lone parent, widowed, divorced, never married, terraced, dwelling size, on benefit, farmers/small holders, small proprietors with and without employees, and lower-grade professionals.

### Transition after 10 years

In comparison to the above model, the examination of the model representing the transition possibilities to couple without children households in the future ten years has advised that the independent factors cause the dependent variable to variate by only 15%. In addition to that, this analysis also suggested the insignificance of variables namely; never married and semi-unskilled manual workers. This, in turn, indicates that the employed dataset could not capture the transitions of never married households as well as semi-unskilled manual workers beyond 10 years.

On the other hand, for one additional year on the householder age, the odds ratio for being couples without children after 10 years was higher by 55%. Conversely, for higher and lower-grade professionals, the odds ratios for becoming couples without children in 10 years were lower by 69.6% and 73.1%, respectively. The odds ratio of the square root total benefit income; however, was relatively stable (lower by 0.3%) in comparison to the previous models.

Finally, variables namely; On pension, working full-time, living in purpose built flats, living in converted flats, separated, divorced, widowed, owned with a mortgage, and renting from a private landlord, and renting from local authorities, were not statistically significant.

### 5.5.3 Transition to lone parent households

#### General trend

Table 5-7, Table 5-8, Table 5-9, and Table 5-10 represent the logistic regression models of the transition to a lone parent household in one, five, and ten years. Again, the significance, magnitude and the direction of association of the models' independent variables were not consistent.

During the period from year 1 to year 5, marital status, gender, age, socio-economic, dwelling type, and tenure mode variables, were found significant. In particular, the changes in the odds ratios and direction of these variables over this period could be summarised in four distinct stages. First, stage 1, which entails the model of year 1, was characterised by the weak effect of household type variables. For example, the odds ratio for being lone parents in the next year was lower by 99.37% for couples without children. In addition to that, the variables namely; female, divorced, aged 46-55, household size, renting from a private landlord, routine non-manual, foreman technicians, skilled manual workers, and the square root of annual benefit income had a significant positive effect. However, in stage 2, which encompasses the model of year 2, witnessed the change in the odds ratios of many variables. For instance, there was a decrease of 1.84 and 5.185 in the odds ratios of divorced and household size, correspondingly. Apart from that, the variable "living in an end-terraced dwelling" was significant. In stage 3, which includes the model of year 3, the majority of household types variables, except "couple with children", were not significant. Apart from that, the following factors; Living in terraced houses, Farmers- small holders and renting from local authorities, were found statistically significant. Finally, the odds ratios of the variables namely; living in terraced houses and aged 46-55 decreased by 53% on average. On the other hand, phase 4, which entails the fourth and fifth years of transition, was marked by the insignificance of many variables. For example, in year 4, the following variables namely; rented from local authorities, farmers smallholders, working full-time, and age variables, were insignificant. This phase has also witnessed the significance of the factor "higher-grade professionals". Moreover, a dramatic increase in the odds ratio of the variable "divorced" which went from 4.334 in year 4 to 27.83 in year 5. The reason behind this surprising rise could be because couples divorcing at the early years of their marriage are less likely to have children (ONS, 2013b). As a result, they are unlikely to move to a lone parent household.

In contrast to the above, the changes from year 6 to year 10 could be reported in two different stages. First, in phase 1, which involves the models of years 6 and 7, new variables were found significant in comparison to the models of years 1-5. For instance, the factors namely; aged 26-35 and dwelling size were significant in year 6. However, other factors such as living in terraced houses were insignificant. In addition to that, this phase has also known changes in the odds ratios of certain variables. For instance, there was a 2.209 increase in the odds ratio of the variable "single non-elderly" in year 6. In stage 2, which covers the models of years 8, 9, and 10 witnessed the significance of new variables namely; skilled manual workers, working full-time, small proprietors without employees, aged 36-45 and aged 46-55. Furthermore, unlike the previous phase, many variables such as female, higher-grade professionals, skilled manual-workers and working full-time, were consistent in terms of significance and magnitude. Finally, the factors namely; on pension, rented from local authorities, and small proprietors with employees were inconsistent during this phase. For example, the variable "on a pension" was significant in the model of year 9 but insignificant in the models of year 7 and 8.

Table 5-7.Logistic regression transition models to lone parents in the next year, five and ten years (part1)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Lone parent | Lone parent | Lone parent |
| Household type and marital status variables | Single non-elderly | 0.130** | 3.736* | - |
| | | (0.0691) | (2.059) | |
| | Couple without children | 0.00626** | - | - |
| | | (0.00792) | | |
| | Couple with children | 0.000639** | - | - |
| | | (0.000729) | | |
| | Other households | 0.00888** | - | - |
| | | (0.0123) | | |
| | Female | 3.558* | 4.542* | 3.261* |
| | | (2.184) | (3.160) | (2.826) |
| | Divorced | 4.415* | 27.83** | - |
| | | (3.206) | (18.98) | |
| | Separated | 3.739 | - | - |
| | | (4.257) | | |
| **McKelvey & Zavoina's R2** | | 0.3032 | 0.181 | 0.125 |
| **Type of the used model** | | Random effect | Random effect | Random effect |
| **Sample size** | | 5551 | 3670 | 2229 |

Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table 5-8.Logistic regression transition models to lone parents in the next year, five and ten years (part2)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Lone parent | Lone parent | Lone parent |
| Household age and size variables | Aged 26-35 | - | 1.727 | - |
| | | | (0.948) | |
| | Aged 36-45 | 1.833 | 0.7152 | 0.359* |
| | | (0.926) | (0.5568) | (0.481) |
| | Aged 46-55 | 5.615** | 1.13523 | 0.260* |
| | | (2.969) | (0.61640) | (0.208) |
| | Household size | 7.733** | 2.93** | - |
| | | (2.398) | (0.612) | |
| Dwelling related variables | Renting from private landlords | 7.455** | - | - |
| | | (4.558) | | |
| | Renting from local authorities | - | 0.115** | 0.412 |
| | | | (0.0875) | (0.400) |
| | Living in end-terraced dwellings | - | 2.968 | 2.011 |
| | | | (2.252) | (1.680) |
| | Living in terraced dwellings | - | 4.408* | - |
| | | | (2.557) | |
| | Living in purpose built flats | - | 6.684* | - |
| | | | (4.981) | |

Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table 5-9.Logistic regression transition models to lone parents in the next year, five and ten years (part3)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Lone parent | Lone parent | Lone parent |
| Socio-economic class variables | Higher-grade professionals | - | 0.0701* | 0.187* |
| | | | (0.0759) | (0.138) |
| | Routine non-manual employees | 3.903* | 0.263 | - |
| | | (2.550) | (0.232) | |
| | Small proprietors with employees | 2.093 | | 51.33** |
| | | (2.150) | | (62.04) |
| | Small proprietors without employees | 0.102 | | 10.96* |
| | | (0.132) | | (12.00) |
| | Farmers/small holders | 0.0893 | 0.0244* | - |
| | | (0.146) | (0.0411) | |
| | Foreman technicians | 7.191* | 3.009* | - |
| | | (7.217) | (2.416) | |
| | Skilled manual workers | 6.047* | | 8.692* |
| | | (5.288) | | (9.392) |
| | Semi-unskilled manual workers | 7.136* | - | - |
| | | (5.498) | | |

Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table 5-10.Logistic regression transition models to lone parents in the next year, five and ten years (part4)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Lone parent | Lone parent | Lone parent |
| Income related variables | On Pension | 0.186** | 0.267* | - |
| | | (0.117) | (0.216) | |
| | Square root of total annual benefit income | 1.023** | 1.0134* | - |
| | | (0.00723) | (0.006863) | |
| | Square root of annual investment income | 0.991 | 0.984 | - |
| | | (0.00856) | (0.00932) | |
| | Working full-time | 1.380 | 1.2363 | 0.559* |
| | | (0.671) | (0.644) | (0.323) |

Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

## Transition after 1 Year

The analysis of the logistic regression model representing transition probabilities to lone parent households in one year has advised that the impendent variables explained 30.32% of the total variation of the dependent factor. Apart from that, the subsequent points were concluded.

First, although they were significant, the impact of household type variables on the dependent variable was very small. More specifically, for single non-elderly households, the odds ratio for becoming lone parents in one year was lower by a factor of 0.130. Similarly, the odds ratios were 0.00626 and 0.000639 for couples without children and with children, correspondingly. Also, the variable "Other households" had an odds ratio of 0.00888. Apart from that, for households on pension, the odds for being lone parents in the next year was lower by 81.4%. This aligns with the findings of couples no children transition models, where 48% of the UK pensioners were couples without children (ONS, 2012). Another evidence is based on the fact that lone parent households had the lowest pension wealth in comparison to other household types, such as couple without children (ONS, 2015b).

On the other hand, marital status, age, and socio-economic variables had a considerable positive effect on the dependent variable. In particular, for female householders, the odds ratio for becoming a lone parent household in the coming year was higher by a factor of 3.558. Indeed, lone female parents accounted for approximately 91% of total lone parent households (ONS, 2014b). Divorced householders were also very likely to be in a lone parent household as the odds ratio of this variable was 4.415. This was consistent with the government report on lone parents, suggesting that roughly 30% of lone parents are divorced (ONS, 2015f). Even though the continuous age variable was not found significant, householders aged between 46 and 55 years were more likely to be lone parents next year with an odds ratio of 5.615. This could be attributed to the fact that the around 50% of all the UK divorces were for householders aged between 40 and 55 (ONS, 2013b). In addition to that, for one additional member in the household, the odds ratio for being a lone parent household in the coming year was higher by a factor of 7.733. This could imply that lone parent households with more dependent children are less likely to move to other household types. Households renting their accommodations from private landlords for one year are likely to be lone parents with an odds ratio of 7.733. Although this is not the most common tenure mode for this household type, some of them, who cannot access social housing or are awaiting decisions on their applications, can still benefit from housing benefit schemes while living in private accommodations (GOVUK, 2017c). Indeed, almost 65% of lone parents in the UK rent their accommodations (ONS, 2011b).

Apart from that, the odds ratios were 3.903, 7.19, 6.047, and 7.136 for the following socio-economic status namely; routine non-manual, foreman technicians, skilled manual workers, and semi-unskilled manual workers, correspondingly. According to the UK 2011 census report (ONS, 2011c), around 20%, 10%, and 25% of lone parents belonged to the subsequent classes; routine occupations, lower-supervisory and technical occupations, and semi-routine occupations, respectively. Finally, the square root of total benefit income was significant but with a weak effect since the odds ratio was higher by 2.3%.

In contrast to the above, dwelling type variables in addition to the remaining tenure mode and income variables were not found statistically significant.

### Transition after 5 years

The investigation of the model representing the transition possibilities to a lone parent household in the next five years has indicated that the independent factors cause the dependent factor to variate by only 18.1%. Besides, the below conclusions have been made.

First, like in the previous model, most of the statistically significant factors had a positive impact on this transition type. More precisely, for the following variables namely; single non-elderly, females, household living in terraced and purpose-built flats, the odds ratio were 3.735, 4.542, 4.408, and 6.84, correspondingly. The same applies to the following continuous variables namely; the square root benefit income and household size, in which a one-unit increase in these variables will contribute to a rise of 1.0134 and 2.93 in their odds ratios, respectively. However, despite having a positive effect, the odds ratio for divorced was unexpectedly very high (27.83%). This could be attributed to the fact that almost 48% of divorcees in the UK had at least one child under 16 in the household (ONS, 2013b).

On the other hand, the odds ratios were lower by 0.267, 0.0244, 0.0701, 0.115 for households on a pension, farmer small holders, higher-grade professionals, and the ones renting from local authorities, respectively. The income stability of this socio-economic class in the main reason behind these very low odds ratios. Indeed, around

94% of them are divided between couples with children and couples without children, including empty nesters (ONS, 2013c).

Finally, the variables; end-terraced, age variables, household type factors, in addition to the rest of tenure mode and income variables, were not found significant in this model.

### Transition after 10 years

The inspection of the regression model representing the possibility of becoming lone parents after ten years has advised that the model independent factors are responsible for only 12.5% of the variation in the dependent variable. Apart from that, the below points were reported.

First, it is clear that the direction of association of most variables with the dependent variable was consistent with the precedent models. However, there was an exception for the variables; on a pension and renting from local authorities, as they were not found statistically significant. Variables with a positive impact on the dependent factor, including female, small proprietors with employees, and skilled manual workers, had the following odds ratios; 3.261, 51.33, and 8.692, respectively. In addition to that, for households belonging to the socio-economic class "small proprietors without employees", the odds ratio for being a lone parent family in the next 10 years was higher by a factor of 10.96. This might be attributed to the fact that a lot of lone parents consider setting up their businesses to have more time flexibility; consequently, dedicate more time to their children. Indeed, a recent report from CitizenAdvice (2015) suggested that a large proportion of those who are on part-time self-employment were women. Furthermore, among those, a considerable proportion is aged 45 or older, which aligns with the findings suggesting that most divorces occur at an age between 40 and 55 and with at least one child.

In contrast the above, the odds ratios for the negatively significant variables namely; working full-time, higher-grade professionals, aged 36-45, and aged 46-55 were 0.559, 0.187, 0.359, and 0.260, respectively.

### 5.5.4 Transition to couple with children households

#### General trend

Table 5-11,Table 5-12, Table 5-13, and Table 5-14 depict the logistic regression models outlining the transitions to couples with children in the next year, five and ten years. Apart from the inconsistency of the independent variables' significance, magnitude, and direction of association, which was shared across all transition models, the subsequent changes are reported.

During the period from year 1 to year 5, the effect of marital status and income variables was minimal. Apart from that, other changes in these models could be further summarised in 3 phases. First, stage 1, which covers the model of year 1, was characterised by a large number of negatively significant variables which was almost the double of the positive ones. Conversely, in stage 2, which entails the models of year 2, 3, and 4, witnessed the significance of some variables as well as the insignificance of others while certain factors remained constant. For instance, in the model of year 2, the variables namely; on benefits, lone parents, higher-grade professionals, and routine non-manual were found significant. On the other hand, the factors; living in semi-detached houses, farmers/smallholders, and agriculture workers, were significant in the model of year 3. However, the majority of significant factors in the models of years 2-3 were found insignificant in year 4. Unlike phase 1, stage 3, which entails the model of year 5, was marked by the dominance of positively significant factors such as aged 46-55 and living in detached dwellings. In addition to that, the odds ratio of the variables namely; unskilled manual workers and lower-grade professionals doubled in comparison to the previous stage. However, the odds ratios of some variables such as other households and aged 46-55 remained consistent while other factors were not found significant. These include the following; farmers small holders, living in purpose-built flats, living in converted flats, and rented from local authorities.

In contrast to the above, the changes between year 6 and year 10 could be analysed in five distinct stages. In stage 1, which covers the model of year 6, the majority of independent factors had a positive effect on the dependent variable. In addition to that, some variables related to the marital status, tenure mode, dwelling type, and income were insignificant. However, the factors namely; on benefits and aged 36-45 were found significant in this phase. Stage2, which covers the models of year 7, was marked by the positive impact of the variable "divorced". As discussed in the section pertaining to this model, this might be because 1 in 7 marriages in the UK are for those remarrying after divorce (ONS, 2013e). Similarly, the variables; aged 36-45, aged 46-55, owned outright, higher-grade professionals, on benefit, and working full-time were consistent with the previous phase. As for stage3, which represents the model of year 8, some variables with a negative effect in the precedent phase were insignificant. This is the case for; dwelling owned outright, working full-time, and skilled-manual workers. As a result, the number of positively significant variables was equal to the one of the negative factors. On the other hand, in stage 4 which encompasses the model of year 9, involved only positively significant variables. However, in stage 5, which represents the model of year 10, there was a balance between negative and positive variables.

Table 5-11.Logistic regression transition models to couple with children households in the next year, five and ten years (part1)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Couple CH | Couple CH | Couple CH |
| Household type, age, and marital status | Lone parent | - | 4.212 | - |
| | | | (4.273) | |
| | Other households | 0.130* | 7.813* | |
| | | (0.132) | (7.550) | |
| | Aged 26-35 | 0.238** | 0.0884** | |
| | | (0.0578) | (0.0387) | |
| | Aged 36-55 | - | - | 0.437* |
| | | | | (0.147) |
| | Aged 46-55 | 2.416* | 5.556** | |
| | | (0.879) | (1.972) | |
| | Cohabiting couples | 0.131** | 0.148** | 0.0636** |
| | | (0.0481) | (0.0789) | (0.0567) |
| | Divorced | 0.00252** | 0.0925* | - |
| | | (0.00227) | (0.0869) | |
| | Never Married | 0.00660** | 0.259 | 0.0620* |
| | | (0.00361) | (0.226) | (0.0700) |
| | Not main householder | 4.145** | 1.973 | - |
| | | (1.429) | (0.812) | |
| **McFadden's R2** | | 0.554 | 0.3 | 0.165 |
| **Type of model** | | Fixed effects | Fixed effects | Fixed effects |
| **Sample size** | | 1337 | 781 | 522 |

Couple CH: couples with children / Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table 5-12.Logistic regression transition models to couple with children households in the next year, five and ten years (part2)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Couple CH | Couple CH | Couple CH |
| Accommodation related variables | Living in detached dwellings | 2.844* | 2.601* | - |
| | | (1.284) | (1.305) | |
| | Living is semi-detached dwellings | 1.801 | 3.025* | - |
| | | (0.733) | (1.346) | |
| | Living in Terraced dwellings | 0.459* | - | 3.055 |
| | | (0.216) | | (2.205) |
| | Living in Purpose built flats | 0.130** | 0.403 | - |
| | | (0.0735) | (0.309) | |
| | Living in converted flats | 0.0679* | - | - |
| | | (0.0717) | | |
| | Dwelling size | 1.866** | - | - |
| | | (0.2658) | | |
| | Dwelling owned outright | 8.214** | 3.225* | - |
| | | (4.170) | (1.852) | |
| | Dwelling Rented from employer | 0.0371** | - | - |
| | | (0.0443) | | |
| | Dwelling rented from private landlords | 0.245* | | 0.0769* |
| | | (0.146) | | (0.0789) |

Couple CH: couples with children / Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table 5-13.Logistic regression transition models to couple with children households in the next year, five and ten years (part3)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Couple CH | Couple CH | Couple CH |
| Socio-economic class related factors | Higher-grade professionals | 0.524 | 3.817** | 7.721** |
| | | (0.187) | (1.749) | (5.044) |
| | Lower-grade professionals | 0.397** | 10.18** | 3.799* |
| | | (0.126) | (4.969) | (2.101) |
| | Routine non-manual employees | - | 7.44** | - |
| | | | (3.56) | |
| | Small proprietors without employees | - | 4.656* | - |
| | | | (3.586) | |
| | Semi-unskilled manual workers | 0.310* | 4.096* | - |
| | | (0.164) | (2.817) | |
| | Foreman technicians | 0.242** | - | - |
| | | (0.122) | | |
| | Farmers/smallholders | 0.305 | 2.482 | - |
| | | (0.283) | (2.159) | |
| | Agriculture workers | 0.0920 | - | - |
| | | (0.173) | | |

Couple CH: couples with children / Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table 5-14.Logistic regression transition models to couple with children households in the next year, five and ten years (part4)

| | | (Year 1) | (Year 5) | (Year 10) |
|---|---|---|---|---|
| | | Couple CH | Couple CH | Couple CH |
| Income related variables | On benefit | - | 1.394 | 1.053 |
| | | | (0.472) | (0.539) |
| | On pension | 0.650 | - | - |
| | | (0.482) | | |
| | Square root of annual benefit income | 1.024*** | 1.021** | 1.026** |
| | | (0.00592) | (0.00693) | (0.00827) |
| | Square root of annual investment income | 0.973*** | 0.980* | - |
| | | (0.00757) | (0.00772) | |
| | Working full-time | - | 2.191* | 1.151 |
| | | | (0.851) | (0.519) |

Couple CH: couples with children / Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

## Transition after 1 year

From analysing the regression model pertaining to household transitions to couples with children in the coming year, it is clear that the independent factors explained 55.4% of the total variation in the dependent variable. Moreover, it is evident that the number of variables with negative effect was the double of the ones with positive impact. Furthermore, income and marital status related variables had a minimal effect, since their odds ratios were minimal.

First, for other household structures, including 2 or more unrelated adults, the odds of being couples with children in the next year was lower by 87%. This was in good agreement with the findings of couple without children transition models. Furthermore, the work of Stone et al. (2011) and Daly (2005), advising that individuals living in non-family arrangements are more likely to become couples without children. Householders aged between 26 and 35 were also found unlikely to be couples with children in one year given the low odds ratio (0.238). This is because the average age of both parents at first child birth in the UK is approximately 31.6 (ONS, 2015a). Similarly, for cohabiting couples, divorced, and never married, the odds ratios were lower by 0.131, 0.00252, and 0.0067, respectively. The lower odds ratio for cohabiting couples could be associated with the fact that only 19% of this category had children according to (ONS, 2013c). On the other hand, the lower odds ratios for never married and divorced were expected since householders with these marital statuses frequently move to couples without children and lone parents, correspondingly, as indicated in the previous models. The same applies to households living in terraced dwellings, purpose built, and converted flats, in which the odds ratios were 0.459, 0.130, and 0.0679, correspondingly. Again, these results clearly demonstrate their consistency with the findings of other transition models. For instance, lone parents and single non-elderly households were more likely to reside in terraced dwellings, whereas couples without children, preferred to live in purpose built or converted flats. In addition to that, the odds ratios were lower by 75.5% and 96.29% for those who lived in privately rented accommodations and dwellings provided by employers, respectively. Premised on the other regression models, it was logical that households residing in private accommodations are unlikely to be couples with children. Similarly, given the fact that individuals living in employers provided accommodations are usually nurses, security agents, carers, and are usually working night shifts, it could be argued that it is quite challenging for them to have children especially at the beginning of their career. Indeed, a recent study conducted by the Berlin social science centre claimed that children of non-standard workers (e.g. weekend and nights) are more likely to have behavioural problems which in turn could potentially damage their development (Li et al., 2014). Finally, as expected given this short time prediction (1 year), the odds ratio for lower-grade professionals and semi-unskilled manual workers were lower by 0.397 and 0.310, respectively.

In contrast to the above, for householders aged between 46 and 55, the odds ratio was higher by a factor of 2.416. Considering that the average age of both parents at the first child birth, this outcome was in line with the recent UK government report suggesting that almost 50% of those aged between 20 and 24 lived with their parents in 2015 (ONS, 2016). Additionally, for household members other than the main householder, the odds for being couples with children in the coming year was high by 4.145. This could be associated with the fact that almost half of individuals other than the main householder in the BHPS dataset were mostly females and moved to couple without children households, as indicated at an earlier stage. Apart from that, certain dwelling type and tenure mode variables had also a positive effect on the dependent variable. More precisely, the odds ratios were higher by 2.844 and 8.214 for households living in detached dwellings and possessing their accommodations outright, respectively. The former result was in good agreement with the UK home ownership report suggesting that half of outright homes were occupied by at least two people (ONS, 2013d). Conversely, a possible interpretation of the latter finding could be related to the fact that detached dwellings are the second most common type for couples with dependent children after semi-detached (ONS, 2013d). Finally, income variables namely; the square root of benefit income and log10 of annual investment income, were significant. However, their effect was negligible since their odds ratios were 1.024 and 0.973, respectively.

### Transition after 5 years

The examination of the regression model representing the transition possibilities to couples with children in the next five years has suggested that the independent factors cause the dependent factor to variate by 30%. In addition to that, it is clear that the majority of significant variables had a positive impact on the dependent factor. Moreover, it has advised certain changes, which include the significance of new variables namely; semi-unskilled manual workers and working full-time. What is more, the variables namely; farmers/smallholders, purpose built flat, converted flat, and rented from local authorities, were found insignificant.

In contrast to the above, the remaining variables were consistent with the previous models, despite some changes in their odds ratios. For examples, for households living in non-family arrangements, the odds for moving to a couple with children family was higher by a factor of 7.813. Similarly, the odds ratios were 5.556, 2.601, 3.025, and 3.225 for the variables; aged 46-55, living in detached houses, living in semi-detached dwellings, and owned outright, correspondingly. Similarly, socio-economic status variables had also a significant effect. However, the odds ratio for lower-grade professionals almost doubled from the model of year 4 to reach 10.18. This implies that this socio-economic class is more likely to become couples with children with time. In this respect, it could also be argued that there should be a possible relationship between savings and planning to have children. Indeed, a recent report by the insurance company Aviva (2013) claimed that couples who plan to have children are more likely to be on a saving plan, and around 31% of them save more than 2000 £ annually. Apart from that, for households working full-time and farmers smallholders, the odds ratios were 2.191 and 4.096, respectively. This was in line with the UK report on families in the labour market suggesting that around 60% of couples with children were in full-time employment (ONS, 2014b).

On the other hand, like on the previous models, the effect of marital status variables was very weak. For instance, for divorced and never married householders, the odds for becoming couples with children in 5 years were lower by 0.0925 and 0.259, correspondingly. Similarly, for a one-unit increase in the square root of investment income, the odds ratio for being a couple with children in 5 years was lower by 2%. Finally, the odds ratio of householders aged between 26 and 35; however, was slightly higher with a value of 0.158.

### Transition after 10 years

The analysis of the model representing the transition probabilities to couples with children in the next years has advised that the independent factors are responsible for 16.5% of the total variation in the dependent variable. Moreover, it was noticed that the number of variables with a positive effect was nearly equal to the ones with negative impact. Other changes are reported below.

First, the odds ratios were higher by 7.724 and 3.799 for higher and lower-grade professionals, respectively. Similarly, for a one-unit increase in the square root of annual benefit income, the odds ratio for being a couple with children household in 10 years was higher by 2.6 %. Finally, the variable receiving benefits was not found significant.

On the other hand, the marital status variables namely; never married and cohabiting couples, were significant, but their effect was very negligible. More precisely, the odds ratios were 0.0636 and 0.0620 for cohabiting couples and never married, correspondingly. The same applies to the variable "rented from private landlords" in which the odds ratio was 0.0769. This could be explained by the fact that usually, couples with children are more likely to live in dwellings owned outright or with mortgage according to the UK office for National statistics (2013b). Indeed, 65% of the privately rented accommodations were occupied by 1-2-person households. Finally, unlike the previous models, for householders aged between 36-45 years the odds of being couples with children in 10 years was lower by 56.3%. This could be attributed to the fact that around 50% of UK householders aged between 40 and 55 are prone to divorce, as indicated at an earlier stage (ONS, 2013b). As a result, they should move to other household types in the future such as lone parents or single non-elderly.

## 5.6 Conclusion

Chapter five has investigated the demographic and socio-economic factors influencing the various transitions of single non-elderly households in the next ten years. To accomplish this aim, existing transition patterns in the British household panel data survey (BHPS) between 1991 and 2008 have been first explored using descriptive statistics. This has, in turn, permitted to understand the overall change in the proportion of different households over this period. Moreover, allowed to inspect for the hidden inner transition patterns between various household types and which has, in turn, informed the change in the number of different families.

After understanding the existing transition mechanisms in the BHPS dataset, the demographic and socio-economic factors affecting those transitions have been explored using fixed and random effects binary logistic regression. Although the results of the transition models were in good agreement with the literature, the significance and magnitude of their independent factors were not consistent across all years. This provides a strong indication of the complexity of household demographic transitions and suggests the non-feasibility of statistical approaches in interpreting such a phenomenon (Carter and McGoldrick, 1988; White and Klein, 2008). In fact, it is believed that a combination of statistical and other qualitative approaches is indispensable for a holistic interpretation of factors affecting household transitions.

In addition to the above, the results' analysis suggested that the BHPS dataset cannot and should not be employed to predict transition patterns beyond the 10 years. Indeed, even though most models' independent factors were good predictors of the dependent variables, the models' R2 decreased from being 20-56% in years 1-7 to below 20% after the 9$^{th}$ year of transition. Similarly, given the models low R2 values, the employed dataset could not capture, to some extent, factors influencing the transitions to lone parent families.

The next chapter investigates the impact of the predicted transition variables on the annual electricity and gas consumption of households.

# 6

# THE IMPACT OF HOUSEHOLD TRANSITIONS ON DOMESTIC ENERGY USAGE PATTERNS

*The sixth chapter explores the effect of various households' transition patterns on their annual electricity and gas usage. Based on those findings, a prediction model which forecasts the annual electricity consumption in function of household transitions, demographic, and socio-economic factors, is developed.*

**Contents**

## 6.1 Introduction

The chapter aims to investigate the effect of the predicted transitions on the household energy consumption patterns for both gas and electricity. This will be followed by predicting the electricity consumption of different family types in the next ten years in line with their estimated transition patterns. Moreover, their significant socio-economic and demographic characteristics.

## 6.2 The impact of household transitions on their annual energy consumption

As indicated in the methodology chapter, the nature and strength of association between the predicted household transition models and their annual gas and electricity consumption variables were explored using point-Biserial correlation. Based on that, this chapter aims to analyse in detail the outcomes of the four generated point-Biserial correlation tables. Each table encompasses ten columns of a predicted transition type (e.g. couple without children) and two rows representing the square root of annual gas consumption and the log 10 of annual electricity usage. From analysing each table, it is evident that the majority of household transitions do have an impact on both gas as well as electricity consumption figures. However, this effect is weak since the majority of point-Biserial correlation coefficients $r_{pb}$ did not exceed a value of 0.2. The below subsections attempt to discuss the effect of each transition type in more detail.

### 6.2.1 The effect of transition to single non-elderly households

In the light of Table 6-1(below), it is clear that the impact of the predicted single non-elderly transitions on the log 10 of electricity consumption was significant across all years. However, this was not the case for the impact on the square root of annual gas consumption as it was not significant after the fourth year of transition. In addition to that, the magnitude of all significant correlation coefficients was positive, despite the fact it decreased over time.

Table 6-1. The impact of single non-elderly transitions on the annual gas and electricity consumption

| | Single 1 year | Single 2years | Single 3years | Single 4years | Single 5years | Single 6years | Single 7years | Single 8years | Single 9years | Single 10 years |
|---|---|---|---|---|---|---|---|---|---|---|
| Log 10 annual electricity usage | 0.18** | 0.154** | 0.153** | 0.142** | 0.07** | 0.06** | 0.07** | 0.052** | 0.05** | 0.04** |
| Sig.(2-tailed) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 |
| Square root of annual gas usage | 0.11** | 0.094** | 0.069** | 0.062** | 0.015 | -0.003 | 0.01 | -0.001 | -0.012 | -0.013 |
| Sig.(2-tailed) | 0.000 | 0.000 | 0.000 | 0.000 | 0.246 | 0.815 | 0.421 | 0.928 | 0.353 | 0.301 |
| Sample size | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 |

Note: * Significance at the 95% level ** Significance at the 99% level

First, the change affecting the point-Biserial serial correlation coefficients of the log 10 annual electricity consumption variable could be reported in 4 different phases. The first stage, which covers the coefficients of year 1 and 2, has known a 14.45 % decrease in the $r_{pb}$ coefficient. This implies that the transition to single non-elderly after two years has less effect on the annual household electricity usage than the transition after 1 year. This was logical given the 0 % transition rates from other household types such as couples with children, to single non-elderly over the first year in the BHPS dataset. Indeed, in the second year, there were around 6.5 % transitions from other household types, including 1.4% of lone parent households. Conversely, the second stage, which entails the 2nd, 3rd, and 4th years of transition, has known a slight decline in the $r_{pb}$ coefficient. More precisely, from year 2 to year 3, the $r_{pb}$ coefficient remained fairly stable, whereas from year 3 to year 4, it decreased by 7.19%. On the one hand, this is due to the fact that the transitions to single non-elderly after 3 years from other household types dropped from 6.5 % to

5.33%. On the other hand, the reason behind the fall in the $r_{pb}$ coefficient of the 4th year of transition could be attributed to the high proportion of transitions from families with larger sizes such as couple with children. In other words, despite the decline in the transition rate in this year by 0.78%, transitions from households with 2 or more individuals represented more than 52% of these transitions. As for the third phase, which includes years 4 and 5, $r_{pb}$ coefficient dropped dramatically by 49.30 % in year 5. This was reasonable given the fact that wave 1996 was omitted from the BHPS dataset and replaced by wave 1997 as it lacks energy consumption variables. In other words, the change in $r_{pb}$ was due to the transition rate of nearly 15% and which covers two waves. Finally, phase 4, which encompasses the $r_{pb}$ coefficients of year 5 to year 10, witnessed a fluctuating pattern of minor decline (0.01 on average). Following an inspection of transition rates to single non-elderly from other family types in the BHPS dataset, it was found that they also fluctuated between 6% and 4%.

In contrast to the above, the change in the $r_{pb}$ coefficients pertaining to the square root of annual gas usage variable could be described over three stages. First, stage 1, which includes the first and second year of transition, witnessed a 14.5% decrease in the $r_{pb}$ initial correlation coefficient. This was almost identical to the change in the $r_{pb}$ of the variable log 10 annual electricity usage in stage 1. In the second phase; however, the $r_{pb}$ decreased by 26.6%. Again this could be due to a 1% increase in the transition rate to single non-elderly from other household types (e.g. couple without children). In phase 3, which entails the third and fourth transitions, there was only 10.14% decline in the $r_{pb}$ coefficient. This could be because the transition rate from other family types to single non-elderly dropped by 2.13% from year 3 to 4. Finally, the $r_{pb}$ coefficients from year 5 to year 10, were not significant which implies that there is not enough evidence on the nature of the relationship between both variables.

After analysing the above facts, it is clear that the change in the point-Biserial correlation coefficients across both electricity and gas was associated with the transition from other family types such as couple without children. This, in turn, is strongly affected by the change in certain demographic and socio-economic factors such as household size, income, age, and marital status, as discussed previously. Therefore, it is believed that transitions to single non-elderly households do have an indirect impact on their future gas and electricity usage patterns.

### 6.2.2 The effect of transition to couple without children

Table 6-2.The impact of couple without children transitions on the annual gas and electricity consumption

| | CN 1 year | CN 2 years | CN 3 years | CN 4 years | CN 5 years | CN 6 years | CN 7 years | CN 8 years | CN9 years | CN 10 years |
|---|---|---|---|---|---|---|---|---|---|---|
| **Log 10 annual electricity usage** | 0.11** | 0.093** | 0.098** | 0.094** | 0.08** | -.034** | 0.04** | 0.03** | 0.02 | 0.008 |
| **Sig.(2-tailed)** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.005 | 0.000 | 0.006 | 0.074 | 0.538 |
| **Square root of annual gas usage** | 0.114** | 0.091** | 0.068** | 0.057** | 0.05** | -.09** | 0.013 | 0.001 | -.008 | -.010 |
| **Sig.(2-tailed)** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.290 | 0.930 | 0.553 | 0.419 |
| **Sample size** | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 |

Note: CN: Couple no children * Significance at the 95% level ** Significance at the 99% level

Table 6-2 represents the point-Biserial correlation coefficient matrix between couple without children transition variables and household energy consumption factors. Overall, it is clear that the resulting point-Biserial correlation coefficients were not statistically significant over the ten years. Furthermore, the direction of these associations was positive except for the 6th year of transition. Finally, the $r_{pb}$ related to both the log10 annual electricity usage and the square root of annual gas consumption decreased in function of time. This variation was marginally affected by the change in the proportion of couples without children, which suggests that other factors, including demographic and socioeconomic, could also have an influence.

First, the change in the $r_{pb}$ coefficients of the log 10 of annual electricity consumption could be reported in 4 distinct stages. The first stage, which encompasses the $r_{pb}$ coefficients of year 1 and 2, has witnessed a decrease of 27.27% in the initial $r_{pb}$ coefficient. This outcome was unexpected given the fact that transition to couple without children is usually accompanied by an increase in the household size and income levels. However, due to the high proportion of full-time employment among this family type (67.6% on average), this decline could be attributed to their patterns of presence at home.

Similarly, the second phase, which covers the 2nd, 3rd, and 4th years of transition, has known a 1.08% increase in the $r_{pb}$ coefficient. More specifically, this was split into 5.38% increase from year 2 to 3 as well as 4.08% decline from year 3 to 4. The increase of $r_{pb}$ coefficient from year 2 to 3 could be due to a 2% increase in the proportion of couples without children over this period. However, the decrease in the point-Biserial correlation coefficient over the period from year 3 to 4 was not likely to be related to the change in the proportion of couples without children, which increased by 15% over this period.

As for the third phase, which entails the 4th, 5th, 6th years of transition, there was not only a decline in $r_{pb}$ coefficients but also a change in the direction of the point-Biserial correlation coefficient of year 6. After a careful examination of the transition rates in the BHPS dataset over this period, it was found that the proportion of couples without children over this period remained stable. This reinforces the argument made previously with regards the possible effect of other socio-economic variables, including patterns of presence at home, income, dwelling type, and others.

In contrast to phase 3, phase 4, which includes the 6th, 7th, and 8th transitions models, did not only witnessed a 213% increase in the $r_{pb}$ coefficient of year 6 but also a change in its direction of association with the Log 10 of annual electricity consumption variable. Considering the 2.4% decrease in the percentage of couples without children from year 6 to 7, it is evident that the variation in the proportion of this family type has little or no effect on the $r_{pb}$ correlation coefficients.

On the other hand, the variation in the $r_{pb}$ correlation coefficients of the square root of annual gas consumption could be discussed across three distinct stages. In the first stage, which includes the 1st, 2nd, 3rd, and 4th years, there was around 20% yearly decrease in the point-Biserial coefficient of this variable. Like the variation in the $r_{pb}$ coefficient for the variable log 10 electricity, these results were controversial given the increasing percentage of this family type over this period, which implies that other factors are responsible for this decline. Similarly, in stage 2, which comprises year 4 and 5, the decrease in the $r_{pb}$ coefficient was not as significant as in the previous phase with a value of 12.28%. Finally, stage 3, which entails the 5th and 6th years, witnessed a 280% decline in the point-Biserial correlation coefficient over this period and the change in the direction of its association with year 6. This was in line with the 4th stage depicting the change in the point-Biserial coefficient of log 10 annual electricity usage. After analysing the above facts, it is evident that the variation in the portion of couples without children over time does have a minor effect on the $r_{pb}$ coefficients of the variable square root annual gas consumption.

### 6.2.3 The effect of transition to lone parent households

Table 6-3 outlines the association matrix between lone parent transition variables and annual energy usage variables. In general, it is evident that point-Biserial correlation coefficients were not significant across all the ten years and their direction was mostly negative. In addition to this, the $r_{pb}$ correlation coefficients pertaining to both the log10 of annual electricity usage and the square root of annual gas consumption variables decreased over the course of 10 years. Finally, the change in the proportion of lone parent households does have a negative effect on both variables. However, it is smaller on the log 10 of annual electricity consumption.

Table 6-3.The impact of lone parent transitions on the annual gas and electricity consumption

| | LP 1year | LP 2years | LP 3years | LP 4years | LP 5years | LP 6years | LP 7years | LP 8years | LP 9years | LP 10years |
|---|---|---|---|---|---|---|---|---|---|---|
| Log 10 annual electricity usage | 0.11** | 0.12** | 0.01 | 0.03* | 0.01 | -.034** | -.05** | -.05** | -.06** | -.053** |
| Sig.(2-tailed) | 0.000 | 0.000 | 0.363 | 0.03 | 0.494 | 0.005 | 0.000 | 0.000 | 0.000 | 0.000 |
| Square root of annual gas usage | 0.008 | 0.005 | -.05** | -.06** | -.05** | -.09** | -.13** | -.13** | -.12** | -.116** |
| Sig.(2-tailed) | 0.509 | 0.701 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Sample size | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 |

Note: LP: Lone parent * Significance at the 95% level ** Significance at the 99% level

First, the variation in the point-Biserial coefficient of the log 10 annual electricity consumption could be reported in two distinct stages. In the first stage, which covers the period from year 1 to year 5, the $r_{pb}$ coefficients of year 3 and 5 were found statistically insignificant. However, there has been a 72.73% decline in the initial $r_{pb}$ coefficient from year 1 to 4. In particular, this was split into a 9% increase in the point-biserial coefficient from year 1 to year 2. Moreover, a 75% decline from year 2 to year 4. Based on the existing transition patterns in the BHPS dataset, it is believed that the rise in the $r_{pb}$ in year 2 was caused by an increase in the number of lone parents as a result of single non-elderly transitions. Therefore, it could be argued that the rise in the household size could have affected the $r_{pb}$ correlation coefficient over this period. On the other hand, considering the 30.76% increase in the proportion of this family type over the period from year 2 to 4, the decline in the corresponding $r_{pb}$ correlation coefficient should be related to other demographic and socio-economic factors such as income. Secondly, in phase 2, which embodies the period from year 6 to 10, there was a negative association between the transition variables and the log 10 of yearly electricity usage. This implies that households becoming or remaining lone parents after 6 years should have lower electricity figures as a result of a significant decrease in their income as well as a possible change in their socio-economic circumstances. Indeed, lone parent households had the second lowest average annual labour income over the 18 BHPS waves, as shown in Figure 6-1. Moreover, the ones with low-income consume the least amount of electricity per year (2500KWh) according to a recent report from Ofgem (Ofgem, 2014).

In contrast to the above, significant associations between the square root of yearly household gas usage and lone parent transition variables were negative. Moreover, their magnitude was prone to fluctuating patterns of decline over this period. However, despite this fact, the $r_{pb}$ coefficients did also increase and remain stable over certain years. Therefore, it would be beneficial to report this variation in 4 different phases. First, phase 1, which entails the period from year 3 to year 5, witnessed a stability in the magnitude of the $r_{pb}$ coefficient despite the 23.53% decline in the proportion of this family type. Conversely, phase 2, which covers the 5$^{th}$, 6$^{th}$, and 7$^{th}$ years of transition, knew a 260% decrease in the $r_{pb}$ correlation coefficient of year 5 while the proportion of lone parents remained stable over this period. This, in turn, suggests the influence of other factors, including socio-economic and demographic. For example, lone parents who have consistently been in fuel poverty due to their income insecurity (DECC, 2015a). However, this was not the case for phase 3, in which both the point-Biserial coefficient $r_{pb}$ and the proportion of lone parents remained stable over the period from year 7 to 8. In phase 4, which covers the 8$^{th}$, 9$^{th}$, and 10$^{th}$ years of transition, there was a 10.77% increase in the point-Biserial correlation coefficient of year 8. This was attributed to a 30.77% decline in the proportion of lone parents over this period. After analysing these facts, it is clear that there is a negative relationship between the percentage of lone parent households and the point-Biserial coefficients of their square root of annual gas energy consumption.

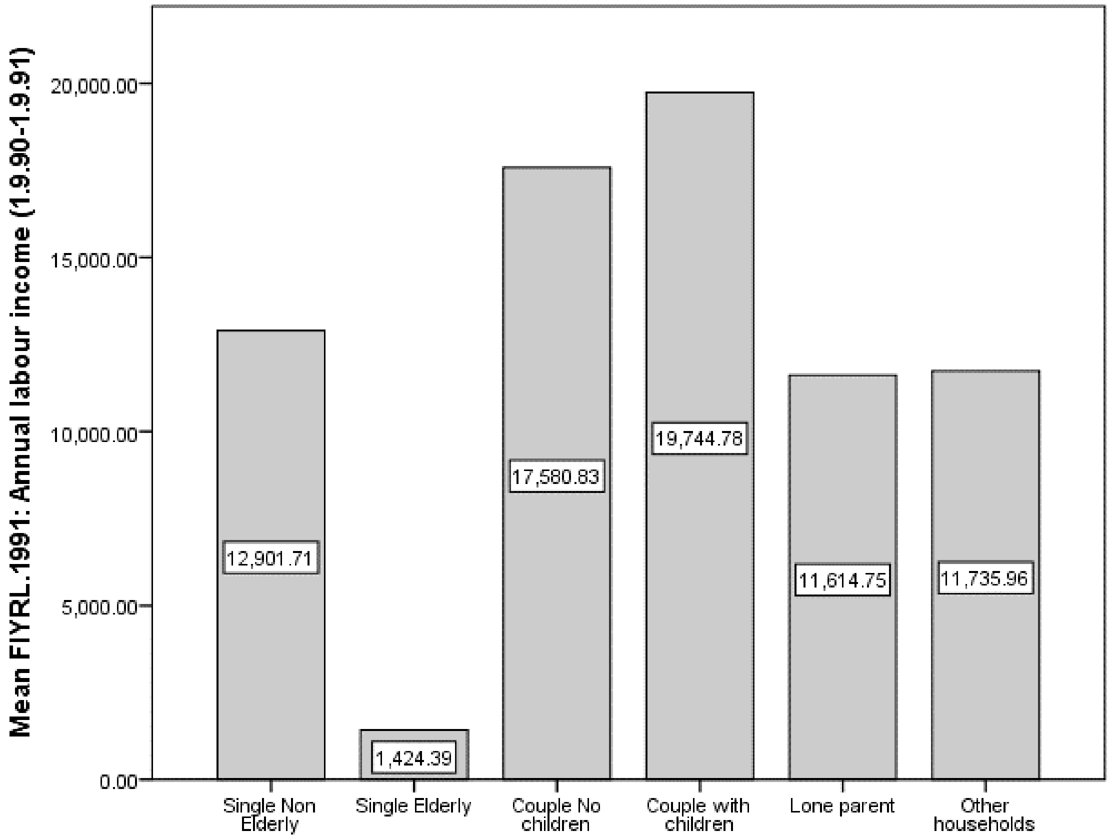



Figure 6-1.The average annual labour income of different household types in the BHPS panel data survey (1991-2008)

### 6.2.4 The effect of transition to couple with children households

Table 6-4 shows the point-Biserial correlation coefficients between energy consumption variables and couple with children transition factors. Overall, it is evident that the existing associations were positive except for the 6$^{th}$ and 7$^{th}$ years. Furthermore, they were prone to a decline over ten years. Finally, the change in the proportion of couples with children did not have an impact on the $r_{pb}$ coefficients pertaining to the log10 annual electricity consumption. However, it had a minor effect on the $r_{pb}$ coefficients of the variable square root yearly gas usage but was very inconsistent across all variables.

Table 6-4.The impact of couple with children transitions on the annual gas and electricity consumption

| | CWC 1year | CWC 2 years | CWC 3years | CWC 4years | CWC 5years | CWC 6years | CWC 7years | CWC 8years | CWC 9years | CWC 10years |
|---|---|---|---|---|---|---|---|---|---|---|
| **Log 10 annual electricity usage** | 0.16** | 0.13** | 0.14** | 0.13** | 0.11** | -0.03* | -.04** | 0.07** | 0.05** | 0.05** |
| **Sig.(2-tailed)** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.015 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Square root of annual gas usage** | 0.160** | 0.135** | 0.11** | 0.1** | 0.09* | -.09** | -.12** | 0.03* | 0.010 | 0.011 |
| **Sig.(2-tailed)** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.011 | 0.443 | 0.382 |
| **N** | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 | 6700 |

Note: CWC: Couple with children * Significance at the 95% level ** Significance at the 99% level

First, the variation in the point-Biserial correlation coefficients of the log 10 annual electricity consumption could be summarised in 5 distinct phases. In phase 1, the $r_{pb}$ coefficient of year 1 decreased by 18.75% in year 2 while the proportion of couple with children households almost doubled over this period. In stage 2, however; the point-Biserial correlation of year 4 remained stable in comparison to year 2, although it increased by 7.7% in year 3. At first glance, it was thought that the increase in the $r_{pb}$ coefficient of year 3 could be associated with the decrease in the proportion of non-dependent children in the BHPS dataset over this period. However, conducting a correlation analysis between children dependency and the annual electricity consumption variables, revealed no significant association between both factors. In phase 3, which encompasses the period from year 4 to 7, witnessed a 130% decline in the $r_{pb}$ and a change in its direction in years 6 and 7. In parallel to these changes, there was a 6.38% increase in the proportion of couple with children. In phase 5, the direction of the $r_{pb}$ coefficient of year 8 became positive, and its magnitude almost tripled in comparison to year 7. This was also accompanied by a 12% rise in the percentage of this family type. Conversely, in phase 5, the point-Biserial correlation coefficient of year 8 decreased from 0.07 to 0.05 in year 9 and then remained stable in year 10. In the light of the above facts, it is evident that the change in the proportion of couple with children over time does not have an impact on their electricity consumption figures.

In contrast to the above, the change in the $r_{pb}$ correlation coefficients of the variable square root of annual gas usage could be analysed in 3 different stages. First, stage 1, which comprises the period from year 1 to 5, knew a 43.75% decrease in the initial $r_{pb}$ coefficient. In more details, from year 1 to 2, there has been a 15% decline in the $r_{pb}$ coefficient while the proportion of couples with children almost doubled. Similarly, from year 2 to year 3, the percentage of couples with children rose by 29.16%, and the $r_{pb}$ coefficient decreased by 18%. Based on those facts, it could be argued that there should be a negative relationship between the proportion of this family type and the point-Biserial correlation coefficients of the square root yearly gas usage variable over the first five years of transition. As for stage 2, the $r_{pb}$ correlation coefficient declined from 0.09 in year 5 to -0.12 in year 7. Furthermore, the proportion of couple with children rose by 8.51 % from year 5 to 6, whereas, it declined by 1.96% from year 6 to 7. This indicates that variation in the proportion of this household type did not have a consistent effect on $r_{pb}$ correlation coefficients during this period. Finally, in stage 3, there was a change in the direction of $r_{pb}$ correlation coefficient and a 125% increase in its magnitude from year 7 to 8. Considering the 12% rise in the percentage of this family type over this period, it is evident that it did not influence the increase in the corresponding $r_{pb}$.

In the light of the above point-Biserial correlation coefficient tables, it is clear that the change in the proportion of a given household type might have an impact on the variation in the point-Biserial correlation coefficient. However, this depends on the type of household. For instance, the increase in the proportion of single non-elderly could lead to a rise in the $r_{pb}$ coefficient of log 10 yearly electricity consumption, whereas the change in the percentage of couples with children, does not. As indicated at an earlier stage, it is believed that other socio-economic and demographic factors should also have an impact. Therefore, this subsection does not only aim to predict future energy usage patterns in function of the forecasted transition models but also the change in the households' socio-economic as well as demographic circumstances.

## 6.3 Electricity consumption prediction model

Table 6-5 to Table 6-11(below) represent the estimation results of the multiple linear regression model of electricity consumption. Overall, it is evident that the majority of the independent variables had a significant and a positive impact on the dependent factor. However, there was an exception of few variables such as the log10 of household total annual income and routine non-manual employees. In addition to that, all the model's independent factors only explain 25% of the variation in the log 10 annual electricity consumption in which the predicted transition variables account for only 6% of the total.

### Transition variables

First, for individuals becoming (or remaining) single non-elderly in the next two years, their annual electricity consumption is more likely to increase by around 13% holding other variables constant. However, this effect is more likely to decrease reaching a value of 2.4% for the transition variable corresponding to year 10. Similarly, for households becoming couples without children in one year, their corresponding annual electricity consumption rises by 16.1% while holding the other independent factors fixed. As expected from analysing the point-Biserial correlation coefficients of this household type in the previous subsection, the 16.1% variation in the dependent factor witnessed a fluctuating pattern of decline over the ten transition factors. For instance, the annual electricity consumption in year 5 is expected to rise by 16.41%, as a result of households moving to a couple without children family in this year. Conversely, for the ones making this transition in the following year, their annual usage is more likely to decrease by 4% while making the other variables constant. The same applies to the remaining transition variables, including couples with children and lone parents. For example, households becoming lone parents after 6 years could experience a 5.47% decline in their annual electricity consumption, whereas the ones becoming couples with children in the same period, could use 3.5 % less electricity per year.

### Age of household reference person (HRP)

In contrast to the above transition factors, other socio-economic and demographic variables also had a significant impact on the household annual electricity usage patterns. First, the yearly electricity consumption is more likely to be lower by 28% and 9.9 % for householders aged 16-25 and aged 26-35, correspondingly. On the other hand, if the household reference person (HRP) is aged 46-55 and 56-65, their yearly electricity usage should increase by 8.7% and 4.95%, respectively. These results were in good agreement with the work of Jones and Lomas (2015), Leahy and Lyons (2010), McLoughlin et al. (2012), and Brounen et al. (2012), suggesting that electricity usage is usually higher for HRP aged between 50 and 65. As discussed in the literature review chapters, this is because middle age householders usually have a high number of occupants and longer presence at home as they usually live in couple with children households. Apart from age, occupants other than the main householder could potentially consume 22.34% more electricity per year while holding other variables constant. This is reasonable given the fact that the main householders are usually the ones in charge of paying the utility bills. As a result, they are more willing to reduce their consumption than the ones who do not.

### Location

In addition to the above, the location variables had a significant effect. For instance, the electricity energy consumption of householders living in east Anglia is expected to increase by 50% while holding the rest of independent factors fixed. According to the most recent official report on energy consumption in England and Wales (ONS, 2011a), this is attributed to the fact that this region has the highest proportion of economy 7 (dual tariff) meters, in which the average electricity consumption is around 5176 KWh. This is almost 56% more than the average medium yearly electricity consumption suggested by Ofgem (3300KWh) (Ofgem, 2014). On the other hand, householders living in the region of South Yorkshire should use 32% less annual electricity while the remaining factors are held constant. These figures were in line with a report from the department of trade and industry (DTI, 2005) advising that the following South Yorkshire areas namely; Barnsley, Sheffield, Rotherham, and Doncaster, had average yearly electricity figures below 3800 KWh.

### Dwelling type and size

Dwelling type variables were also found significant but with a negative impact. For example, for households living in terraced houses, the corresponding yearly electricity consumption is more likely to decline by 16.30% holding other variables constant. Similarly, a decrease of 11.9%, 12.9%, and 14.29% is expected for households dwelling in semi-detached houses, end-terraced, and purpose-built flats, respectively. These results were expected given the fact that terraced dwellings and flats have small floor areas. Moreover, they are mostly occupied by a small number of occupants, as indicated in chapter five. Furthermore, they stand out as being more efficient than other dwelling types such as detached houses (BEIS, 2016). Therefore, the estimated decrease in the annual electricity consumption of households living in end-terraced and semi-detached houses was higher as opposed to terraced and purpose-built flats. Figure 6-2 (below), which shows the average annual electricity consumption per UK dwelling type, clearly reinforces these findings. On the other hand, for a one additional room in the accommodation, the household yearly electricity consumption is more likely to increase by 7.9%. This result was in excellent agreement with the work of Longhi, (2014) who suggested that an extra room in the dwelling is associated with 6% increase in the annual household expenditure on electricity. Other studies, including (Bedir et al., 2013; Baker and Rylatt, 2008; Leahy and Lyons, 2010), were also in line with this outcome.

### Heating fuel type and previous consumption figures

From analysing chapter three, fuel variables were expected to have a significant impact on the annual electricity consumption. For households using gas and solid as their main heating fuel, their yearly electricity usage is likely to be lower by 15.28% and 19.54%, correspondingly, while holding the rest of factors constant. Conversely, the ones with central electric heating are more likely to have a 61% increase in their annual electricity consumption while holding other variables fixed. These findings were in line with the Ofgem report (Ofgem, 2015), advising that households heating their homes using electricity, consume 6500 KWh on average as opposed to 3200 KWh for the ones using gas. The household electricity consumption of the precedent year had also an effect on the rise in electricity usage of the subsequent one. More specifically, for a 1 KWh increase in the actual electricity consumption, the future one is more likely to rise by 15.35% while holding other impendent variables constant.

### Employment mode and socio-economic class

As expected, householders working part-time are more likely to experience a 5.4% increase in their electricity usage due to their long presence at home during the day. Similar findings were also reported by Baker and Rylatt (2008) and Fell and King (2012). Apart from that, householders who belong to the socio-economic class small-proprietors without employees are more likely to witness a 13.60% rise in their electricity consumption while holding other independent variables constant. As indicated at an earlier stage, this could be owing to the fact that a large proportion of this class are couples with children (CitizenAdvice, 2015). Another possibility could be related to the high income of some of them as indicated by Ofgem (2014). This report also suggested that 30% of this socio-economic class should consume around 5600 KWh of electricity annually. However, the remaining socio-economic class variables were not found significant which aligns with the work of Yohanis et al. (2008) and Cramer et al. (1985).

### Income

The log 10 of household total annual income was not found statistically significant. However, the square root of annual benefit income had a minor significant effect on the dependent variable. More specifically, for a one-unit increase in the annual benefit income of the household, the yearly electricity consumption decreases by 0.12% while holding other independent variables constant. Considering the proportion of couples with children on benefits (IFS, 2015), this result is consistent with the study of van den Brom et al. (2017). In particular, this latter suggests that the electricity usage of couples with children, who are receiving benefits, is higher than those in employment. Again, this could also be linked to the fact that these households are more likely to have a longer presence at home during the day; consequently, consume more electricity (Brounen et al., 2012; McLoughlin et al., 2012; Guerra Santin, 2011). Another factor could be the old age of the dwellings, poor ventilation, and inefficiencies of the appliances of certain benefits claimers especially those with low-income levels (NRDC, 2016; Tunstall et al., 2013). On the other hand, for households on a pension, the annual electricity consumption decreased by 8.38% holding the rest of independent factors fixed. Given the fact that those receiving a pension are mostly aged above 65, this result was in good agreement with the work of Yohanis et al. (2008), Cramer et al. (1985), Leahy and Lyons (2010), McLoughlin et al. (2012), Kavousian et al. (2013).

Table 6-5.The multiple regression model of the household annual electricity consumption (part1)

| | Independent variables | Electricity consumption | 95 % Confidence intervals |
|---|---|---|---|
| Single non-elderly transition variables | Single non-elderly transition variables | | |
| | | | |
| | Single after 2 years | 0.0539*** | 0.04 / 0.068 |
| | | (0.00457) | |
| | Single after 3 years | 0.049*** | 0.036 / 0.062 |
| | | (.0048633) | |
| | Single after 4 years | 0.048*** | 0.035 / 0.0616 |
| | | (.0030135) | |
| | Single after 5 years | 0.027*** | 0.0143 / 0.045 |
| | | (0.00495) | |
| | Single after 6 years | 0.02*** | 0.011 / 0.048 |
| | | (0.0035) | |
| | Single after 7 years | 0.023*** | 0.0134 / 0.0485 |
| | | (0.005) | |
| | Single after 8 years | 0.013*** | 0.0094 / 0.0376 |
| | | (0.00257) | |
| | Single after 9 years | 0.0115*** | 0.00962 / 0.0379 |
| | | (0.00616) | |
| | Single after 10 years | 0.0103*** | 0.00862 / 0.0354 |
| | | (0.00681) | |

Note: Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6-6.The multiple regression model of the household annual electricity consumption (part2)

| | Independent variables | Electricity consumption | 95 % Confidence intervals |
|---|---|---|---|
| Couples without children transition variables | Couple without children after 1 years | 0.0647*** | 0.025 / 0.095 |
| | | (.0070651) | |
| | Couple without children after 2 years | 0.064*** | 0.0224 / 0.0945 |
| | | (.0075472) | |
| | Couple without children after 3 years | 0.0662*** | 0.0226 / 0.0947 |
| | | (.0078574) | |
| | Couple without children after 4 years | 0.067*** | 0.02254 / 0.0948 |
| | | (0.0084042) | |
| | Couple without children after 5 years | 0.066*** | 0.0224 / 0.096 |
| | | (.0088884) | |
| | Couple without children after 6 years | -0.01778*** | -0.009 / -0.0352 |
| | | (.0063607) | |
| | Couple without children after 7 years | 0.0370*** | 0.0124 / 0.0748 |
| | | (.0085751) | |
| | Couple without children after 8 years | 0.0265*** | 0.0094 / 0.0678 |
| | | (.0092059) | |
| | Couple without children after 9 years | 0.0210 | 0.0097 / 0.0679 |
| | | (.0099222) | |
| | Couple without children after 10 years | .0135575 | 0.007 / 0.0629 |
| | | (.01073) | |

Note: Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6-7.The multiple regression model of the household annual electricity consumption (part3)

| | Independent variables | Electricity consumption | 95 % Confidence intervals |
|---|---|---|---|
| Lone parents transition variables | Lone parent after 1 years | 0.1131*** | 0.0914 / 0.3897 |
| | | (.01241) | |
| | Lone parent after 2 years | 0.1272*** | 0.0917 / 0. 391 |
| | | (0.0135) | |
| | Lone parent after 3 years | 0.00591 | 0.00198 / 0.00783 |
| | | (.0064) | |
| | Lone parent after 4 years | 0.0139* | 0.0095 / 0.037 |
| | | (.0064) | |
| | Lone parent after 5 years | .0043518 | 0.0011 / 0.007 |
| | | (.0063556) | |
| | Lone parent after 6 years | -0.01778*** | -0.009 / -0.0323 |
| | | (0.0063607) | |
| | Lone parent after 7 years | -0.02445*** | -0.0095 / -0.0324 |
| | | (.0063496) | |
| | Lone parent after 8 years | -0.0285*** | -0.0421 / - 0.041 |
| | | (.0064141) | |
| | Lone parent after 9 years | -0.0308*** | -0.0443 / -0.0405 |
| | | (.0065652) | |
| | Lone parent after 10 years | -.0296*** | -0.0098 / - 0.053 |
| | | (.0068216) | |

Note: Standard errors in parentheses $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 6-8.The multiple regression model of the household annual electricity consumption (part4)

| | Independent variables | Electricity consumption | 95 % Confidence intervals |
|---|---|---|---|
| **Couples with children transition variables** | Couple with children after 1 years | $0.1064^{***}$ | 0.085 / 0.240 |
| | | (0.0078) | |
| | Couple with children after 2 years | $0.101^{***}$ | 0.0865 / 0.231 |
| | | (0.00846) | |
| | Couple with children after 3 years | $0.1012^{***}$ | 0.0858 / 0.227 |
| | | (.00874) | |
| | Couple with children after 4 years | $0.1001^{***}$ | 0.0832 / 0.215 |
| | | (.00916) | |
| | Couple with children after 5 years | $0.0932^{***}$ | 0.0752 / 0.121 |
| | | (.009832) | |
| | Couple with children after 6 years | $-0.0155^{***}$ | -0.007 / -0.026 |
| | | (.0064) | |
| | Couple with children after 7 years | $-0.0225^{***}$ | -0.0081 / -0.032 |
| | | (.0063468) | |
| | Couple with children after 8 years | $0.06^{***}$ | 0.03 / 0.1 |
| | | (.0106408) | |
| | Couple with children after 9 years | $0.051^{***}$ | 0.0114 / 0.0862 |
| | | (.0118341) | |
| | Couple with children after 10 years | $0.0504^{***}$ | 0.012 / 0.0816 |
| | | (.0118202) | |

Note: Standard errors in parentheses $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 6-9.The multiple regression model of the household annual electricity consumption (part5)

| | Independent variables | Electricity consumption | 95 % Confidence intervals |
|---|---|---|---|
| Age and household reference person related factors | Aged between 16 and 25 | -0.143*** | -.1830092 / -.1024026 |
| | | (0.02056) | |
| | Aged between 26 and 35 | -0.0454*** | -.0642296 / - 0.265398 |
| | | (0.009613) | |
| | Aged between 46 and 55 | 0.0364212*** | 0.0196 / 0.054 |
| | | (0.0086) | |
| | Aged between 56 and 65 | 0.021** | 0.0024 / 0.039 |
| | | (0.00874) | |
| | Not the main householder | 0.0876*** | 0.0588 / 0.1163 |
| | | (0.0132) | |
| Location and accommodation related factors | Living in east Anglia | 0.179*** | 0.0557 / 0.302 |
| | | (0.0563) | |
| | Living in South Yorkshire | -0.153** | -0.328 / 0.0206 |
| | | (0.0771) | |
| | Living in terraced houses | -0.0773*** | -0.1004 / -0.039 |
| | | (0.0142) | |
| | Living in semi-detached dwellings | -0.055*** | -0.0808 / -0.0281 |
| | | (0.0122) | |
| | Living in end-terraced houses | -0.06*** | -0.0985 / -0.0211 |
| | | (0.0174) | |
| | Living in purpose built flats | -0.067*** | -0.097 / -.0362 |
| | | (0.0142) | |
| | Number of rooms in the dwelling | 0.033*** | 0.025 / 0.04089 |
| | | (0.00372) | |

Note: Standard errors in parentheses $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table 6-10.The multiple regression model of the household annual electricity consumption (part6)

| | | Electricity consumption | 95 % Confidence intervals |
|---|---|---|---|
| Socio-economic class and income related variables | Working part-time | 0.0228*** | |
| | | (0.0124) | |
| | Routine non-manual workers | 0.0166 | -0.0345 / 0.029 |
| | | (0.0141) | |
| | Farmers smallholders | -0.0782 | -0.220 / -0.011781 |
| | | (0.0493) | |
| | Small proprietors without employees | 0.0554*** | 0.0188 / 0.0921 |
| | | (0.0186) | |
| | Log10 of annual household income | -0.0191 | -0.0144 / 0.0527 |
| | | (0.0153) | |
| | Square root of non-labour income | -0.00052** | -.0008 / -0.00024 |
| | | (0.000127) | |
| | Receiving benefit | 0.033*** | 0.01104 / 0.0545 |
| | | (0.00975) | |
| | On pension | -0.038** | -0.071 / -0.0045 |
| | | (0.0150) | |
| **R2-Sq** | | 0.25 | |
| **Constant** | | 3.113*** | 2.938752 / 3.286249 |
| | | (0.0651) | |
| **Sample size *N*** | | 4393 | |
| **Type of Model** | | Fixed effects | |

Note: Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6-11.The multiple regression model of the household annual electricity consumption (part7)

| | Independent variables | Electricity consumption | 95 % Confidence intervals |
|---|---|---|---|
| Energy consumption related factors | Using gas as the main heating fuel | -0.072*** | -0.10047 / -.0432 |
| | | (0.0132) | |
| | Using electricity as the main heating fuel | 0.207*** | 0.171 / 0.243 |
| | | (0.0159) | |
| | Using solid fuel as the main heating fuel | -0.0944** | -0.023 / -0.212 |
| | | (0.0299) | |
| | Log10 electricity consumption of the previous year | 0.062*** | 0.03123 / 0.0921 |
| | | (0.016) | |

Note: Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

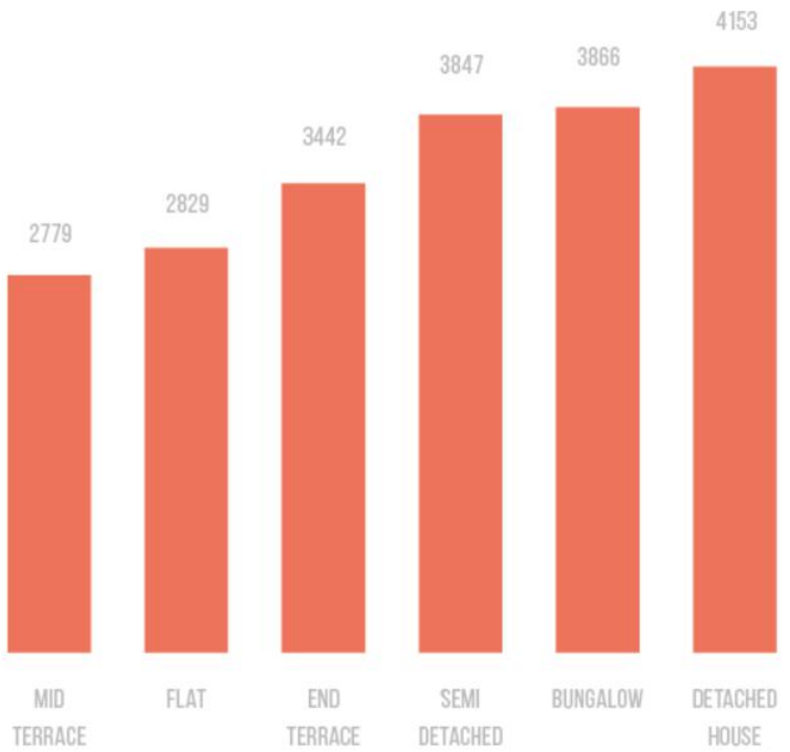



Figure 6-2. The average annual electricity usage in KWh per dwelling type (OVOenergy, 2014)

## 6.4 Conclusion

This chapter aimed to answer the following research questions;

- *Is there a significant relationship between household transitions and energy usage patterns?*
- *If so, to what extent it does affect the household energy consumption figures?*

To answer the first research question, point bi-serial correlation analysis was adopted. The examination of the point bi-serial correlation coefficients has shown that the majority of future transition patterns, except the ones to lone parents, had a positive but weak effect on the yearly gas and electricity consumption.

On the other hand, multiple linear regression analysis was used to answer the second research question. Besides the good agreement of the models' coefficients with the literature, the analysis of the model also suggested that the transition variables are only responsible for 6% of the variation in the residential electricity usage. In this respect, it is argued that the generation of this new knowledge opens the door to many possibilities. One of the opportunities is to develop a 3D urban energy forecasting tool that maps the household future energy usage onto their various transition probabilities. Thus, we present an example (Figure 6-3) showing the variation in energy usage patterns of single non-elderly households during their transitions to other family types after five years.

Overall, analysis of the findings indicated that the transition probabilities from singles non-elderly to couples without children, couples with children, lones parents, singles elderly, and other households after five years were 19.9%, 12.1%, 3.1%, 46.7%, and 1.7%, respectively. For the majority of households transitioning to single elderly households in 5 years, 67.5% should consume between 1000 and 3000 kWh electricity annually. On the other hand, of those that become couples without children households, 4.76%, 42.85%, and 26.19% are expected to consume less than 1000 kWh, 1000–3000 kWh, and 3000–4000 kWh, respectively. As for the majority of households transitioning to couple with children status (53%), their use should be between 2000 and 4000 kWh annually, whereas 19.6% should consume more than 5000 kWh. Concerning households making transitions to lone parent households, around 57% should use 2000–4000 kWh, whereas the remaining should use more than

5000 kWh per year. Interestingly, the annual electricity consumption of the majority of those who remained single non-elderly households after five years and consumed between 1000 and 4000 kWh should increase by an average of 1.02%. Conversely, for those who consumed 4000–5000 kWh and more than 5000 kWh, a decrease of 1.1% and 3%, respectively, is expected.

The next chapter aims to discuss the socio-economic and demographic characteristics of the pilot area. Moreover, discusses in depth its CityGML LOD3 modelling process.

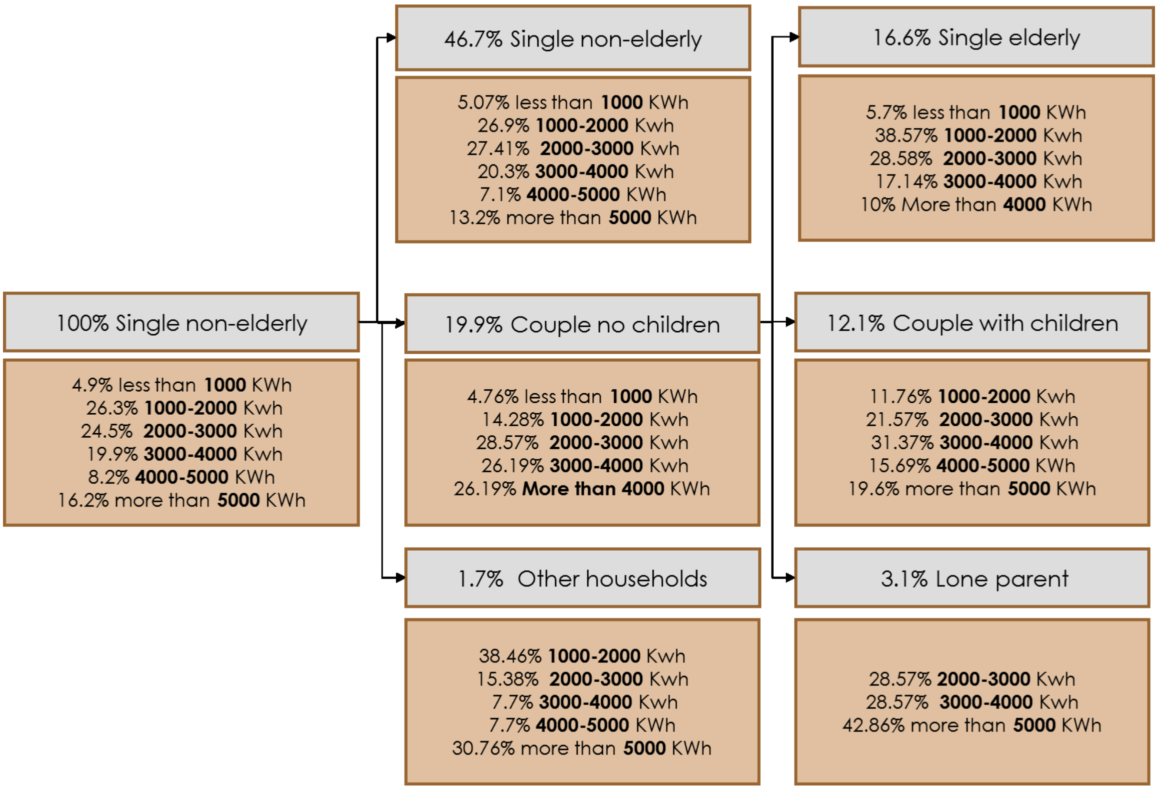



Figure 6-3.Expected annual electricity consumption figures of single non-elderly households before and after their transition to different family types in 5 years.

# 7

# Analysis of the pilot area and the initial development of its 3D urban energy prediction tool

*The seventh chapter analyses the demographic and socio-economic characteristics of the Sneinton area in relation to Nottingham city. Furthermore, discusses in depth its CityGML modelling process.*

**Contents**

## 7.1 Introduction

The previous chapter has investigated the effect of household transition variables on their annual gas and electricity usage. Based on that, a linear regression model, which encompasses the household demographic, socio-economic, and physical factors, was developed to enable the prediction of future annual electricity usage patterns accordingly. All in all, the findings of chapter six were in good agreement with the literature. However, given their current format, they are not enough to utilise because some urban planners are unfamiliar with raw statistical models. Instead, they are used to the interaction with 2D maps or 3D models.

In the light of the above, chapters seven and eight address primarily how these findings are incorporated into an intuitive and user-friendly system. This, in turn, takes into account the strengths and weaknesses of urban energy forecasting approaches discussed in chapters two and three. In particular, chapter seven sheds light on the demographic and socio-economic profile of the pilot area residents. Moreover, discusses some physical characteristics of its dwelling typologies. However, this chapter only focuses on the initial development of the pilot area 3D semantic model.

## 7.2 Chapter structure

This chapter has been carefully organised into two major sections (Figure 7-1). In the first one, the rationale behind choosing the Sneinton area, in particular, is introduced. This is followed by a brief historical overview of this area and a demographic and socio-economic analysis of its residents. More precisely, the analysis comprised their distributions of age, gender, marital status, ethnic background, household type, educational qualifications, socio-economic class, and employment mode. In addition to that, their dwelling types and tenure modes are also outlined.

Finally, based on the Cities Revealed (CR) database developed by GeoInformationGroup (2017), the physical characteristics of the Sneinton dwelling typologies are discussed. These encompass the year of construction, number of floors, wall and roof type.

On the other hand, the second section examines, in depth, the process of building a CityGML LOD3 model. This will, in turn, shed light on the employed data, interoperability issues and strategies, 3D modelling techniques and technologies, data cleaning and processing. Each phase will be explicitly supported by illustrations to allow the reader to follow the sequencing and development of this section easily.

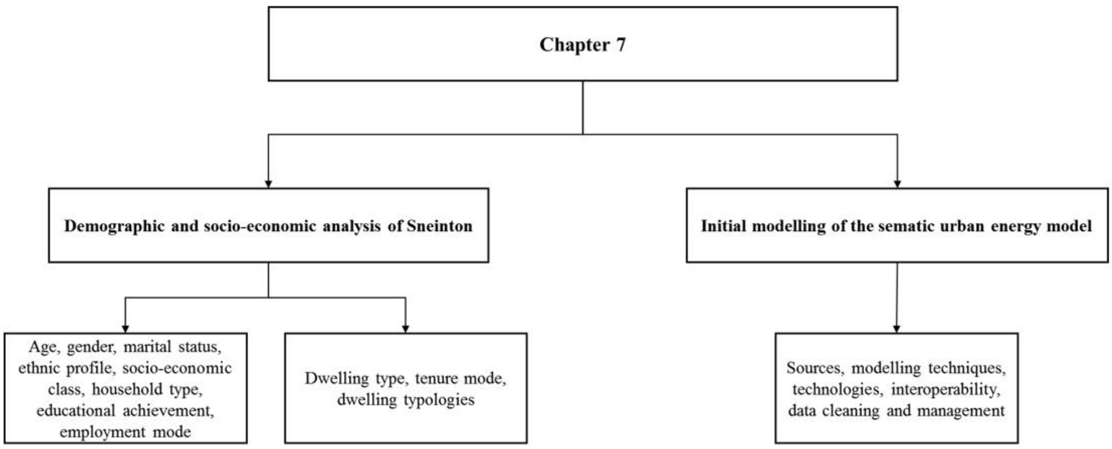



Figure 7-1. Structure of chapter 7

## 7.3 Selection process of the pilot area

Figure 7-2 illustrates the areas that are located within a 1-mile radius from the researcher location (Nottingham Trent University, city campus). These include the following neighbourhoods namely; Arboretum, The park, Hockley, Sneinton, St Ann's, Forest Fields, and Hyson green. On the other hand, Table summarises the proximity of these areas from the researcher location and indicates the availability of secondary data on the residents' various characteristics and their past energy related figures.

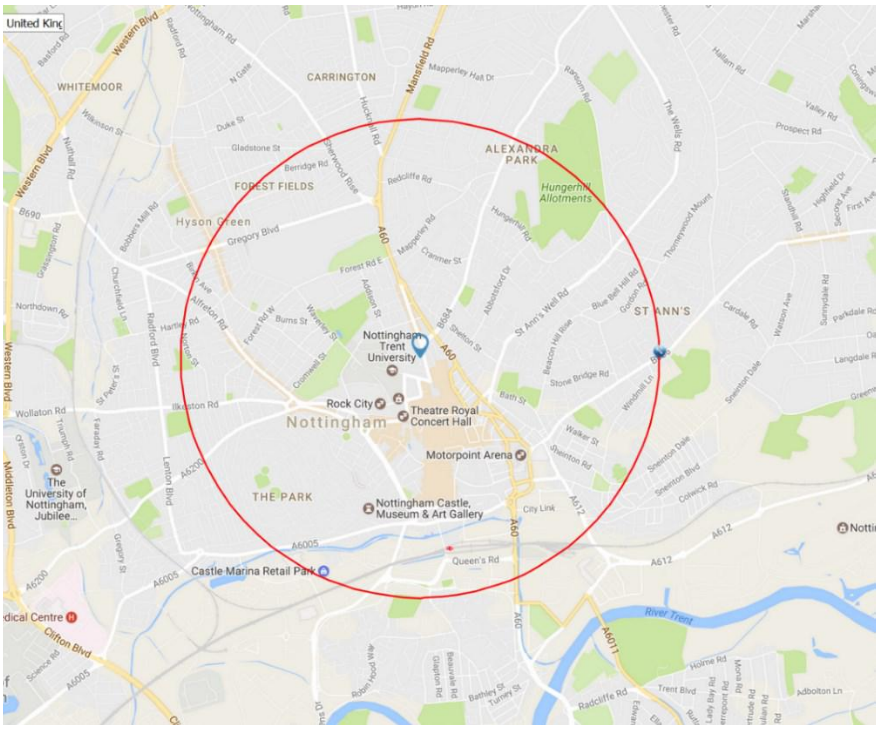



Figure 7-2. Neighbourhoods located within 1 mile of the researcher location, Nottingham Trent University.

Overall, from analysing Table 7-1, it is evident that none of the areas have fulfilled the three criteria. For examples, the Arboretum is the closest area and covered by the 2011 census data, whereas Sneinton, is the furthest but with a wide range of data. However, given the prominence of secondary data in this study, other parameters such as proximity, were compromised. Therefore, the Sneinton area was selected as a pilot area.

The advantage of choosing the Sneinton area as a case study lies in the availability of extensive information on the local residents' socio-economic characteristics and energy consumption data. Moreover, the thermal performance of their dwellings in the form of energy audit reports. This is because the area was selected to accommodate the European project Rumourban future city demonstrator. Over the course of this project, around 3 million pounds will be spent on improving the thermal quality of dwellings of the Sneinton area. This should help the local residents reduce their energy bills; consequently, reduce the CO2 emission levels in the city of Nottingham (Remourban, 2017).

The subsequent section aims to address in detail, the demographic and socio-economic characteristics of this area based on the 2011 and 2001 census data. In the same time, the resident profiles will be compared to those of the city of Nottingham.

Table 7-1. Summary of most suitable areas based on their location and availability of related data

| | Distance from Nottingham City centre | Availability of secondary data | Availability of energy related data from Nottingham Energy Partnership |
|---|---|---|---|
| **Arboretum** | 0.4 mile (9 mins walking) | Yes (census data) | Not for the entire area |
| **Hockley** | 0.6 mile (12 mins walking) | No | No |
| **The park** | 0.8 mile (19mins walking) | Yes (census data) | No |
| **Sneinton** | 1 mile (25 mins walking) | Yes (census data) | Yes |
| **St Ann's** | 1 mile (23 mins walking) | Yes (census data) | Not for the entire area |
| **Forest fields** | 1 mile (24 mins walking) | Yes (census data) | Not for the entire area |
| **Hyson green** | 1 mile (22 mins walking) | Yes (census data) | Not for the entire area |

## 7.4 Demographic and socio-economic characteristics of the Sneinton area

### 7.4.1 Location and brief history

Sneinton is a Nottingham suburb, bounded by Nottingham city centre to the West, Bakersfield to the North, Colwick to the East, and River Trent to the south, as

shown in Figure 7-3. Its population was estimated to be around 13,000 inhabitants in 2011, which is around 5% of the total Nottingham population (OurNottinghamshire, 2015).

This area, which was initially a Saxon settlement, remained a small village until the early 19th Century (industrial era). This is because it witnessed major developments comprising a plan to build houses that accommodate factories and workshops' workers. This local industry was dominated mainly by framework knitting and bobbin net making. However, this expanded into machinery manufacturing, lace making, and malting (OurNottinghamshire, 2015).

Sneinton had a mixed social structure staying in a wide range of houses ranging from back to backs, terraced houses, to genteel houses. However, some older buildings were demolished in the first half of the 20th Century to make way for new ones. In addition to that, in 1941 many of the Sneinton buildings were bombarded in an air raid. A few years later, in 1975, more precisely, the old village of Sneinton, which contains the Green's Windmill, was nominated a conservation area.

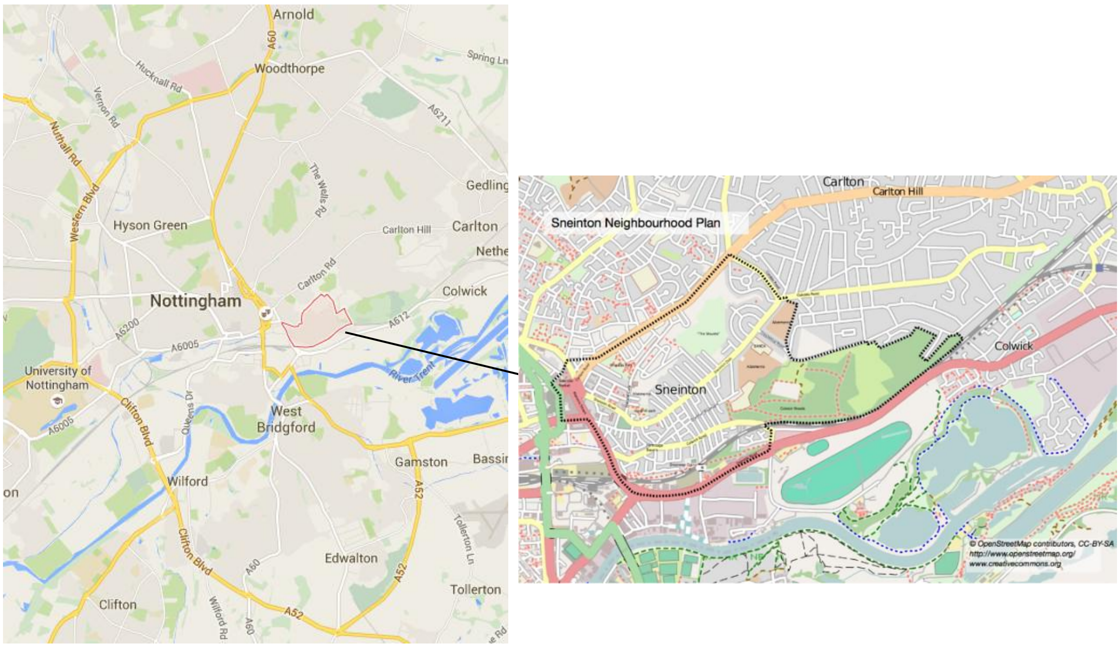
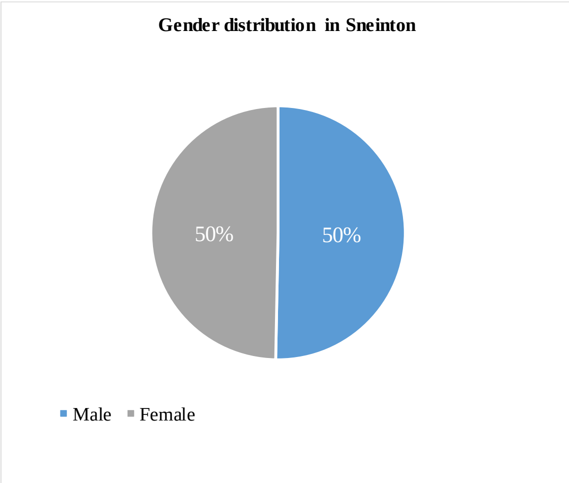



Figure 7-3.Location of the Sneinton area and the surrounding neighbourhood

### 7.4.2 Gender and age distribution in Sneinton

Pie charts in Figure 7-4 outline the gender distribution of Sneinton and the city of Nottingham. Overall, it is clear that the percentage of males in the Sneinton area is equal to the ones of females, which perfectly matches the gender's distribution in Nottingham city.



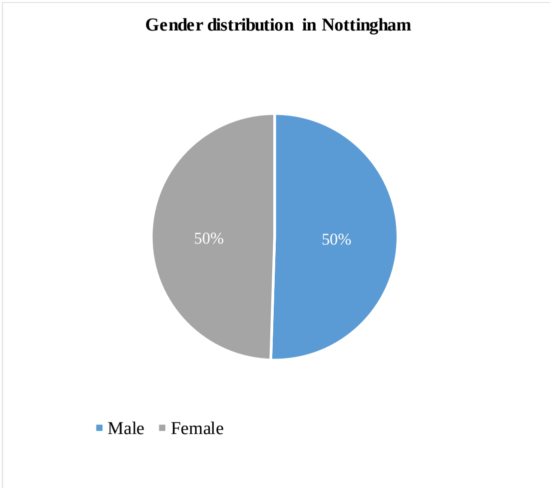




Figure 7-4.Percentage of males and females in Sneinton and Nottingham city

On the other hand, the analysis of the age distribution of the Sneinton area in comparison to the age distributions in Nottingham city and England (Figure 7-5) has highlighted the following points. First, around 60 % of the Sneinton population is aged between 20 and 59-year-old. This was divided into *8.5%* aged 20-24, *12.2%* aged 25-29, *25.9%* aged 30-44, and *14.8%* aged 45-59. However, there were discrepancies of 5-6% between the average percentage of age groups in Nottingham and England. For example, the percentage of residents aged between 20 and 24 was about 2 % higher than the one of England but 5% lower than the one of Nottingham city. This is because 15% of Nottingham city residents are full-time University students aged 18 and over (Nottingham Council, 2016).

The proportions of children aged between 0-4 years and 5-7 years were 3% and 0.5% higher than the ones of Nottingham city and England. Given the impact of children presence on energy consumption, higher residential energy figures should be expected in the Sneinton area.

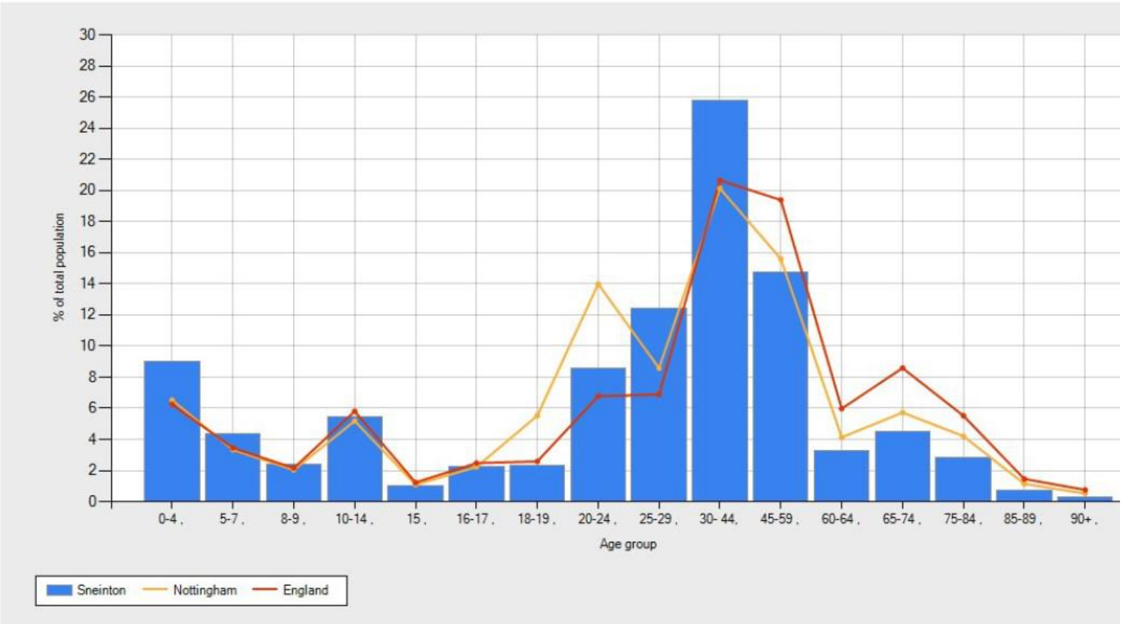
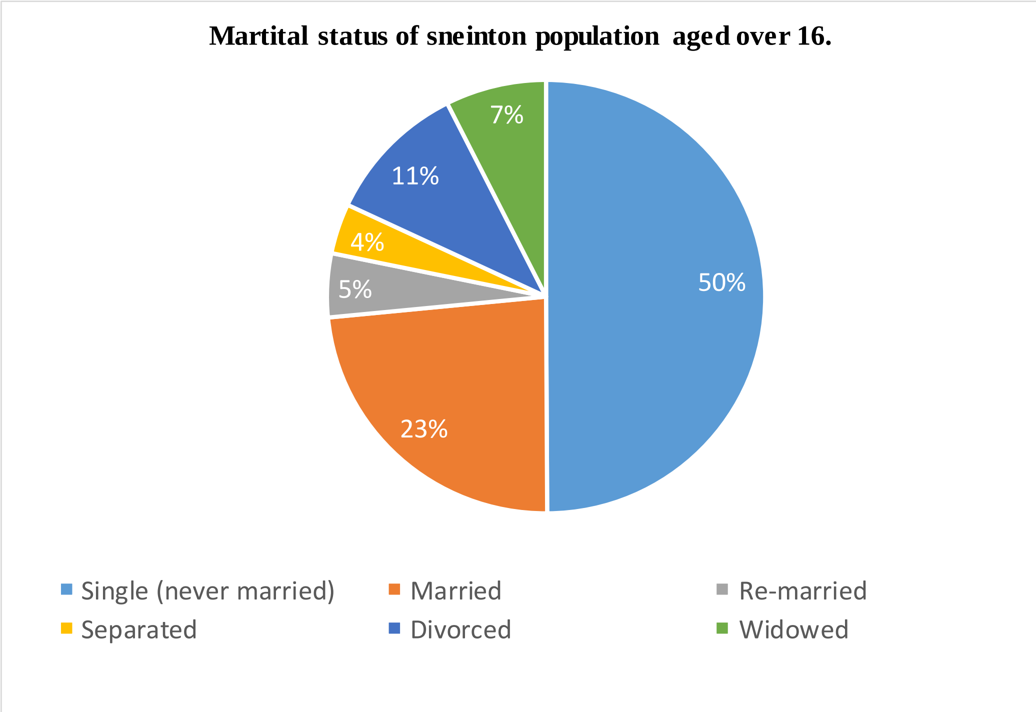



Figure 7-5.Distribution of age groups in Sneinton, Nottingham, and England

### 7.4.3 Marital status

Figure 7-6 (below) illustrates the marital status of the Sneinton and Nottingham city population aged 16 years and over. Overall, it is evident that the majority of the Sneinton residents were single never married (50%), which was around 6% higher than the proportion of the ones in Nottingham city. However, there was a 7.9% discrepancy between the percentages of married of both locations. For example, Sneinton had 23% married, whereas Nottingham city, had 30.9%. In contrast, the proportions of the remaining marital statuses namely; divorced (11%), widowed (7%), re-married (5%), and separated (4%), were in good agreement with the ones of Nottingham city residents.

In the light of the above, it is clear that the marital status profile of the Sneinton area is to a large extent representative of the one of Nottingham city. However, given the dominance of single occupancy in the Sneinton area, its average energy usage, especially electricity, is expected to fall below the national average.



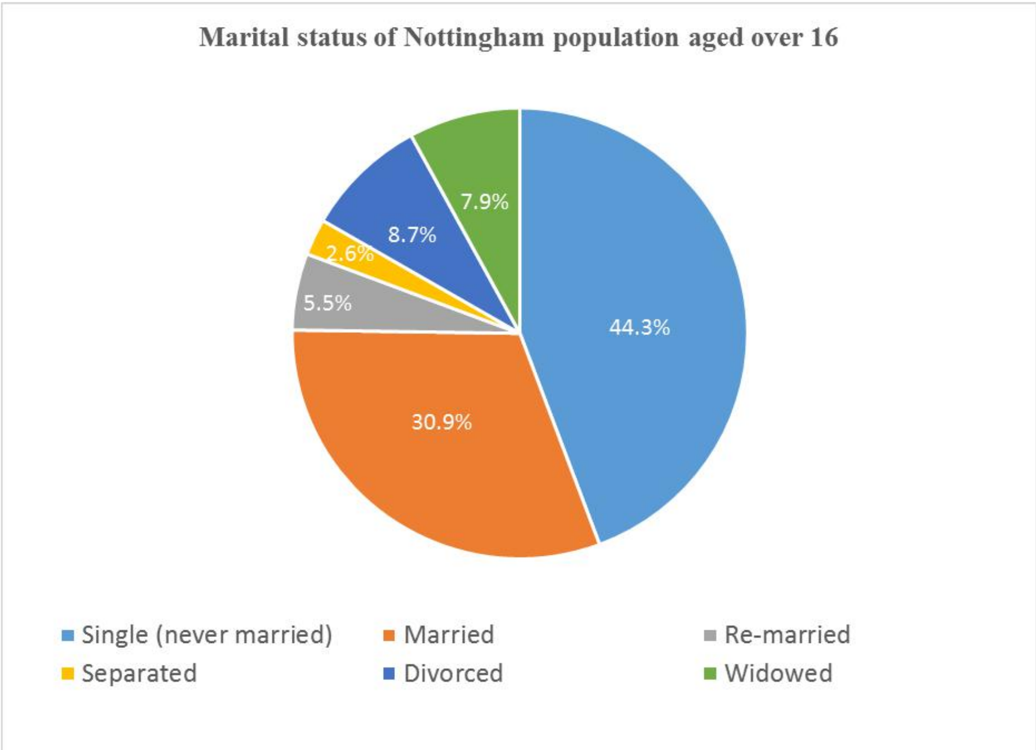
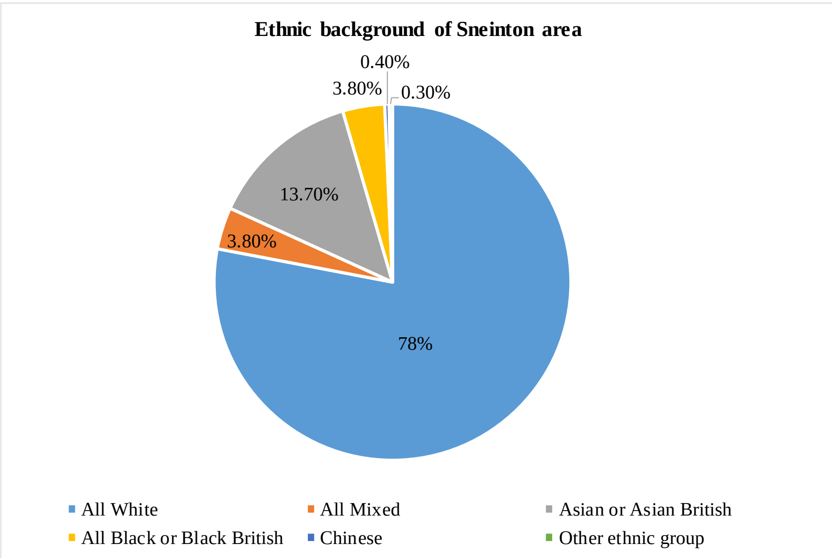




Figure 7-6.Marital status of Sneinton and Nottingham city population aged over 16, respectively.

### 7.4.4 Ethnic background

Figure 7-7 depicts the ethnic profile of Sneinton and Nottingham city. In general, the majority of the Sneinton residents are white 78% of which 72.8% are white British. This figure was 7 % higher than the one of Nottingham city, where white residents constituted 71.1%. Similarly, the percentages of Black British or Black (3.8%), as well as mixed ethnicities (3.8%), were around 3% below the ones of Nottingham city. However, the percentages of Asians or Asians British (13.7%) as well as other ethnicities (0.75%), were almost identical over both locations.

After analysing the above facts, it is evident that the ethnic profile of the Sneinton area is close to the one of Nottingham city.



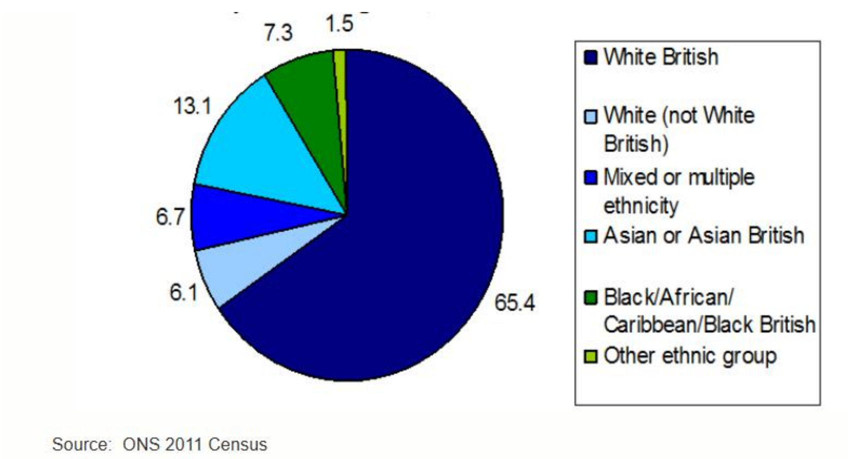




Figure 7-7.Ethnic profile of the Sneinton area and Nottingham, Correspondingly (census 2011)

### 7.4.5 Household type

Figure 7-8 depicts the households' composition of the pilot area (Sneinton) in relation to Nottingham city occupants' structure. Overall, both Sneinton and Nottingham city were dominated by a high proportion of single households aged 64 or below, 32% and 24.6%, reciprocally. As discussed at an earlier stage, this is due to the high proportion of University students living in this city (around 15% of the population). This figure was followed by 15.6% of other households, 14.4% of couples with dependent children, and approximately 11% of lone parents with dependent children and couples no children, each. In comparison to the percentages of those household types in Nottingham city, there were minor discrepancies of 1 to 3%.

After analysing the above facts, it is clear that household type distribution of Sneinton area is to a great extent representative of the family structure of Nottingham city.

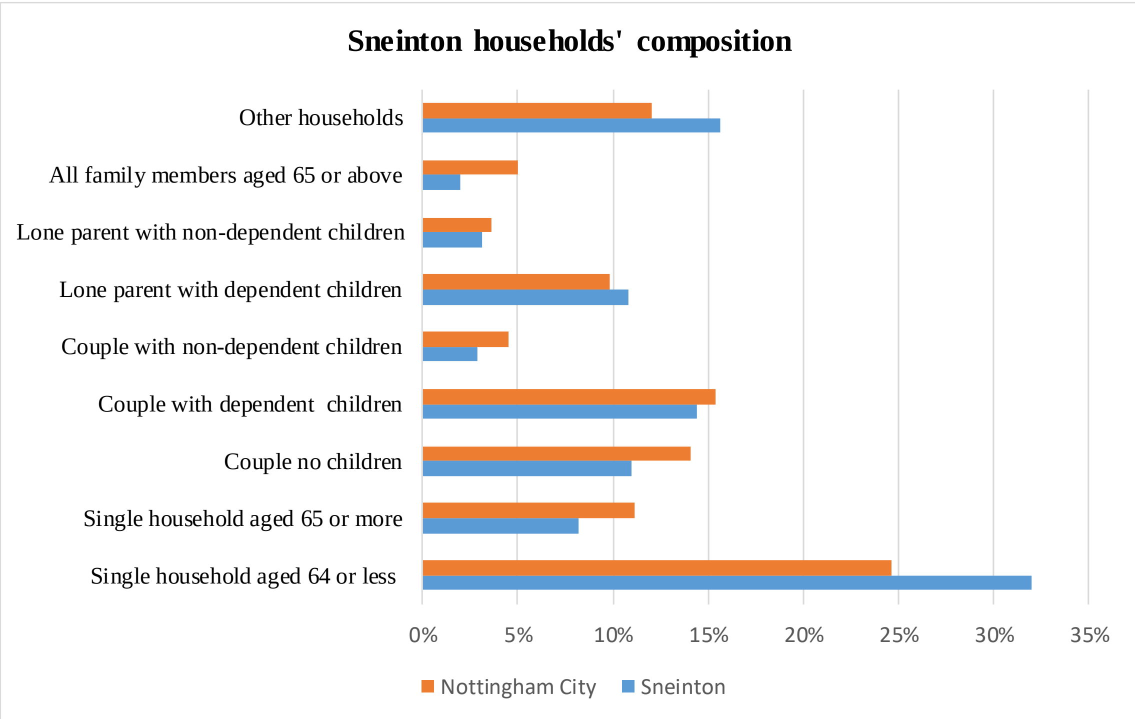



Figure 7-8. Distribution of household types in Sneinton and Nottingham city

### 7.4.6 Types of dwelling and housing tenure

Figure 7-9 and Figure 7-10 represent the dwelling and the tenure types of Sneinton and Nottingham city, correspondingly. From analysing the dwelling types of the Sneinton area, it is clear that it is characterised by a high residential density. This is because of the large proportion of terraced houses (49.40%), which is about 25% higher than in Nottingham city. This was followed by 21% and 20% of semi-detached and purpose built flats, correspondingly, which was around 10% lower than the proportions of Nottingham City. Conversely, the figures of converted flats (3%) and other dwelling types (1.2%) in Sneinton were in good agreement with the ones of Nottingham city.

As for the tenure mode, it is clear that most Sneinton households (30.3%) lived in privately rented accommodations, which was in disagreement with Nottingham city. This is because this latter was characterised by the dominance of home owners with mortgage who represented around 25%. This figure was followed by around 20% renting from local authorities both in the Sneinton area and Nottingham city. However, the percentage of dwellings rented from housing associations in Sneinton was about 6% higher than in Nottingham city. Conversely, the proportion of dwellings owned outright (around 13%) was 7% lower than the one of Nottingham city. These discrepancies might be an indication of the difference in income/savings levels between the residents of both areas. Indeed, Ball (2010) argued that one of the reasons behind the increasing demand for private housing is the unstable or low-levels of income of occupants that prevent them from becoming homeowners. The authors also stressed the importance of current stage of household life-cycle, since it indicates the household saving level.

In the light of the above, it is evident that Sneinton area is not representative of Nottingham city dwelling type and tenure mode distributions.

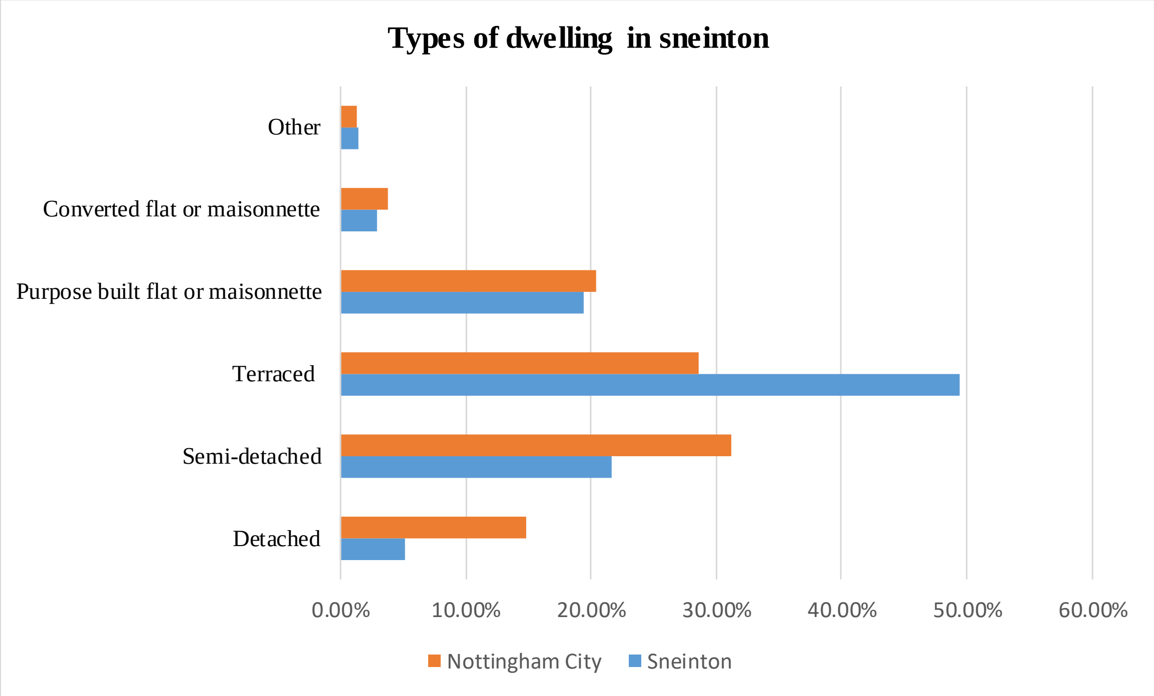



Figure 7-9. Percentage of dwelling types in Sneinton in comparison to Nottingham city

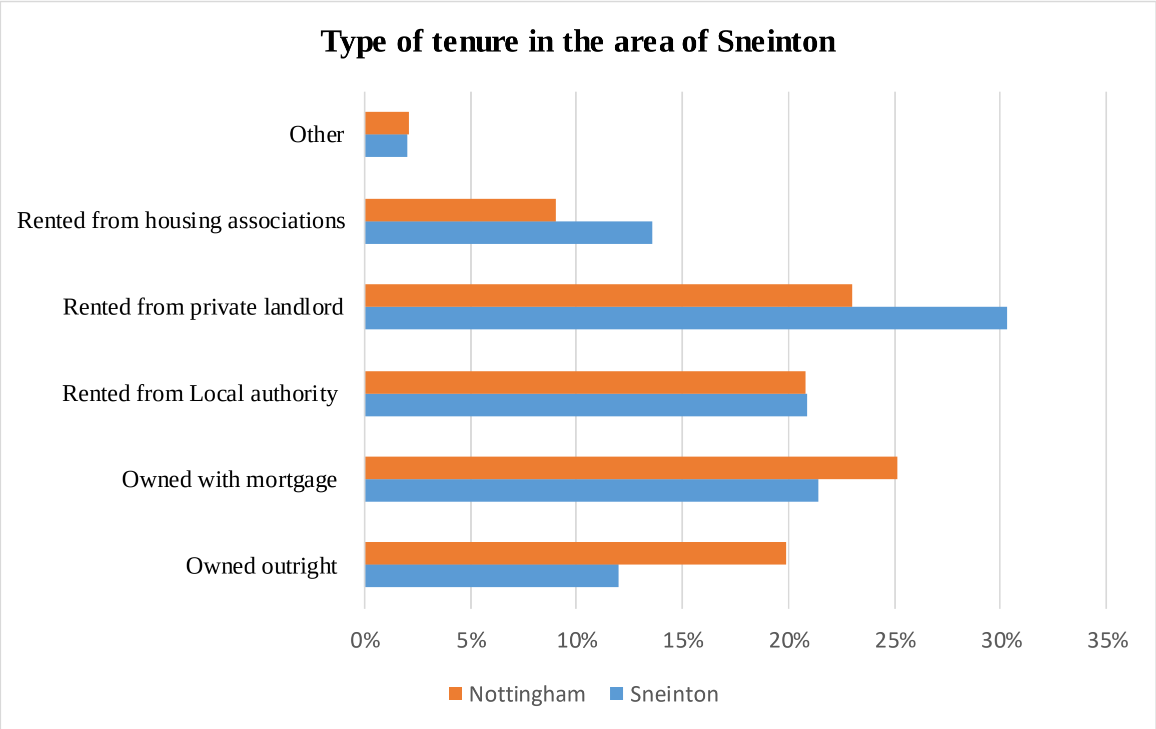



Figure 7-10. Tenure type in Sneinton and Nottingham city

Figure 7-11. Terraced is the most dominant dwelling type in the Sneinton area

### 7.4.7 Dwelling Typologies

As addressed in chapter two, the majority of research on energy sustainability in the building sector create typologies based on two key factors; building age and type (InSMART, 2013). In the UK, Cities Revealed (CR) is the only data source that provides both dwelling age and type under one database. Considering the most common dwelling age bands in the city of Nottingham, six age groups can be identified. These are;

- Pre-Victorian and Victorian (Pre 1914)
- Inter-war period (1914-1945)
- Post-war era (1946-1964)
- Sixties/ Seventies houses (1965-1979)
- Modern pre-2000 (1979-2000)
- Modern Post 2000 (Post 2001).

The area of Sneinton comprises 16 dwelling typologies namely; Victorian flat, Victorian terrace, Victorian semi-detached, inter-war semi-detached/terraced, inter-war detached, post-war semi-detached/terraced, post-war detached, 60s/70s flats, 60s/70s terrace, 60s/70s semi-detached, 60s/70s detached, post-1979 flats, post-1979 terrace, post-1979 semi-detached, post-1979 detached.

Table 7-2 represents the distribution of the above typologies in the area of Sneinton. Overall, it is evident that the percentage of Victorian terraced houses was the highest by almost 33%, whereas the proportion of Victorian detached, was the lowest (0.48%). Interwar semi-detached/terraced houses had also an important share in the area by 25.11%. This figure was followed by interwar detached (11.52%) and 60s/70s flats (9.13%). On the other hand, the proportion of the remaining typologies ranged between 0.48% and 3.95%. For example, the proportions of Victorian flats and 60s/70s terraced were 3.95% and 3.53%, correspondingly.

As for the wall structure of these typologies, it is evident that most of them possess a brick masonry cavity wall. However, there was an exception for Victorian dwellings under the typologies T1, T2, T3, T4 and which have a solid brick wall structure. However, 60/70 flats possess a brick and concrete wall structure.

Regarding the roof type of these typologies, the majority of them have sloped roof. However, flat roofs could also be encountered in Victorian flats, 60s/70s dwellings, and post 1979 flats. Furthermore, shed roofs can be found in 60s/70s terraced houses.

After analysing the above facts, it is evident that the majority of dwelling stock at the Sneinton area have a poor thermal quality. This, in turn, should have a negative impact both on households' health and their energy consumption figures, especially during winter. This justifies why the Rumourban project mostly concentrated on dwellings' retrofit in the Sneinton area.

Table 7-2. Distribution and basic physical characteristics of existing typologies in the Sneinton area

| | Typology | Percentage in the Sneinton area | Year of constructio n | Number of floors | Wall type | Roof type |
|---|---|---|---|---|---|---|
| **Victorian flat** | T1 | 3.95% | Pre 1914 | 3-5 | Brick/Stone (no cavity) | Flat/sloped |
| **Victorian terraced** | T2 | 32.96% | Pre 1914 | 2-3 | Single brick | Sloped |
| **Victorian semi-detached** | T3 | 0.64% | Pre 1914 | 2-3 | Single brick | Sloped |
| **Victorian detached** | T4 | 0.48% | Pre 1914 | 2-3 | Single brick | Sloped |
| **Interwar semi-detached/terra ced** | T5 | 25.11% | (1914-1945) | 2 | Cavity wall | Sloped |
| **Interwar detached** | T6 | 11.52% | (1914-1945) | 1-2 | Cavity wall | Sloped |
| **Post-war semi-detached/terra ced** | T7 | 0.64% | (1946-1964) | 1-2 | Cavity wall | Sloped |
| **Post-war detached** | T8 | 1.82% | (1946-1964) | 1-2 | Cavity wall | Sloped |
| **60s/70s flats** | T9 | 9.13% | (1965-1979) | 2-20 | Concrete/ Brick | Flat/sloped |
| **60s/70s terraced** | T10 | 3.53% | (1965-1979) | 2-3 | Cavity wall | Flat/sloped/she d |
| **60s/70s semi-detached** | T11 | 2.18% | (1965-1979) | 2 | Cavity wall | Flat/sloped |
| **60s/70s detached** | T12 | 0.99% | (1965-1979) | 1-2 | Cavity wall | Sloped |
| **post-1979 flats** | T13 | 3.60% | post-1979 | 2-10 | Cavity wall/concre te panel | Flat/sloped |
| **post-1979 terraced** | T14 | 0.81% | post-1979 | 2 | Cavity wall | Sloped |
| **post-1979 semi-detached** | T15 | 2.1% | post-1979 | 1-2 | Cavity wall | Sloped |
| **post-1979 detached** | T16 | 0.55% | post-1979 | 2 | Cavity wall | Sloped |

### 7.4.8 Economic profile

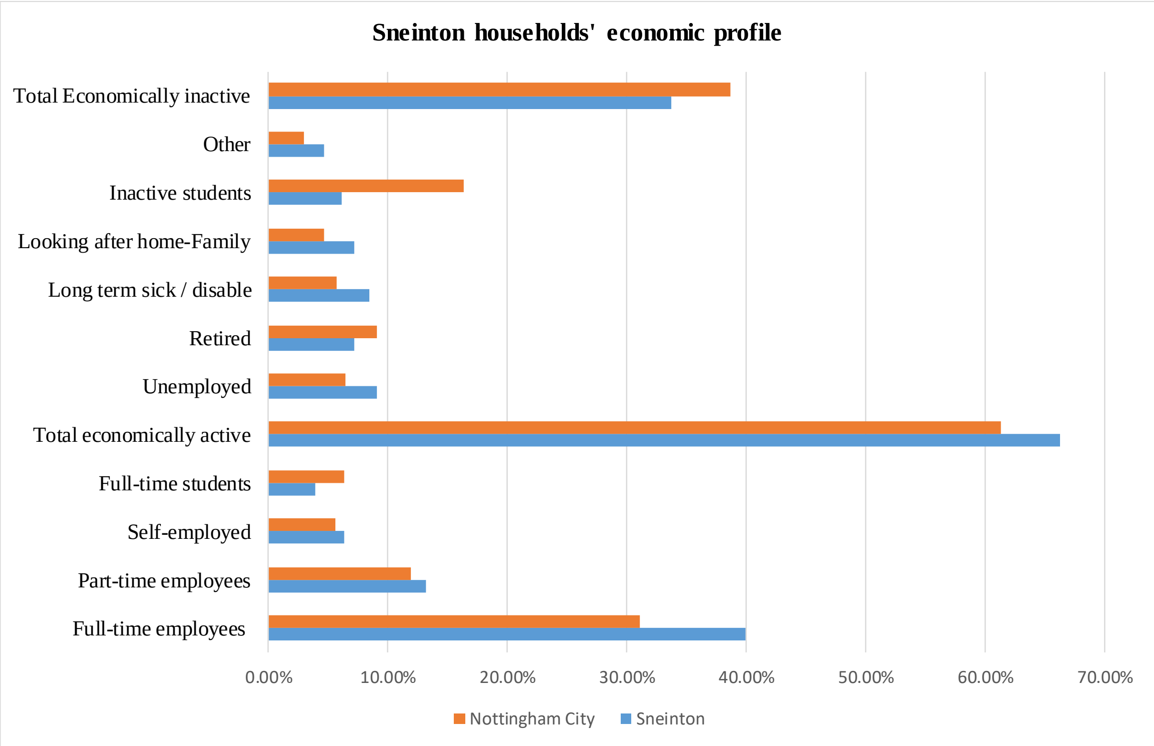



Figure 7-12.Socio-economic profile of Sneinton residents in comparison to Nottingham city inhabitants

Figure 7-12 outlines the economic profile of Sneinton residents and contrasts it with the one of Nottingham city. Overall, it is clear that the majority of Sneinton households (66.3%) are economically active, which was approximately 5% higher than the figure of Nottingham city but lower by 12.2% than the UK one. Thus, the Sneinton area is excepted to contain households living in poverty. Amongst the economically active Sneinton households, 40% were in full-time employment, 13.2% working part-time, and 6.3% self-employed. Except for full-time economically active students, these figures were in good agreement with the ones of Nottingham city.

In contrast to the above, there was 33.7% of economically inactive households, which was around 5% lower than the figure of Nottingham city. This overall proportion is shared almost equally between retired (7.2%), disable or long-term sick (8.4%), looking after home family (7.85%), inactive students (6.1%), and unemployed households (9.1%).

### 7.4.9 Qualifications and socio-economic classes

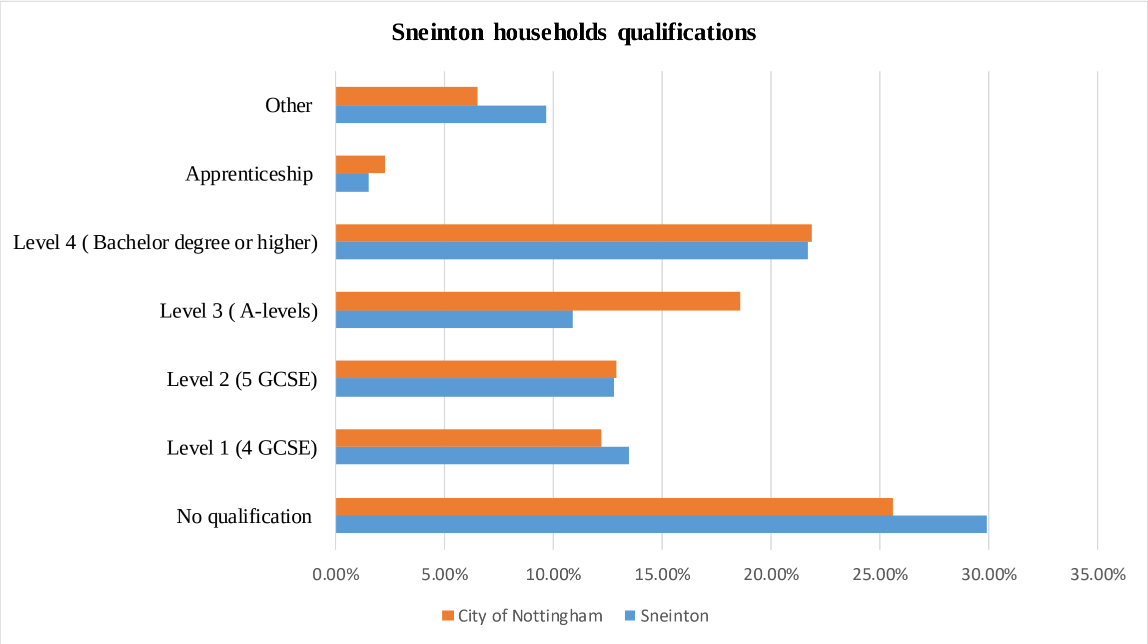



Figure 7-13. Sneinton households' level of education in comparison to the one of Nottingham city residents.

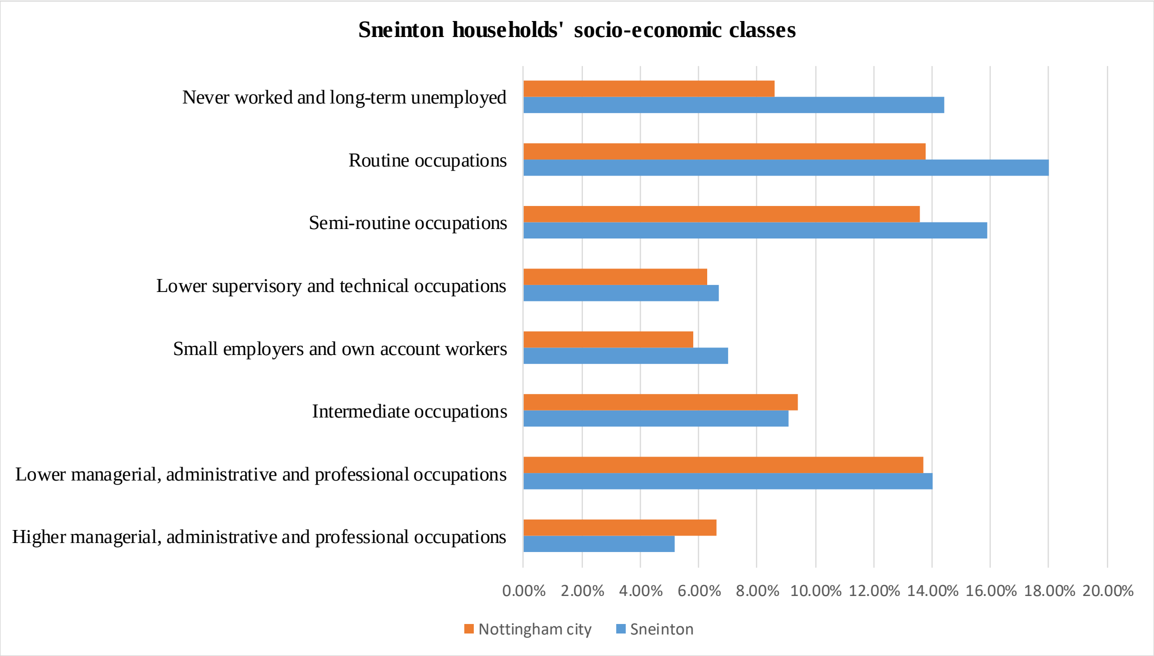



Figure 7-14. The socio-economic profiles of Sneinton and Nottingham city households.

Figure 7-13 and Figure 7-14 represent the Sneinton households' qualifications and socio-economic classes, respectively. In general, the majority of the pilot area residents, approximately 30%, had no formal qualifications. This justifies why a large proportion of them (around 50%), either performed routine occupations (e.g. cleaner), semi-routine (e.g. Shop assistant) or even have never previously worked. Larger discrepancies of up to 10% were observed between these figures and the ones of Nottingham city.

On the other hand, 21.7% had level 4 qualifications (bachelor degree or above) which are believed to be mostly split between "Higher managerial, administrative and professional occupations" and "Lower managerial, administrative and professional occupations" socio-economic classes. In contrast to the above statistics, a good agreement was found between these figures and the ones of Nottingham city.

## 7.5 Development of the urban 3D semantic model

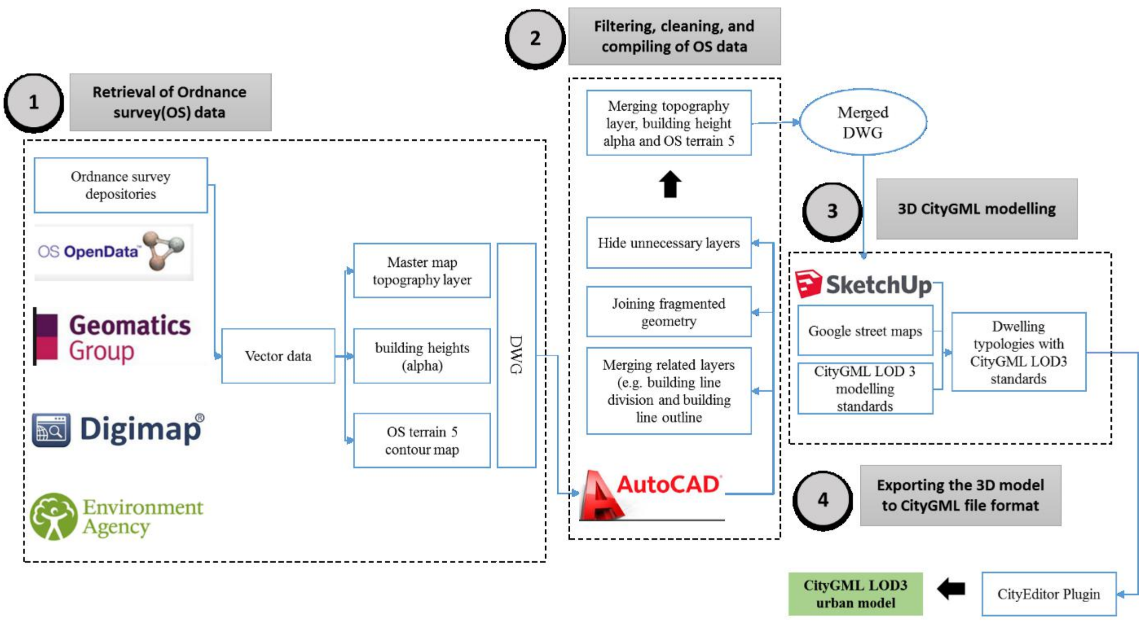



Figure 7-15. strategy of the initial semantic urban energy modelling

The analysis of chapter two has unravelled the modelling principles of the 3D urban energy semantic model of the pilot area. The developed strategy takes into account all necessary stages to create an integrated a 3D urban energy prediction tool. These comprise the following; data collection, CAD/CityGML modelling, energy simulation, socio-economic modelling, validation, and visualisation. For more details on the developed modelling framework, please refer to chapter two.

This section; however, aims to address in more depth a prominent aspect of this framework, which is urban 3D semantic modelling. This latter does not only spot light on the modelling pipeline of 3D CityGML LOD3 energy models but also consider the associated technologies (e.g. software, hardware, and database). However, given its importance in the undertaken research, urban 3D Semantic modelling would be briefly defined while highlights its importance in urban planning.

### 7.5.1 Urban 3D semantic modelling and its relevance

In linguistics, semantics refer to the study of meanings. However, in software engineering, a semantic model is a conceptual model which describes the meaning of its instances (data items) and reflects the relationship between them (Koehl and Grussenmeyer, 2000). Semantic models are adequate for making the relationship between the stored data items and real world not only explicit but also meaningfully interpreted with minimal or even without human intervention (SEMAGIX, 2009). For simplification purpose, let us consider a database composed of two items; a table and a cup. The corresponding semantic data model should contain three parts namely; cup, table, and their relationship which is cup sitting on table. Thus, both objects are interpreted in function of their relationship.

In the context of this research, urban 3D semantic modelling stands for the process of representing the thematic and spatial attributes of urban entities and the relationship between them (Chaturvedi et al., 2017). Building volumes, area, and perimeter, are examples of spatial attributes. On the other hand, building name, function, usage, wall and roof types, are examples of thematic attributes.

In urban planning, 3D urban semantic models can be effectively utilised to support the decision-making process as well as promote collaboration and data integration (Chaturvedi et al., 2017). This is evident in the growing adoption of CityGML international standards in the 3D semantic modelling of many cities worldwide such as Vienna and Berlin. Moreover, the initiatives to integrate the 3D semantic modelling approach  used at the building scale, IFC models more precisely, with those of urban scale (CityGML models) (Kardinal Jusuf et al., 2017).

Unlike 3D CAD models, 3D semantic  models do not only allow the development of applications supporting various aspects of urban planning (e.g. disaster management and environmental simulation) but also promote interoperability of shared data and domain knowledge models (Howell et al., 2017). As such, this facilitates the exchange of information pertaining to urban entities with users from different disciplines or across other application domains (El-Mekawy et al., 2011).

In the subsequent sections, the process of urban 3D semantic modelling, which comprises four distinct phases (Figure 7-15), will be discussed.

### 7.5.2 Ordnance survey (OS) data retrieval

To begin with, obtaining geospatial data is without any doubt the first step towards developing a semantic urban model supporting energy planning decision-making. One of the benefits of such data is **geolocation**, which, in turn, enables an accurate positioning of dwellings on site and also performing semantic querying. Geospatial data is in turn classified into two distinct categories namely; raster and vector (Figure 7-16). However, vector data were preferred because they have a better representation of reality given their high accuracy. Furthermore, they offer the ability to alter the scale of analysis and observation, and their file size is smaller than raster data (OrdnanceSurvey, 2017). Finally, vector data facilitates the spatial analysis process and allows flexible manipulation of different entities such as layer management and graphic override. For more information on the difference between raster and vector data, please refer to appendix A.A.1.

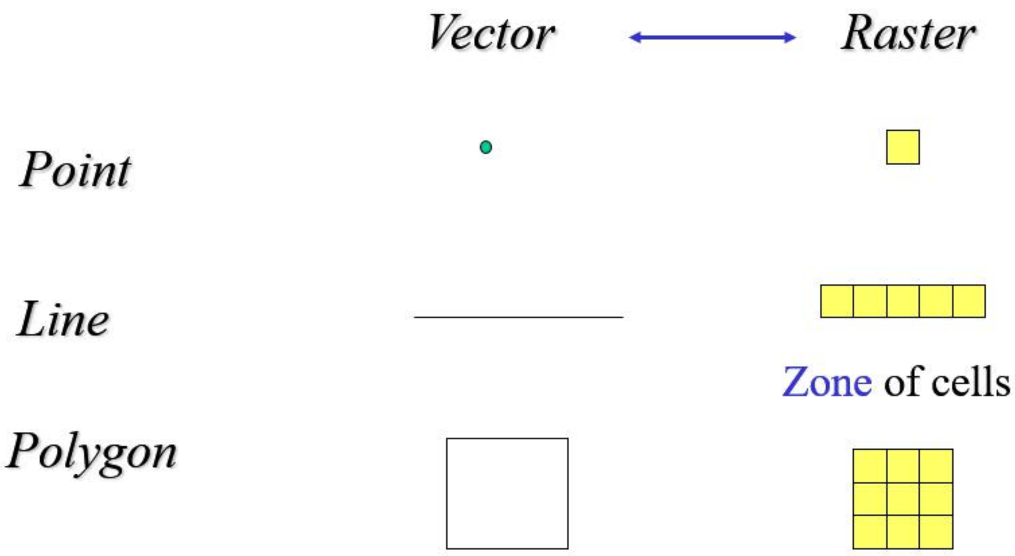



Figure 7-16.Main differences between raster and vector data (Fotheringham and Rogerson, 2013)

In the UK, many geospatial data repositories are offering both OS vector and raster data such as ShareGeo Open, Geomatics Group LiDAR, and Edina Digimap. However, the latter was chosen due to its long tradition of providing ordnance survey map to tertiary education (since 2000), including Universities. Furthermore, Edina Digimap group are partners with national agencies such as Ordnance Survey and British Geological Survey (BGS), to provide high-quality maps and spatial data.

After visiting the Digimap website and choosing to download ordnance survey data, a UK map with different OS data types will appear (Figure 7-17). The following step involves querying for the envisaged area using the search bar via keywords or postcode. Once attained, the area boundaries can be defined by using; a rectangle selection tool, easting and northing coordinates, a tile name, or a screenshot of the visible area. Next, the type of OS data needs to be selected before downloading files. In the current research, three types of OS data namely; Master map topography layer, building heights (alpha), and OS terrain 5 contour map, are relevant. The topography layer map encompasses features that represent the physical environment such as buildings, paths, roads, railways, and rivers. On the other hand, Building height alpha consists of height attributes for each building in OS Master map topography layer. OS terrain contour 5 represents a contour dataset, where the interval between ordinary contour lines is 5 meters. Those maps can be downloaded in various formats including, SHAPE, GML2, GML3, KML, and DWG. However, DWG file format was chosen due to its high interoperability across different platforms such as game engines and GIS software packages (Table 7-3).

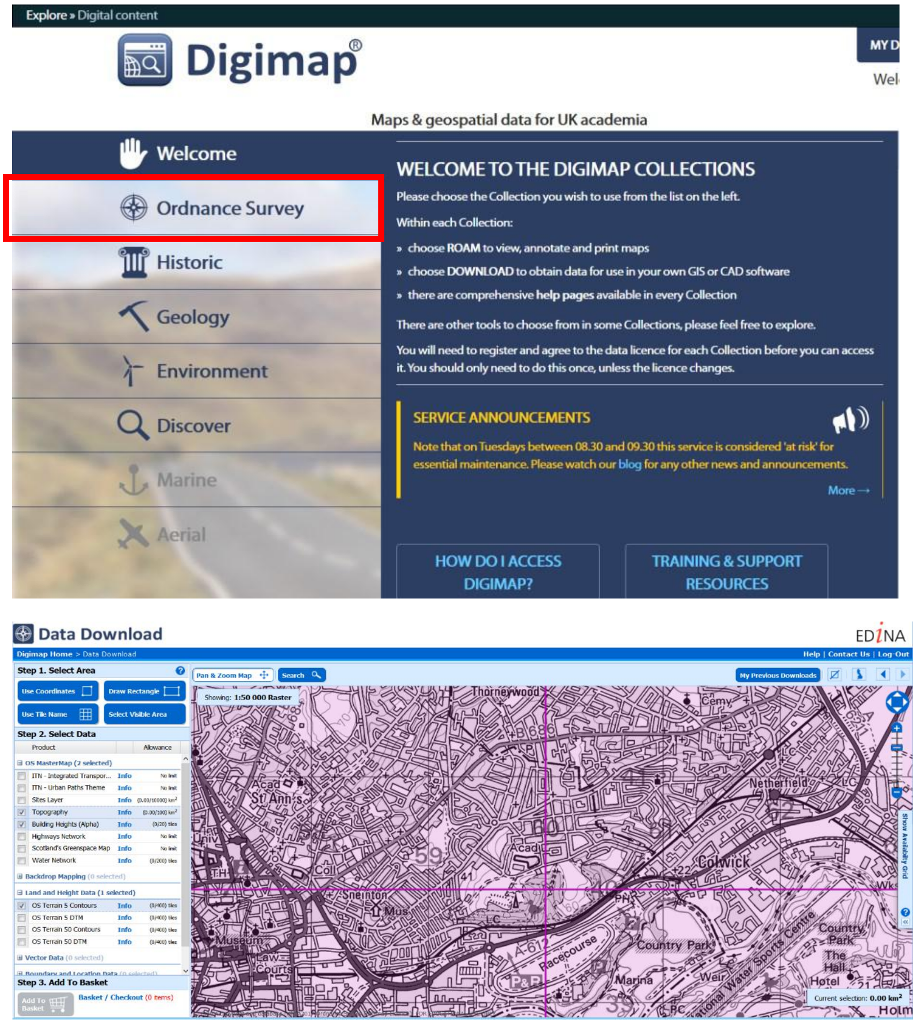



Figure 7-17. The main components of Digimaps ordnance survey module.

Table 7-3. Export file formats offered by Digimaps and their supporting software packages.

| | Stands for | Category | Supported software |
|---|---|---|---|
| **KML** | Keyhole Markup Language File (stored in XML) | GIS | -Google earth<br>- ESRI ArcGIS<br>- Blender<br>-Merkaartor<br>-Keyhole PRO |
| **DWG** | AutoCAD drawing file | CAD, CAM, CAE | plenty of programs including;<br>-Adobe illustrator<br>- ArchiCAD<br>-Revit<br>-Sketchup<br>-Rhinoceros<br>- ESRI ArcGIS |
| **GML** | Geography Markup Language File | GIS | -Merkaartor<br>-TatukGIS Viewer<br>- ESRI ArcGIS<br>- Galdos Systems GML SDK |
| **SHAPE** | N/A | GIS | -Mainly GIS software packages including;<br>-ESRI ArcGIS<br>-QGIS<br>-Terra Explorer Pro |

### 7.5.3 Filtering, cleaning, and compiling of OS data

Since the imported maps contain a lot of information, Autodesk AutoCAD was utilised to filter and clean unnecessary information. The main reasons for adopting this software are its support for GIS data and powerful layer management capabilities. Moreover, its ability to export to FBX, DWG, and DXF formats, which have a high interoperability with different platforms, including games engines.

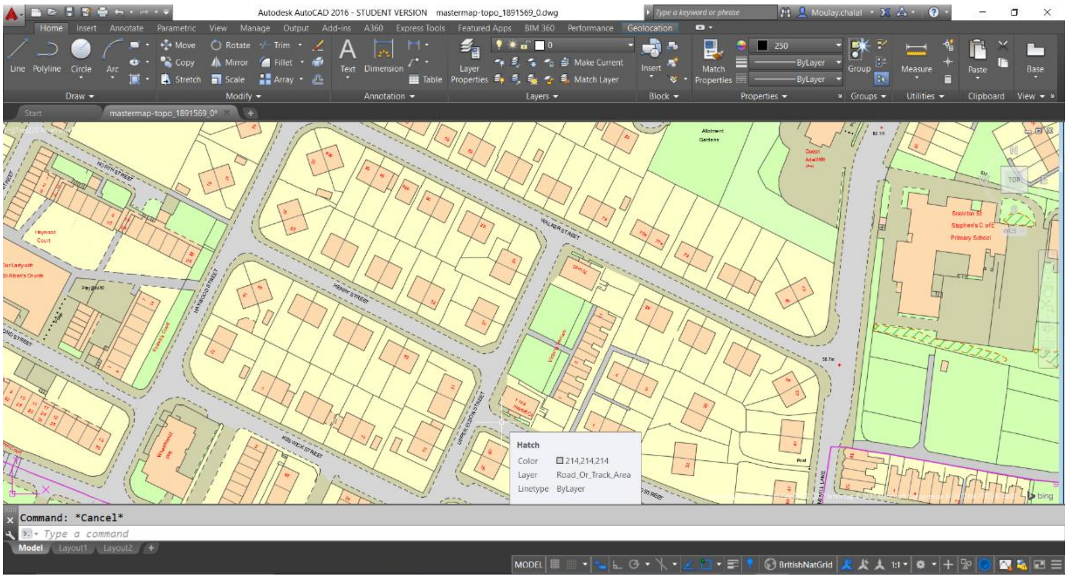


Figure 7-18. OS master map topography layer before being cleaned with AutoCAD

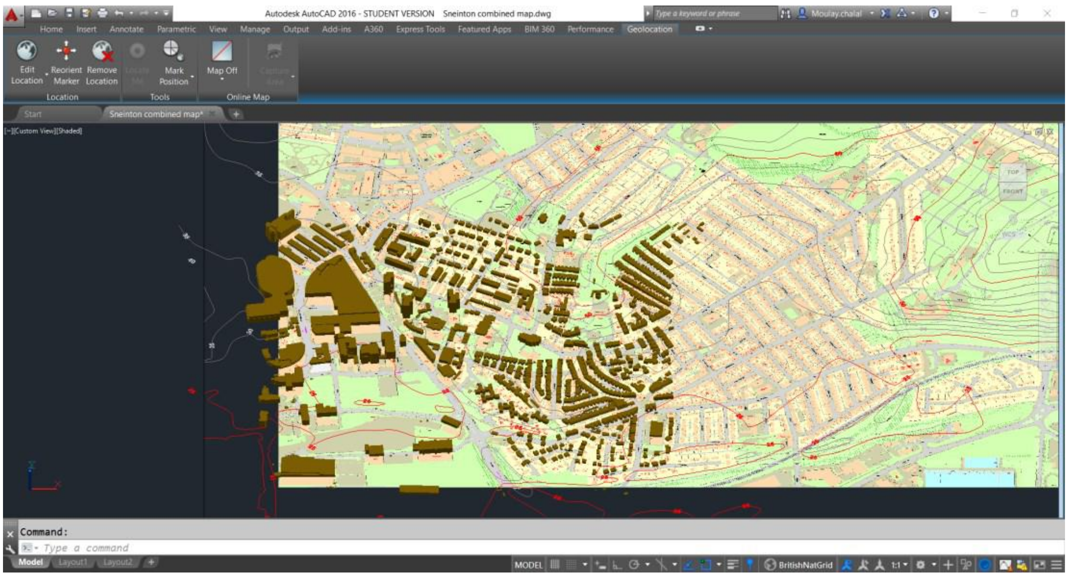


Figure 7-19. A snapshot illustrating the process of merging the topography layer with the Building Alpha and OS terrain contour 5 layers in AutoCAD

Figure 7-18 illustrates OS Master map topography layer of the Sneinton area before being cleaned. First, building line division and building line outline layers were merged under one layer named; building footprint. Secondly, road lines, which were initially fragmented, were joined together to facilitate their draping (projection) onto the area toposurface. Thirdly, the 3D buildings were imported into the OS Master map topography layer file. This was achieved by opening up Building height Alpha, copying them from an origin (e.g. 0,0,0), and pasting them into the OS Master map topography layer. Similarly, the OS terrain contour 5-layer file was imported into the OS Master map topography layer following the same approach (Figure 7-19). Finally, since 3D building height alpha layer contains some discrepancies, as indicated by Edina, some of the buildings were manually adjusted on the map.

### 7.5.4 CityGML modelling

Once the three maps were combined into one single DWG file, the CityGML modelling occurs. For more details on CityGML LOD3 structure and classes, please refer to appendixA.1.3.

Apart from its geolocation support, Sketchup was adopted since it is supported by many CityGML modelling plugins such as CAD to CityGML by 3DIS (2016). In addition to this, it gives the possibility to download realistic maps directly from Google street view and apply them into different dwelling elevations. Given the limited time and resources in this doctoral project, these maps will in turn permit a fast and cost-effective modelling of building elements, including; doors, windows, and even roofs.

The process began by importing the merged file into Sketchup (Figure 7-20). After that, a location was added, and the generated topography map was aligned with the existing one (Figure 7-21). Following the above step, the next challenge consisted of transforming a mass model, which is partially equivalent to CityGML LOD1, into a LOD3 model. Therefore, the subsequent layers namely; roof surface, ceiling surface, wall surface, building installation, windows, doors, floor surface, ground surface, interior wall surface, outer floor surface, were first created. Once accomplished, dwellings with similar typologies were grouped together with the help of Google street maps to speed up the modelling process. After that, the street elevations of each typology were prone to a realistic texture mapping, as shown in

Figure 7-22 . On the other hand, textures of other elevations, which are not covered by Google street maps, were obtained from photos taken during a site visit.

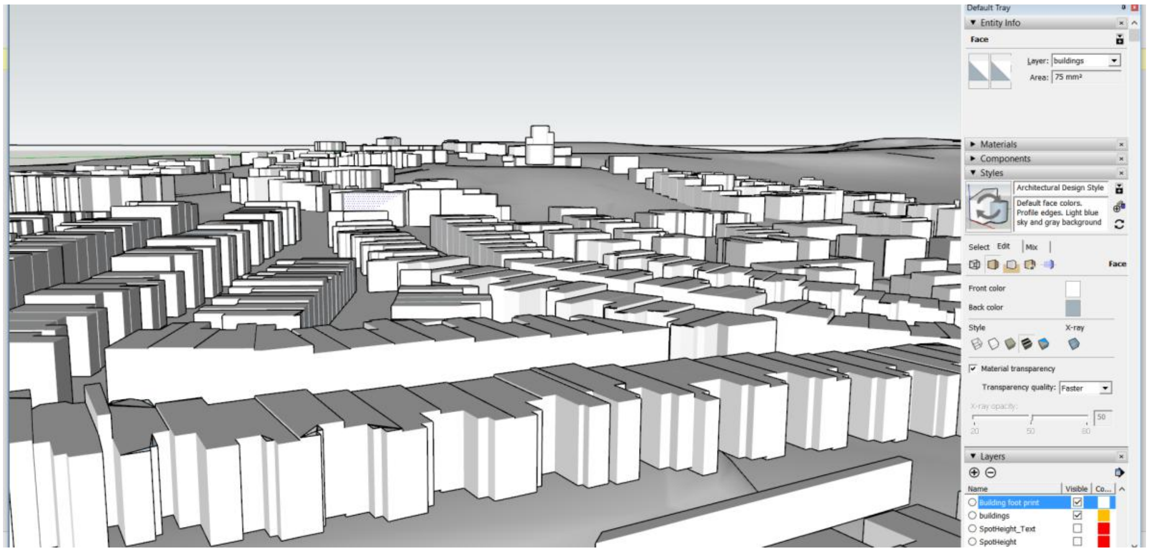


Figure 7-20. A Sketchup snapshot showing the imported merged AutoCAD file

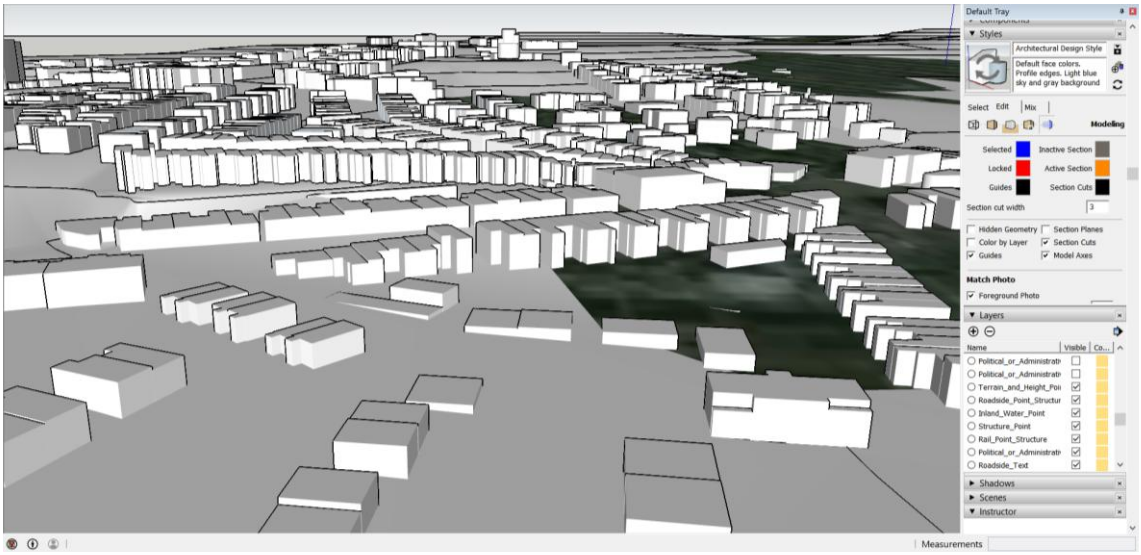


Figure 7-21.A Sketchup snapshot showing the process of aligning Sketchup toposurface with the ones of the imported model.

Once all textures were mapped onto the elevations of each typology, windows and doors were created. Furthermore, roofs were generated using the 1001bitdtandard plugin in Sketchup (1001bitStandard, 2017). Following the CityGML LOD3 modelling guide, which was developed by SIG3D (2014), roof, wall, and ground surfaces were offset, as shown in Figure 7-24 . Moreover, ceiling, floor, and ground surfaces were created. Any elements belonging to the *BuildingInstallation* class such as chimneys, balconies, and conservatories, were also created. Once the modelling of all building components was achieved, each element was assigned to the corresponding layer. For example, outer roof surfaces were assigned to the "roof surface" layer, whereas the inner ones, were allocated to the "ceiling surface" layer. Similarly, inner wall surfaces were appointed to the layer "interior wall surface". However, the outer ones were assigned to the layer "wall surface". Finally, door steps were allocated to the layer "outer floor surface". Figure 7-24 illustrates the different CityGML LOD3 layers of an existing typology in Sneinton.


Figure 7-22. Realistic texture mapping using Google Street View and Sketchup

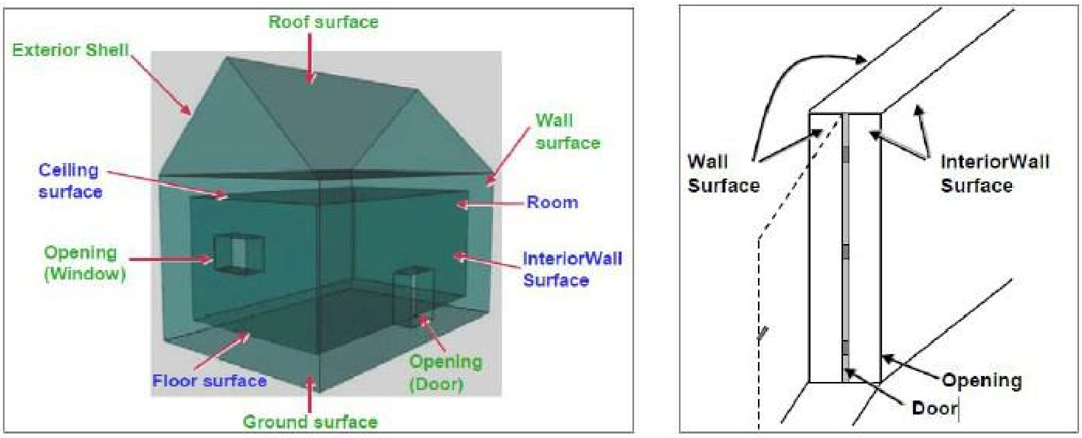
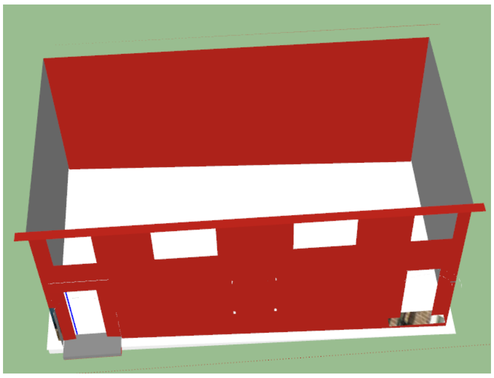
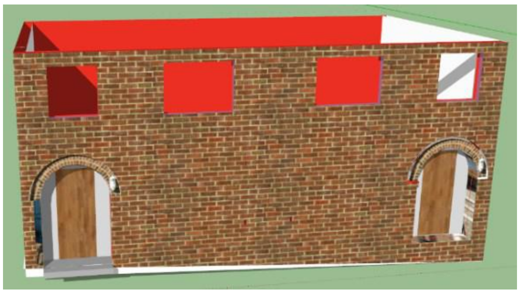


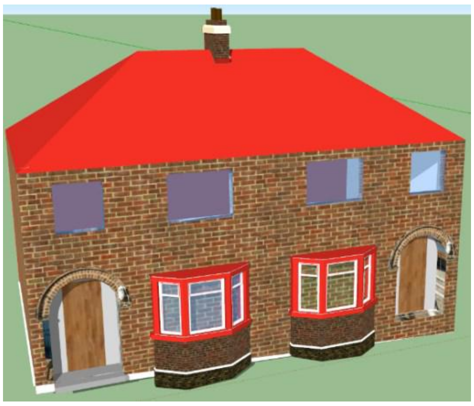

Figure 7-23. CityGML LOD3 modelling principles(SIG3D, 2015)



*InteriorWallSurface and FloorSurface* layers



*InteriorWallSurface, FloorSurface, Doors, ExternalWallSurface* layers



*InteriorWallSurface, FloorSurface, Doors, ExternalWallSurface, windows, CeilingSurface, RoofSurface, and BuildingInstallation* layers

All LOD3 CityGML layers

Figure 7-24. Snapshots illustrating the different CityGML LOD3 layers of a dwelling typology in the pilot area (Sneinton)

### 7.5.5 Exporting the 3D model to CityGML file format

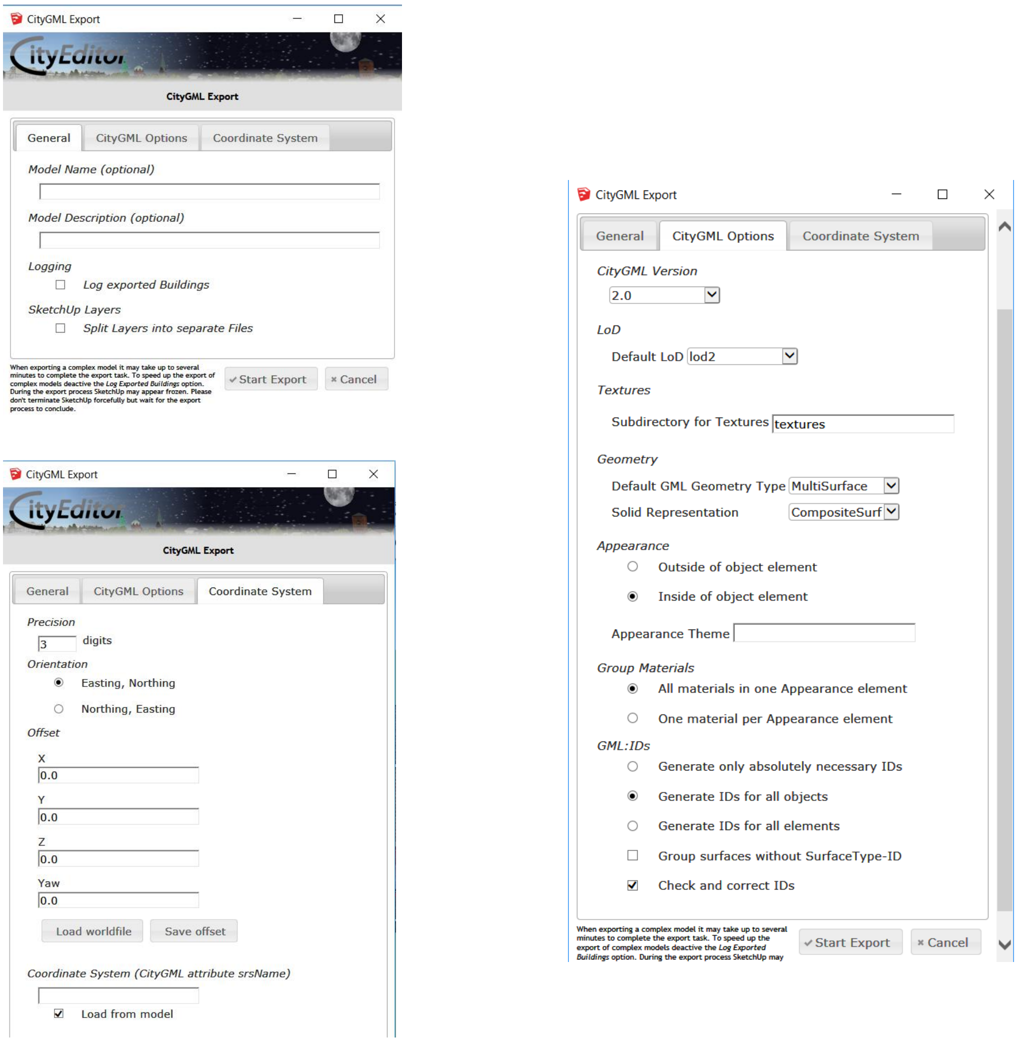



Figure 7-25. The main tabs of the CityEditor CityGML exporter user-interface

After accomplishing the CityGML LOD3 modelling of all typologies, the remaining building masses were replaced by the corresponding typology. Furthermore, some adjustments (e.g. Walls following slope) were applied to comply with the typology of the pilot areas. The final step involved exporting the whole 3D model to a CityGML file, which has an XML format, with the help of the CityEditor plugin developed by (3DIS, 2016). As shown in Figure 7-25, the user interface of CityEditor is composed of three main tabs. First, the general tab comprises fields dedicated for the model name and some description. The second tab; however, encompasses options describing the nature and characteristics of the envisaged CityGML file. A CityGML LOD3 version 2.0 was chosen because it is highly recommended by SIG3D (2015). As for the nature of exported geometry, the model was exported as solid geometry. As indicated in chapter two, solid geometry is better than the multi-surface one for energy simulation purposes because it enables an easier calculation of building volumes. Finally, the 3rd tab contains geo-referencing parameters, where the user may specify the orientation, offset the original model coordinates, and also determine the coordinate system used.

## 7.6 Conclusion

This chapter has reviewed the demographic and socio-economic characteristics of the residents of the Sneinton area in Nottingham. Moreover, addressed the initial development process of the 3D Semantic urban energy model.

First, the analysis of Sneinton residents' profile in comparison to the one of Nottingham city households has suggested a good agreement between both profiles despite minor discrepancies. This implies that the Sneinton area is representative of Nottingham city. Based on that, it is believed that the developed 3D urban energy prediction tool could forecast the residential energy consumption and demand of different neighbourhood areas within the city of Nottingham.

On the other hand, the second section of this chapter has concentrated on the CityGML modelling of the Sneinton area. The lack of software packages that support an automated conversion of 3D CAD models to CityGML LOD3, discussed in chapter two, led the researcher to take an alternative approach. This strategy involves a 4 stage hybrid process, and various digital tools and techniques. In stage1, ordnance survey data of the Sneinton area were imported from Edina Digimap in the form of vector data. After that, the downloaded three vector maps were cleaned and combined. The third phase; however, consisted of converting the mass models of the dwellings into models complying with CityGML LOD3. Finally, the last stage involved exporting the created LOD3 models into XML format using a Sketchup CityGML plugin. However, following this process was not only tedious but also time-consuming. This is because of the encountered interoperability challenges across different platforms and the technical problems resulting from the large size of 3D CAD file. More specifically, it was not possible to handle the 3D model of the whole pilot area in AutoCAD. Instead, it was subdivided based on the pilot area map tiles into sub files to save computational power. Based on these facts, there is a need to develop automated processes to allow an efficient conversion of existing CAD models to CityGML LOD3.

The next chapter discusses the integration of the research findings from chapters five and six into the initial CityGML LOD3 model.

# 8

# The development of EvoEnergy: 3D urban energy prediction tool

*The eighth chapter discusses in detail the process of implementing the findings from chapters five and six into the CityGML model developed in chapter seven. Moreover, analyses the findings from the participatory workshop which attempted to improve the developed urban energy prediction tool.*

**Contents**

## 8.1 Introduction

Chapter 7 has addressed the demographic and socio-economic characteristics of the pilot area (Sneinton). Moreover, it has shed light on the modelling process of its CityGML 3D model. Chapter 8; on the other hand, aims to discuss the implementation of the research findings. More precisely, the CityGML 3D model of the pilot area would be further developed to predict the domestic electricity consumption both at the building and urban scales. Following that, the 3D urban energy prediction tool was evaluated by a group of experts in the field, as indicated in Chapter 4. Finally, some improvements were made as a result of their invaluable feedback.

## 8.2 The Architecture of EvoEnergy

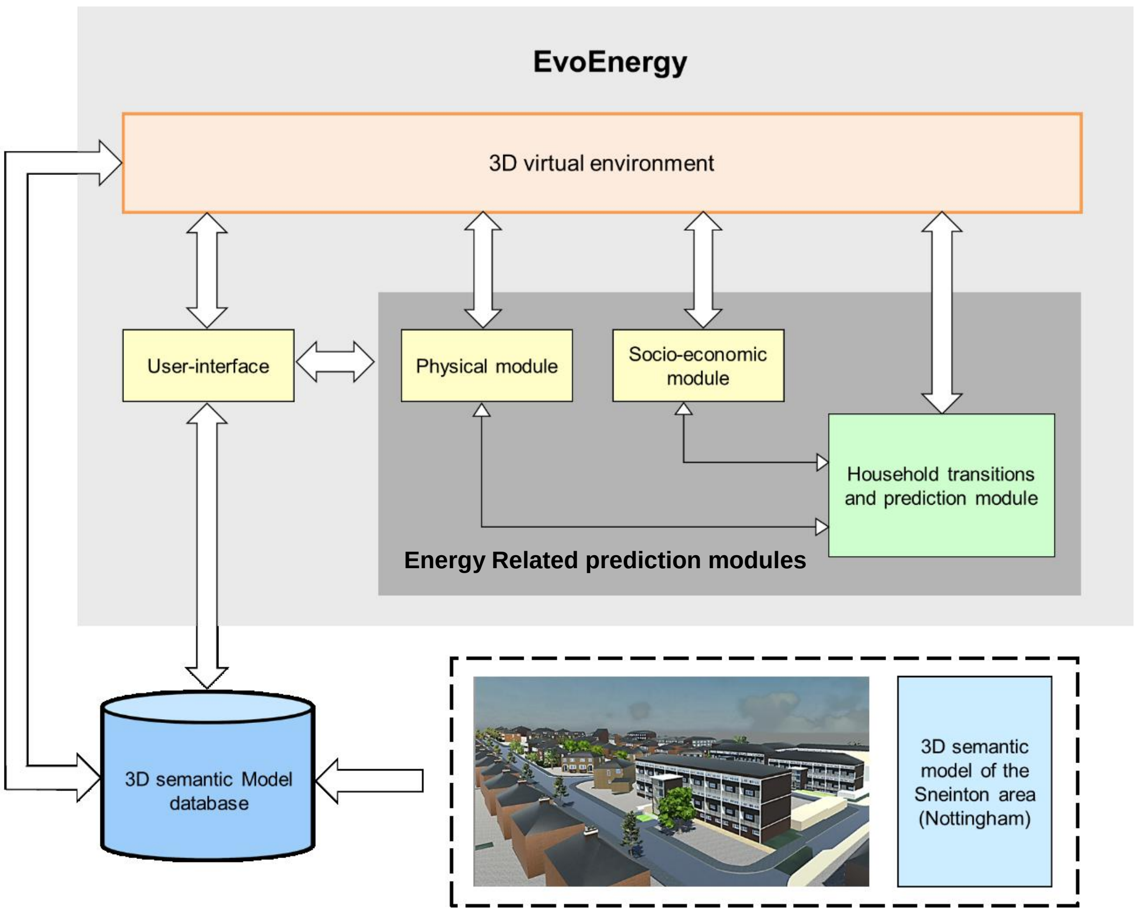



Figure 8-1.The Architecture of the developed 3D urban energy prediction tool (EvoEnergy)

A solid software architecture ensures the efficiency, stability, quality, and longevity of a system. Moreover, reduces its maintenance costs in the long run (IBM, 2007). Based on that, the development of EvoEnergy architecture has carefully taken into account prominent factors related to the requirements of urban planners who necessitate robust, stable, and reliable systems. Furthermore, it has considered the possibility of extending the capabilities of EvoEnergy to cover different energy consumers by contemplating factors such as constraints imposed by the hardware of mobile devices.

Figure 8-1 illustrates the architecture of the 3D urban energy prediction tool EvoEnergy, which is composed of 4 main components namely; **3D virtual environment**, **UI (user-interface)**, **energy related prediction modules**, and **3D semantic model database**.

First, the **3D virtual environment** represents the 3D platform where it is possible to import and interact with any 3D semantic model of a given urban area (e.g. Sneinton) via the user-interface module. Furthermore, it allows the mapping and visualisation of different types of information resulting from the energy related prediction modules such as household yearly energy usage figures and domestic transport emissions. The 3D virtual environment module is closely linked to **the 3D semantic model database** which, in turn, stores the different components of CityGML (3D-GIS) models in a hierarchically structured manner to ensure a stable and reliable data management. In addition to that, it permits to exchange data (e.g. Import, export, modify, and save) with 3D virtual environment module.

On the other hand, the **energy related prediction modules** are primarily responsible for predicting the demographic transition probabilities of different household types in line with their expected annual electricity usage. To facilitate the interaction as well as data exchange between such modules and the 3D virtual environment and the user-interface modules, they were organised into; **physical**,

**socio-economic**, and **household transition and prediction module. The physical module** contains information related to the physical and thermal characteristics of the household dwellings such as type HVAC system and type of walls. These parameters can be both stored in and retrieved from the 3D Semantic model database. One unique feature of the physical module is that it could retrieve geometric parameters (e.g. dwelling size) from the CityGML model (stored in the 3D semantic model database) via the user-interface. Similarly, **the socio-economic module** includes the socio-economic characteristics of the household heads such as income, level of education, socio-economic class, and monthly rent. Those parameters can be stored in and loaded from the 3D semantic model database module via the user-interface module. Finally, the household transition and prediction module encompasses the transitions and energy prediction statistical models developed in chapters 5 and 6. Independent factors of those models are directly retrieved from the physical and socio-economic modules via the user-interface module. As discussed at an earlier stage, the results of the statistical models can be visualised or mapped onto the 3D virtual environment module via the user-interface.

The subsequent section aims to cover the development process of EvoEnergy.

## 8.3 Development of the 3D urban prediction tool (EvoEnergy)

First of all, it should be noted that the name of the developed 3D urban energy prediction tool (EvoEnergy) is a shortened version of (Evolution Energy). It was chosen because it does consider the evolution of households over time in the prediction process. The development of EvoEnergy followed five major steps namely; cleaning of CityGML model, exporting CityGML to Unity3D, Development of a selection tool, development of a user-interface, and integrating household transition and electricity prediction models.

### 8.3.1 Cleaning the CityGML model

Since portability of EvoEnergy across different platforms, including mobile ones, is necessary, minimising the file size is crucial. The first step towards obtaining a light file size is to clean the CityGML model by erasing any hidden geometry, reducing the vertices of imported objects (e.g. chimneys, removing duplicated objects, and purging unused items. For those reasons,Cleanup3 (Figure 8-3), which is a powerful Sketchup plugin developed by (thomthom, 2015), was utilised. This has permitted to collapse the initial file size of the 3D CityGML model of approximately 80 MB to 46.7 MB.

### 8.3.2 Exporting the CityGML model to EvoEnergy

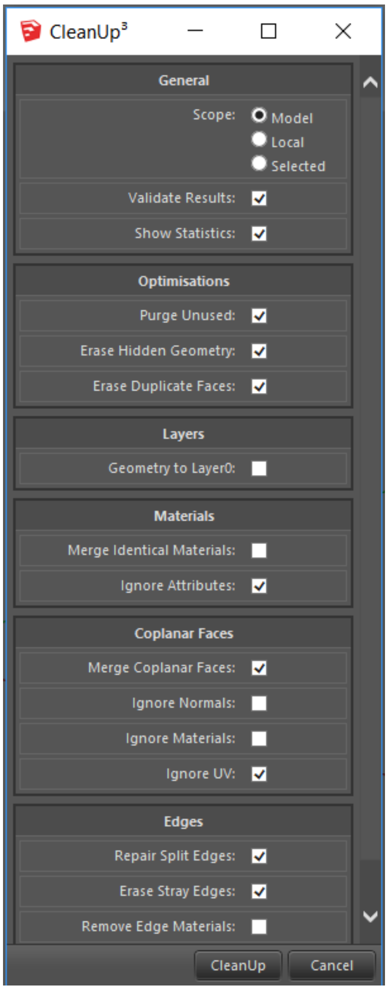



Figure 8-3. User-interface of Sketchup CleanUp plugin

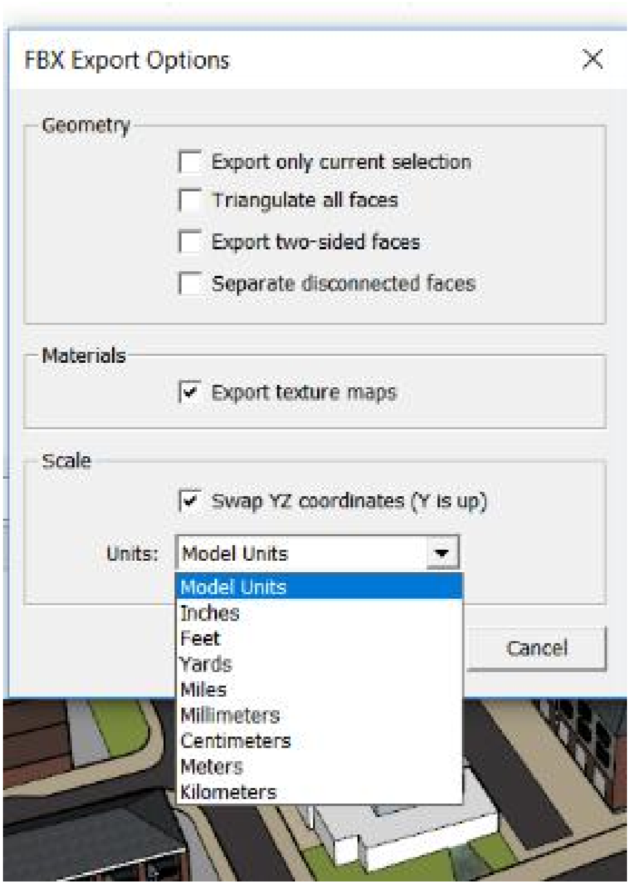



Figure 8-2.Sketchup FBX export dialogue box

After cleaning the CityGML model, the next step consisted of exporting it to unity 3D while maintaining all materials and layer structure. Two widely used formats support the export from CAD to game engines platforms namely; Collada and FBX. First, Collada (collaborative design activity) is an XML based schema which allows interoperability across different 3D based platforms such as CAD, game engines, and others (Trimble, 2017). It is well known for its ability to store complex scenes with multiple U/V sets. However, since it is an XML based, it is prone to precision issues and a slow parsing in comparison to other formats (Metashapes, 2016). On the other hand, FBX (FilmBox), which is a proprietary file exchange format belonging to Autodesk, was designed to support scenes comprising polygonal meshes with possible animations. The possibility to specify the export unit is an advantage of FBX over Collada (Figure 8-2). In this way, no necessary scaling is needed in Unity3D. However, one drawback is its large file size and the lack of specification which can create a lot of issues from a developer perspective (Metashapes, 2016). Nevertheless, FBX format overshadowed Collada to become the standard for importing 3D models in most games engines such as Unity3D and Unreal engine (Ardolino et al., 2014). For those reasons, FBX file format was adopted as an export format.

### 8.3.3 Development of a selection tool


Figure 8-4. Illustration of the developed selection tool projecting a red pattern on mouse hover

Undoubtedly, a vital step towards the development of a 3D urban energy prediction tool is to create a smart selection tool which can perform the following actions;

- Detects a particular dwelling on mouse hover
- Displays the corresponding address of the detected dwelling
- Highlights the selected dwelling through projecting a pattern on the ground

```
    void Update () {
        Ray ray = Camera.main.ScreenPointToRay (Input.mousePosition);
        RaycastHit hitinfo;
        if (Physics.Raycast (ray, out hitinfo)) {

            Debug.Log ("Mouse is over : " + hitinfo.collider.name);

            GameObject hitObject = hitinfo.transform.root.gameObject;

            SelectObject (hitObject);

        } else {

            ClearSelection ();

        }

    }

        void SelectObject(GameObject obj){
            if(selectedObject !=null){
                if(obj==selectedObject)
                    return;

                ClearSelection();
            }

            selectedObject=obj;
        }

        void ClearSelection(){
            selectedObject=null;
        }

}
```

Figure 8-5. The C# selection tool script developed by the researcher

In order to meet the above requirements, a centralised C# script was first created. It is based on the idea of casting a ray beam in every single frame. The origin of the beam is the camera position in the scene, whereas its direction, is based on the mouse position on the screen (Figure 8-5). Once the casted ray beam hits a particular dwelling, it would return its address. Secondly, another script was created to control the scene projector. More precisely, this permits activating/ deactivating the projector based on the centralised mouse manager script. Furthermore, it allows for an automatic positioning, rotation, and scaling of the projected pattern based on the dwelling dimensions as well as orientation, as shown in Figure 8-4.

### 8.3.4 Development of a user-interface

#### A-Tools and techniques

The development of the user-interface entailed different tools and techniques such as the new Unity 3D GUI (graphical interface) system and C# scripting. For more information on the tools and techniques involved in the development of EvoEnergy user-interface, please refer to appendix A.3.4.

#### B-Main components

After establishing a selection tool, the building of an intuitive and user-friendly interface occurred. This process was primarily based on EvoEnergy Architecture (discussed in 8.2) and the analysis of chapters two and three, and the contribution made by the researcher in this study. For those reasons, the user-interface was divided into four major components (Figure 8-6). The first module, which is known as “Dwelling physical and thermal characteristics”, encompasses physical information, such as dwelling age, type, and size. Furthermore, it includes information on the nature of a dwelling’s HVAC system (e.g., combi-boiler), HVAC controls, insulation, SAP rating, and bills inclusion.


Figure 8-6. The main menu of EvoEnergy

The second part, which is known as “Household socioeconomic characteristics”, contains demographic and socioeconomic information. These include the following factors: different income types (e.g., pension), household size, children dependency, householder age, marital status, household type, level of education, socio-economic class, tenure mode, and others. It should be noted that these two modules were designed to work in a complementary manner to preserve the integrity of the overall system. The third module enables consultation of the past energy usage trends of particular households at different granularities (e.g., monthly). Moreover, it allows comparison at different periods or with different users. Finally, the fourth module, which is the "Household transition" module, permits the prediction of future transition probabilities to various household types. Moreover, it enables the estimation of annual gas and electricity consumption figures based on the physical and socioeconomic modules as well as the predicted transitions.

### 8.3.5 Integrating household transition and electricity consumption models


```
public void Transition_Prob()
{

    if (Timeline.value == 1 & SelectedImage.name == "SingleNonElderly" & buttonpressed == true)
    {

        ProbSin1y = (float)(-10 * SingleElderly) - (float)(1.3 * CoupleNoChildren) - (float)(2.440

        float expprobsin1y = Mathf.Exp(ProbSin1y);
        ProbSin1yp = (expprobsin1y) / (1 + expprobsin1y) * 100;

        ShowPrediction.text = ProbSin1yp.ToString(); //+ "   %" +  "  Single non-elderly";

    }
```

Figure 8-7. EvoEnergy household transition module (above) and the script related to the transitions to single non-elderly households after 1 year (below)

After accomplishing the design of user-interface and its main functionalities, the process of incorporating the research findings (chapters five and six) took place. This was first attained by writing a script that defines the envisaged transition model. It consists of a publically referenced timeline slider and clipart images depicting the household type (Figure 8-7). Moreover, a public function which returns a void and takes no parameters. Once a clipart image is clicked, the script will return its name and a "true" statement. Similarly, the timeline slider will also return an integer. A combination of these values will help determine the desired transition model. For instance, to select the transition model of couples without children in the next two years, the user should first click on the couple without children clipart and then set the timeline to a value of 2 (Figure 8-7). This was followed by inputting the corresponding transition and electricity prediction models. In this way, when the process button, which is located under the timeline slider, is clicked, two text boxes will display the envisaged transition probability and average yearly electricity consumption in KWh.

## 8.4 Modus operandi of EvoEnergy

As soon as the application is launched, a 3D model of a particular area is shown (Figure 8-8). The user could walk through this area and select a particular dwelling(s) or perform a semantic query from the existing database by directly typing the address and postcode of a given dwelling. Once the user hovers over a particular house, it will be highlighted with a red selection box (Figure 8-9). To perform specific operations (e.g. Editing and prediction), the user needs to access the main menu, which contains the four modules described previously, with a right mouse click.

Upon launching the first two modules (physical and socioeconomic), a wide range of information on the envisaged household(s) is retrieved from the central database (Figure 8-10). However, the user cannot select or edit this unless they possess a permission password to do so. To consult the energy usage history of the selected occupant(s), the third module has to be employed. Once accessed, it is possible to choose between two different granularities, namely monthly and quarterly.

Furthermore, specify the period to be covered by this module (past year, past five years, etc.), and perform comparisons between different dwellings (Figure 8-11). This can be extremely useful in cases where the energy patterns of homes with similar physical properties and household characteristics are compared. In this way, it is possible to determine the effect of household behaviour on energy consumption.

Finally, the fourth module, which incorporates the powerful concept of household life cycle transitions, offers the potential to predict the demographic transition probabilities to different household types by simply clicking on one of the household clip arts and specifying the envisaged year of prediction on the timeline slider as discussed at an earlier stage (Figure 8-12). This, in turn, will map the energy usage based on the change in the socio-economic characteristics of the household following this prediction. Finally, the 3D urban energy prediction tool sub-module allows a meaningful comparison between the future transition patterns as well as estimated annual energy patterns of different households over a given period (e.g., 5 years), as is depicted in Figure 8-13.

Figure 8-8. A snapshot of the 3D model displayed upon launching EvoEnergy

Figure 8-9. Once the mouse is hovering over a particular dwelling, the projector will project a red selection box to the ground

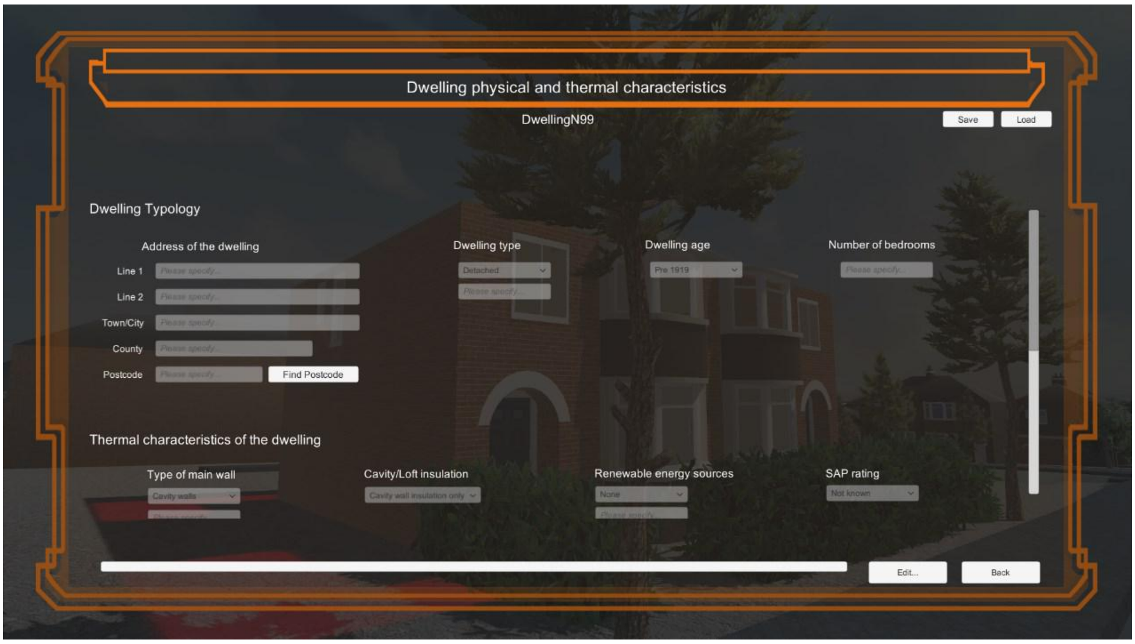



Figure 8-10. The physical module of EvoEnergy


Figure 8-11. The household energy history EvoEnergy module

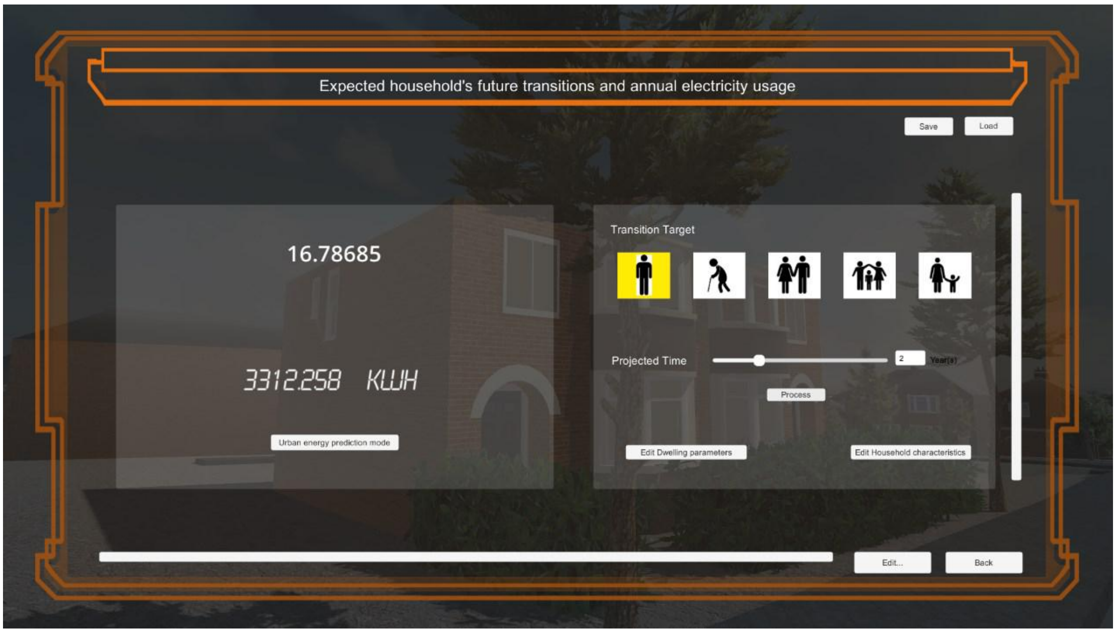



Figure 8-12. Household life cycle transition and energy prediction module (dwelling mode)

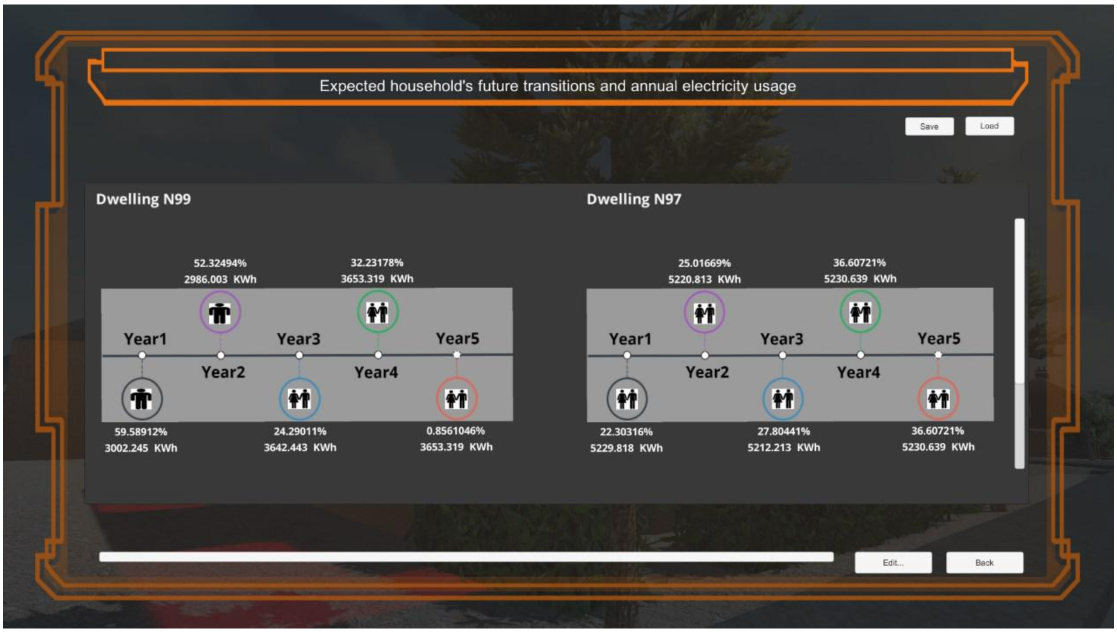



Figure 8-13. Household life cycle transition and energy prediction module (urban mode)

## 8.5 Benchmarking and improvement of EvoEnergy

As addressed in four, a participatory workshop comprising experts in urban energy planning or related fields, was conducted to give them the opportunity to experiment with the developed system. Moreover, share their invaluable opinions, and suggest improvements, or even propose new features. The below subsection will discuss the workshop circumstances.

### 8.5.1 Workshop circumstances

The participatory workshop was held on the 1st March 2017 at 10 am at the creative and virtual technologies research lab, Nottingham Trent University. A total of 6 experts participated in the workshop, in which four of them were urban energy planners, one specialised in residential energy storage, and one was specialised in energy economics and real estate. The researcher facilitated the workshop with the help of the main supervisor Prof. M. Benachir. Overall, both the duration of the whole workshop and the timing of its parts went as planned. For more information on the workshop, please refer to the methodology chapter.

### 8.5.2 Findings

---

**Key finding 1**

- Approaches supporting physical improvement in the residential sector are more dominant than the ones assisting techno-socio-economic measures. This reflects the nature of measures implemented by the UK government to reduce CO2 emissions.
- CityGML approaches in urban energy planning are still in their infancy.
- Survey data used in techno-socioeconomic approaches lacks variables on policy change and building regulations.

---

As discussed in the methodology chapter, a prominent part of the PowerPoint presentation delivered to the participant was to concisely summarise existing approaches supporting urban energy planning in the residential sector. These approaches, which were addressed in detail in chapters 2-3, are divided into physical and techno-socio-economic. Based on that, the participants were asked to work in pairs and answer the subsequent question;

- *Based on your experience, what are the limitations of existing energy planning tools?*

One group interestingly argued that the 2D GIS approach suffers from the lack of information granularity and precise physical models of buildings. Furthermore, claimed that 3D GIS overcome this limitation since buildings are represented in 3D which gives a much more accurate thermal model of the building. In addition to that, they believed that CityGML has no explicit limitations since it is an open standardised data model that allows the development of a wide range of applications. In this respect, one of the participants said *"CityGML is an open source system and you can develop what you want on it. For us there is no limitation as far as you can develop whatever you want on CityGML"*. However, another group shared a different opinion as they acknowledged that the prediction accuracy is also dependent on the level of detail of the CityGML model (3D GIS). Moreover, stressed the lack of urban 3D GIS models with a higher level of detail, which validates the statements made in chapter two with regards the limitations of CityGML approaches. Conversely, the remaining group were not familiar with CityGML. As a result, they could not identify and related limitations.

As for the inconvenience of techno-socio-economic approaches, the group comprising the expert in econometrics suggested that aggregate approaches are not able to capture income distribution, which is a significant disadvantage. Moreover, advised that variables on measures of energy efficiency in the building stock are not usually taken into account. This contributor also claimed that the interpretation of both aggregate and disaggregate data could be shaped by policy change especially when there is a lack of related variables. These include factors that capture when the policy was introduced and when the announcement of policy is going to come into effect. The impact of building regulations was also highlighted through the following statement; *"one of the things that would affect energy consumption would be building regulations and quite often datasets don't really capture a building regulation change"*. Another group believed that aggregate approaches often rely on survey data and their representativeness of survey data can be questionable. Furthermore, argued that disaggregated models (either regression or other econometric models) are exposed to the issue of unobserved heterogeneity. In other words, there should be unobserved factors in the dataset that could affect the actual relationship between the dependent and independent variables. Another group did not only agree with this limitation but also stated *"you might thing as income rises, people might spend more on energy. However, considering unobserved heterogeneity for other variables, energy consumption might rise because the household moved to a bigger house, for instance, and then that would bias the coefficient".* The remaining group did not share any views as they were not aware of techno-socio-econometric approaches.

---

**Key finding 2**

- Most participants expressed a need for an integrated system combining both physical and Techno-socio-economic approaches.
- The inclusion of smart simulation tools is considered essential to the success of an integrated urban energy planning strategies.

---

After the participants identified the limitations of the existing urban energy planning approaches in the residential sector, they were required to express their views by answering the following question;

- *"What is your preferred energy planning tool and why?"*

As expected given their experience in the industry and academia, the first group argued that an integrated system is the best approach. More specifically, they claimed that the perfect urban energy planning support system should consolidate the physical characteristics of the dwelling and the households' techno-socio-economic and behavioural characteristics. Furthermore, they stressed the importance of developing intelligent simulation tools within the physical module. In this respect, one of the group participants argued; *"In CityGML and 2D GIS we need to plug into* *them a bunch of simulation tools. If you don't have intelligent simulation tools, the data is not very useful."*

On the other hand, another group proposed to add energy audits to existing approaches, despite being prone to human error and their dependency on the quality of the field surveyor. To clarify for the audience, the group suggested that the implementation of this tool on a larger scale could be challenging but possible with the help of Archetyping. However, they expressed their concern over the quality of data in the subsequent statement *"The idea is because you usually have x number of archetypes. So, you got a lot terraced which can be mid-terraces and end-terraces and ideally you should be able to clone the data but the data has to be decent. However, this is not usually the case"*. In contrast to the first group, the participants of this one believed that the combination of CityGML and 2DGIS is the best approach. This is because of the nature of measures adopted by their organisations and which revolve mostly on dwelling retrofit. Indeed, a contributor claimed; *"CityGML is very useful for estimating the gas energy demand, whereas 2D GIS would assist in introducing more renewable energy sources"*. In this context, another participant argued that a CityGML approach has a better level of detail and visualisation than the 2D GIS ones as it is possible to break the energy consumption down by floors. The opinions of the remaining group; however, were in line with the first one with regards the integration of 3D CityGML with Techno-socio-economic approaches.

In the light of the above facts, it is evident that the aim of the undertaken research and its developed framework in chapter 2 align with most of the participant's views on the best urban energy planning approach.

**Key finding 3**

- The vast majority of parameters mentioned by the experts were addressed by the researcher in this thesis in chapters 2-3.
- Participants were completely unaware of household transitions.
- Organisations conducting national household surveys should include parameters on cultural context, policy change, and building regulations.

Following the above discussion on the most preferred urban energy planning approach, the participants were asked to work in pairs and list parameters which are thought to be essential for effective residential energy planning decision-making. The main intention behind this question is to investigate to what extent participants are considering household life-cycle transitions as part of urban energy planning. The first group, who previously suggested an integrated approach, believed that a combination of physical, socio-economic, demographic, and technological parameters is an indispensable step towards establishing such as system. However, they felt that criteria related to cultural context such as ethnicity, are often neglected. This was highlighted in the following statement: *"There is one criterion which we haven't seen a lot which is cultural context. If you have a dwelling and you put a Saudian householder into it, at 18 C, he/she will put the heating up. In the same house, you put a Swedish in it, at 18C he/she will go outside to enjoy some sun"*. All the remaining groups found this point very interesting, although the researcher did not encounter any significant relationship between energy consumption and ethnicity. This raises many questions on the sampling procedures and representativeness of ethnic minorities in the BHPS dataset. Indeed, understanding society, which is the organisation responsible for conducting UKHLS (The UK Household Longitudinal Study), acknowledged this limitation and developed sampling mechanisms to enable sub-population groups to be investigated. In addition to this, another group's contributor suggested that energy forecasting models should include a ponderation number where its value is dependent on the cultural context of the households. This is mainly because factors related to cultural context should have an impact on variation in residential energy patterns, heating particularly. On the other hand, the rest of the other groups believed that archetype, the age of the dwelling, energy demand and consumption, thermal efficiency, policy (current policy and previous policy), ethnicity, culture, occupant age, and health condition, were necessary parameters. Furthermore, argued that income and educational attainment are important determinants of energy baselines. In this regards, one of the participants claimed that low-income households living in social housing, who generally have a lower educational attainment, are much quicker in responding to offers of help to reduce their energy bills.

In the light of the above facts, it is evident that none of the experts were aware of household transitions and its potential in urban energy planning.

---

**Key finding 4**

- The concept of household life-cycle transitions and its applicability in urban energy planning was seen prominent by most of the participants.
- The private rented sector should benefit the most from the integration of this concept due to its changing occupancy.
- The experts' opinions on heterogeneity reinforce the robustness of the adopted research methodology and methods in this doctoral study.

---

After discussing the strengths, weaknesses, and input parameters of existing urban energy planning approaches, experts were required to answer the subsequent question in pairs;

- *"Do you think that household transition is an important factor you may consider in residential energy planning?"*

First, group 1 found the idea of incorporating household transitions in relation to their life-cycle very interesting and could potentially lead to the development of effective urban energy planning approaches. Moreover, they argued that this approach could be very useful especially for privately rented properties where the occupancy changes quite often. More precisely, by knowing the energy consumption changes associated with different transitions, energy usage of privately rented dwellings can be reduced by aiming some of them at families instead of students. This is because a lot of students do not have to worry about their energy bills since included in the rent.

On the other hand, the group comprising the expert in energy economics and real estate were unable to answer the question in a definite way due to the issue of heterogeneity. This is because there are different household and property types, which may necessitate focusing on certain household types (e.g. single non-elderly) to capture all of them. In addition to that, one of the participants suggested that the advantage of using datasets covering physical and techno-socio-economic factors is the ability to control for heterogeneity. In this way, it is possible to identify factors with the most effect on energy usage and adapt accordingly. Premised on these arguments, it is evident that utilising a data-driven approach as well as focusing solely on the transitions of single non-elderly in the undertaken research is a solid and robust strategy to control for heterogeneity.

Finally, the last group believed that the concept of household life-cycle transitions is very intriguing and extremely important in urban energy planning. However, variables related to the length of occupancy are needed in the transition models to predict the energy usage accurately.

### Graphical and functional improvements suggested by the experts

After the experts discussed their views on the above four matters, they were offered a short presentation on the impact of household transitions on residential energy consumption patterns. This was followed by a short demo on "EvoEnergy" where each participant had the chance to experiment with the system, its user-interface, and functionalities. Once familiar with the system, the users were required to work in pairs on a PowerPoint file, which contains all the elements of the user-interface, to improve "EvoEnergy". These changes covered both the graphical and functional aspects and were directly related to urban energy planning decision-making.

Overall, most of the experts' comments and suggestions revolved around the operational aspects of the system. More specifically, the first group believed that the Non-diegetic nature of user-interface, which is based on screen overlays, isolates the CityGML(3D-GIS) model from urban energy predictions. As a result, they argued that the 3D-GIS model became redundant when inputting and editing data and also while performing predictions. In this regards, one of the participants, who has a long experience in developing digital applications for architecture and urban planning, suggested that the optimal solution would be to integrate the user-interface with CityGML model itself. More precisely, the main menu should be changed from being screen overlay to a camera perspective, as shown in Figure 8-14 and Figure 8-15. In addition to that, the expert stressed the importance of providing a hover menu containing a summary of key information (e.g. demographic, socio-economic, physical, etc...) on each household.

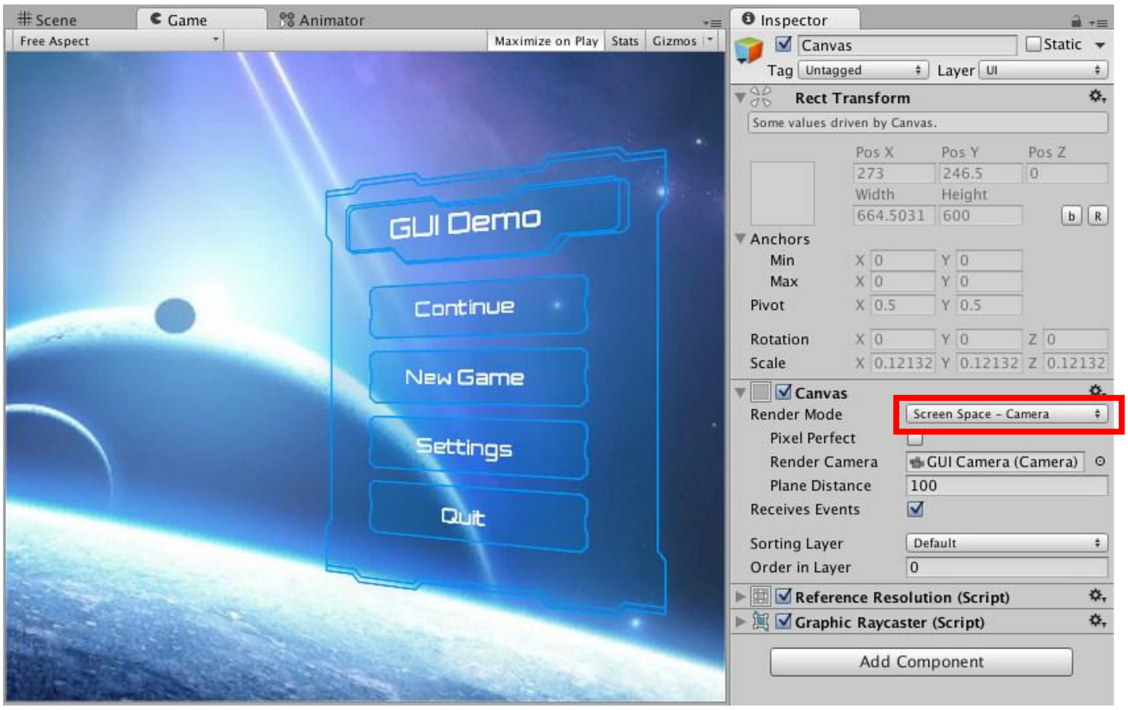



Figure 8-14. An example of a Unity menu with a screen space (camera perspective) Canvas

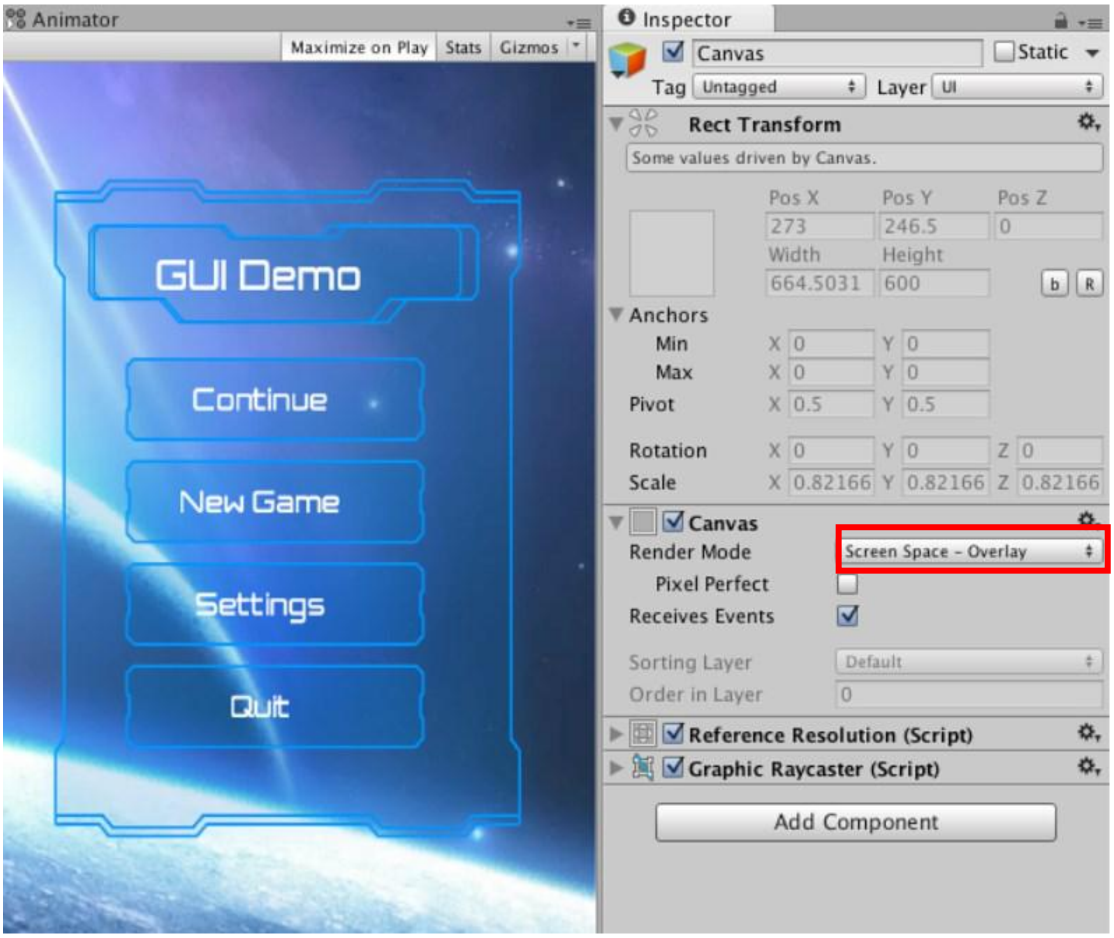



Figure 8-15.An example of a Unity menu with a screen space (Overlay) Canvas

In contrast to group 1, despite being considered by the researcher for future development, the second group recommended a new feature. The main purpose of these functionalities is to facilitate the navigation through different parts of the Sneinton area. This comprises a smart querying system which enables finding a particular dwelling based on its coordinates or the energy consumption figures of its household (e.g. less than 1500 KWh per annum).

Finally, the recommendation of the remaining group entailed incorporating a 2D GIS mode in the form of maps visualising the energy consumption of neighbourhoods thematically. Their argument was based on the fact that the urban prediction mode of EvoEnergy, in its current form, can only be useful at the parcel level.

## 8.6 Conclusion

Chapter eight has discussed in detail the process of integrating the research findings from chapters five and six into the CityGML LOD3 model developed in chapter seven. To attain this objective, a five steps pipeline, which comprises innovative digital technologies such as Unity3D game engine, has been adopted. First, the CityGML LOD3 model of the Sneinton area has been prone to a cleaning process to ensure interoperability and portability across different platforms, including tablets. After that, the model has been exported as an FBX (FilmBox) format to Unity3D game engine. Once achieved, a smart selection tool has been established with the help of C# scripting and Unity3D projector. Next, an intuitive user interface composed of four distinct modules has been generated with the help of Unity3D UI module and C# scripting. Finally, the transition and electricity consumption models have been integrated into the energy prediction module using C# scripting.

In order to evaluate and improve the developed 3D urban energy prediction tool (EvoEnergy), a group of experts in urban planning and related fields has been invited to attend a participatory workshop. The majority of them shared similar views with researcher and the supervisory team with regards the dominance of physical approaches, the infancy of CityGML, the inclusion of smart simulation tools in energy planning, the need for integrated approaches, and the nature of inputs in energy planning tools. Surprisingly, however, some highlighted the lack of variables on the cultural context, policy change, and building regulation in the majority of national household survey data. Furthermore, stressed their influence on the variation in residential energy consumption and demand. As expected, none of the experts was aware of the concept of household life-cycle transition. However, they acknowledged its prominence in urban energy planning, especially in managing the energy demand in the private sector whose occupancy changes frequently.

Finally, after evaluating EvoEnergy, the experts suggested improving the subsequent operational aspects of the 3D urban energy prediction tool.

- Maximise the integration of user-interface in the CityGML model by changing its type from screen overlay to camera perspective overlay.
- Incorporate a hover menu which provides a summary of the household socio-economic, demographic, and actual/expected energy usage.
- Develop a 2D GIS module that enables the mapping of residential energy usage/demand at scales other than the parcel level.

The next chapter intends to discuss the research findings in relation to key studies in the literature. Moreover, evaluates EvoEnergy strengths and weaknesses, and recommends certain areas for future research.

# 9

# Discussion and conclusion

*The ninth chapter discusses the research findings in relation to the relevant literature. Furthermore, concludes the research by summarising key findings, and highlighting its strengths and weaknesses. Furthermore, outlining key areas for future investigation*

**Contents**

## 9.1 SUMMARY OF THE MAIN RESEARCH FINDINGS

The undertaken research aimed to utilise the potentials of innovative digital technologies to develop a 3D urban energy prediction tool that supports sustainable energy planning decision-making in the residential sector. Although the lack of integrated urban energy planning system was identified as a significant gap, the **novelty** of this study consists of predicting future energy usage of households based on their life-cycle transition patterns. This was achieved by answering the below three research questions;

1. How will the different British household structures evolve in the next 5-10 years?
2. Is there a significant relationship between household transitions and energy usage patterns?
3. If so, to what extent it does affect the households' energy consumption figures?

The findings related to each research question will be summarised separately.

- *How will the different British household structures evolve in the next 5-10 years?*

To answer the above research question, the demographic and socio-economic factors affecting different family transitions had to be investigated. This, in turn, helped predict the likelihood of moving to different household types in the future. For reasons related to the feasibility of research and the high proportion of single non-elderly households in the city of Nottingham in general and Sneinton in particular, a focus was placed on the household transition patterns of single non-elderly. Overall, the number, magnitude, and factors affecting household transitions were not consistent across all years. Therefore, only the common ones across all/most years are reported in this section.

### Households becoming/remaining single non-elderly

As depicted in Table 9-1, factors which could affect transitions to single non-elderly in the first five years are; *Living as a cohabiting couple, householder age, single non-elderly, household size, and living in terraced houses.* In particular, *cohabiting couples* and *householders living in terraced houses* are more likely to be single non-elderly over the next five years. On the other hand, families with larger household size and older householders are less likely to be single non-elderly over this period. Householders with a single non-elderly household type; however, are less likely to remain so after 3 years.

In contrast to the above, common variables affecting the transitions to single non-elderly households in the next 5-10 years are as follows; *householder age, single non-elderly, annual benefit income, and gross annual income.* More precisely, older householders, single non-elderly, households with higher gross annual income and benefit income are less likely to be single non-elderly in next 5-10 years.

Table 9-1. Summary of common factors affecting transitions to single non-elderly household

| | Single non-elderly 0-5 years | Impact | Single non-elderly 5-10years | Impact |
|---|---|---|---|---|
| **Key influencing factors** | Living as cohabitating couple | + | Householder age | - |
| | Householder age | - | Single non-elderly | - |
| | Single non-elderly | +/- | Annual benefit income | - |
| | Household size | - | Gross annual income | - |
| | Living in terraced houses | + | | |

### Households becoming couple without children

As shown in the below Table 9-2, transitions to a couple without children household in the first five years were impacted by the following common factors namely; *couples* *with children, other households, divorced, widowed, never-married, age of householder, on a pension, gross annual income, and annual benefit income.* Except for the 5th year, families with children were unlikely to become couples without children over the first five years. Similarly, divorced, widowed, and never-married households in addition to the ones with high annual benefit income were less likely to move to couple without children households. Conversely, older householders with higher gross annual income, in addition to the ones receiving a pension income, are more likely to be couples without children during this period.

Table 9-2.Summary of common key factors affecting transitions to couple without children households

| | **Couple without children 0-5 years** | **Impact** | **Couple without children 5-10years** | **Impact** |
|---|---|---|---|---|
| **Key influencing factors** | Couple with children | -/+ | Couple with children | + |
| | Other households | + | Separated | + |
| | Divorced | - | Never-married | + |
| | Widowed | - | Age of householder | + |
| | Never-Married | - | Higher-grade professionals | _ |
| | Age of householder | + | Semi-skilled manual workers | _ |
| | On pension | + | Annual benefit income | _ |
| | Gross annual income | + | | |
| | Annual benefit income | - | | |

In contrast to the above, *couples with children, separated, never-married, the age of householder, higher-grade professionals, semi-skilled manual workers, and annual benefit income* were the factors affecting transitions to couple without children in the next 5 to 10 years. In more details, couples with children, separated, and never-married, and older householders, are more likely to move to a couple without children family. Conversely, higher-grade professionals and semi-skilled manual workers in addition to the ones with higher annual benefit income are unlikely to be couples without children over this period.

### Households becoming couple with children

Table 9-3.Summary of common key factors impacting the transitions to couple with children households

| | Couple with children 0-5 years | Impact | Couple with children 5-10years | Impact |
|---|---|---|---|---|
| **Key influencing factors** | Aged between 26 and 35 | _ | Aged between 36 and 45 | + |
| | Aged between 46 and 55 | + | Aged between 46 and 55 | + |
| | Living as cohabitating couple | _ | Living as cohabitating couple | _ |
| | Divorced | _ | Higher-grade professionals | + |
| | Never-married | _ | On benefit | + |
| | Living in a detached house | + | Annual benefit income | + |
| | Living in a semi-detached house | + | | |
| | Living in a terraced house | _ | | |
| | Living in a purpose built flat | _ | | |
| | Dwelling owned outright | + | | |
| | Higher-grade professionals | + | | |
| | Lower-grade professionals | -/+ | | |
| | Annual benefit income | + | | |
| | Annual investment income | _ | | |

As illustrated in Table 9-3, transitions to couple with children households in the first five years were affected by the subsequent factors; aged 26-35, aged 46-55, living as a cohabiting couple, divorced, never-married, living in a detached house, living in a semi-detached dwelling, living in a purpose built flat, dwelling owned outright, higher and lower grade professionals, annual benefit income and annual investment income.

Apart from the factors namely; aged 46-55, living in a detached house, dwelling owned outright, higher-grade professionals, and annual benefit income, the remaining variables had a negative impact on this type of transition.

On the other hand, the following factors namely; *aged 36-45*, *aged 46-55*, *higher-grade professionals*, *on benefit*, and *annual benefit income,* positively influenced transitions to couple with children households over the next 5 to 10 years. Finally, cohabiting couples are less likely to be couples with children over this period.

### Households becoming lone parents

Table 9-4.Summary of common key factors impacting the transitions to lone parent households

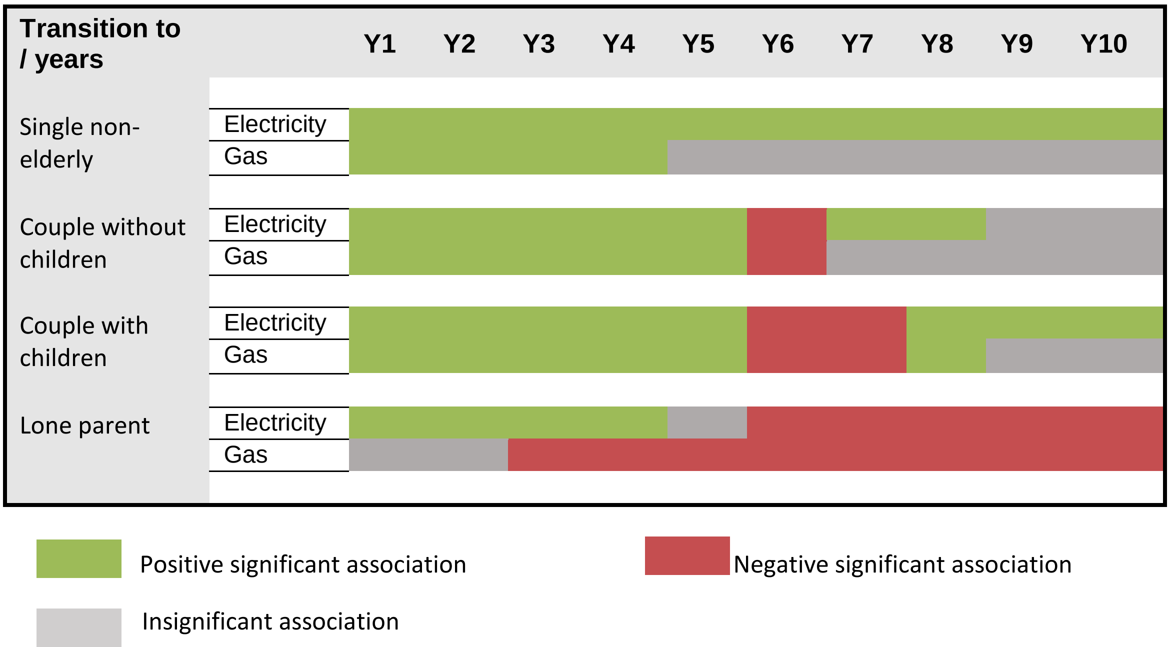
| | Lone parents 0-5 years | Impact | Lone parents 5-10years | Impact |
|---|---|---|---|---|
| **Key influencing factors** | Single non-elderly | -/+ | Single non-elderly | + |
| | Couple with children | _ | Female | + |
| | Female | + | Aged between 26 and 35 | + |
| | Divorced | + | Dwelling rented from local authorities | _ |
| | Aged between 46 and 55 | + | Higher-grade professionals | _ |
| | Household size | + | Skilled manual workers | + |
| | Living in end-terraced houses | + | Working full-time | _ |
| | Living in terraced houses | + | | |
| | On pension | _ | | |
| | Annual benefit income | + | | |

As depicted in Table 9-4, transitions to a lone parent household within the next five years are commonly affected by the following factors namely; single non-elderly, couples with children, female, divorced, aged 46-55, household size, living in an end-terraced and terraced house, on a pension, and annual benefit income. All these factors, except couples with children and on a pension, had a positive effect on this transition type. However, the variable single non-elderly is expected to have a negative effect in the first three years and a positive one afterwards.

In contrast to the above, single non-elderly, female, aged 26-35, dwelling rented from local authorities, higher-grade professionals, skilled manual workers, and working full-time, influenced the transitions to this household type in the next 5 to 10 years. The majority of these variables had a positive effect, except for dwelling rented from local authorities, higher-grade professionals, and working full-time.

- *Is there a significant relationship between household transitions and energy usage patterns?*

Table 9-5. Summary of the impact of household transitions on electricity and gas consumption




In the light of the above Table 9-5, it is so clear that household transitions **do have mostly** a weak significant impact on their electricity and gas consumption patterns. In general, this effect is mostly positive, except for transitions to lone parent households. Moreover, the 6th and/or 7th years of transitions to couples without children and with children. However, the association between family transition variables and the energy consumption ones was not significant in certain years. For example, the effect of moving/remaining a single non-elderly household on gas usage patterns was insignificant in the period between year 5 and year 10.

- *If so, to what extent it does affect the households' energy consumption figures?*

Table 9-6. Summary of the effect of household transitions on their yearly electricity consumption

| Transition target | Average annual change in the next 1-5 years | Average annual change in the next 6-10 years |
|---|---|---|
| **Single non-elderly** | 10.7% | 3.64% |
| **Couple without children** | 16.28% | 3.86% |
| **Couple with children** | 26% | 6.20% |
| **Lone parents** | 13.87% | -5.8% |

In the light of the above Table 9-6, it is clear that household transitions had different effects on electricity consumption. This difference was defined by the time and nature of the occurring transition (e.g. to couple with children). For example, over the next 1 to 5 years, the transition to single non-elderly households was responsible for 10.7% of the variation in the yearly electricity consumption. Conversely, the transition to a couple with children had the highest effect of 26% on the electricity usage variable. Similarly, the influence of transitions in the next 6-10 years was negative on the electricity consumption of lone parents which should decrease by 5.8%. Moreover, transitions to couple with children households remained the most influential on the yearly electricity usage. Based on all the above invaluable findings, a 3D urban energy preidtcion tool (EvoEnergy) was developed with the help of specific digital technologies and techniques such as game engines. Figure 9-1, Figure 9-2, Figure 9-3 (below) depict the user-interface of this system.

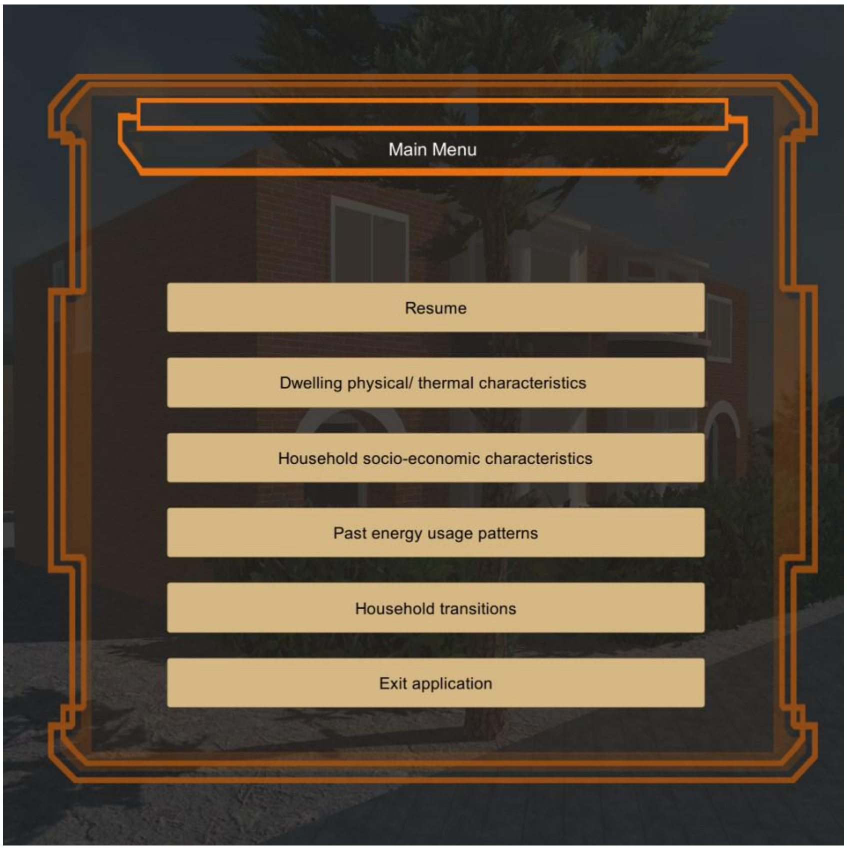



Figure 9-1. Main components of EvoEnergy

Figure 9-2. Summary of a particular household energy usage and socioeconomic profile on mouse hover

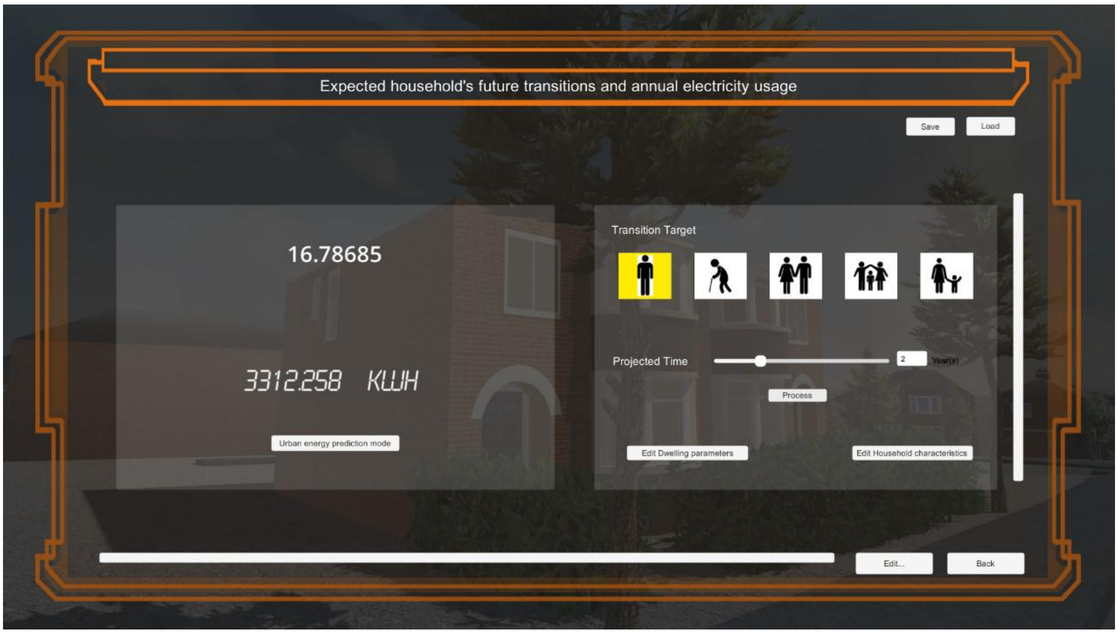



Figure 9-3. Household life cycle transition and energy prediction module (dwelling mode)

## 9.2 Discussion in relation to existing literature

The undertaken research has been built on the invaluable work that has been conducted on energy forecasting tools supporting residential urban energy planning decision-making. This sub-section aims to review how the developed 3D urban energy prediction tool "EvoEnergy" relates to the discussed approaches in chapters two and three as well as the wider context. Moreover, it discusses the values added by EvoEnergy and how it could improve the limitations of these approaches.

### 9.2.1 Discussion in relation to 2D-GIS approaches

As proposed in chapter two, there are three types of approaches supporting physical improvements in the residential sector at the urban scale namely; 2D GIS, 3D-GIS, and statistical. First, in comparison to 2D GIS approaches, EvoEnergy is different in many aspects.

The first difference covers the applicability in the building life-cycle. Unlike 2D-GIS approaches, which mainly support the operational stages of building life-cycle, the developed 3D urban energy prediction tool (EvoEnergy) can be applied to any stage and any residential area. This is mainly because the employed disaggregated data input in this model was based on a representative sample covering the whole UK. This signifies that the predictions of energy consumption of households with particular demographic, socio-economic, and physical characteristics can be generalised to similar households anywhere in the UK. Unfortunately, this is not the case for existing 2D-GIS approaches because they are mainly based on locally collected data. Furthermore, they rely on aggregate data which means that heterogeneity across various households is not captured (Bunn and Rostom, 2015). This is evident in the work of Jones et al. (2001), in which they estimated the change in the energy consumption of Neath Talbot residential area in the UK before and after applying some dwelling retrofit measures. However, the authors concluded that their approach could not be generalised to the whole UK residential sector. More recently, Nielsen and Möller (2013) developed a GIS-based approach to locate residential areas with district heating potentials based on the associated cost. The author acknowledged that their approach is not adequate for decision-making. Moreover, it cannot be applied at a more disaggregated level (e.g. municipality level), since it is based on nationally aggregated data.

Considering the above limitations of 2D-GIS approaches, data disaggregation in "EvoEnergy" does help predict the energy consumption of various users at the early design stages while considering the effect of physical, demographic, and socio-economic factors. This will, in turn, permit the development of proactive decisions to reduce CO2 emissions in the residential sector. A possible scenario includes the estimation of energy consumption of new developments to generate policies/strategies that contribute to reducing the carbon dioxide emissions. These measures will, in turn, take account of the dynamic aspect of household structures based on household transitions in the next 10-15 years.

The second major difference lies in the *frequent purpose of use* of 2D-GIS approaches. More precisely, as discussed in chapter two, they are mainly utilised to evaluate the effectivity of certain physical measures before and after being implemented. Moreover, they are employed to determine potential areas for implementing renewable energy sources, including hydropower, district heating, solar panels, and wind turbines (Aydin et al., 2013; Wróżyński et al., 2016). Although not available in its current release, EvoEnergy was designed to account for most of the 2D-GIS capabilities in addition to the change in energy consumption in relation to household transitions as well as demographic, socio-economic, and physical factors. Of course, there are some practical limitations that are currently or will be addressed in the future, which are mainly related to the integration of external energy plugins.

The third difference is that 2D-GIS approaches are costly and time-consuming to develop because their accuracy heavily relies on building energy simulation tools and the collection of local data. This is not the case for EvoEnergy since it relies primarily on disaggregated household panel data which are regularly collected and maintained by national governmental agencies, including UnderstandingSocitey.

The finale difference lies in the ability of EvoEnergy to retrieve more physical indices from the 3D model such as floor area, volume, glazing ratio, solar exposure, heat gain, orientation, aspect ratio, and azimuth. The majority of these parameters cannot be obtained from 2D-GIS maps. Hence, a lot of assumptions are made which affects the prediction accuracy.

### 9.2.2 Discussion in relation to 3D-GIS approaches

The developed 3D urban energy prediction tool (EvoEnergy) shares many common aspects with 3D-GIS approaches since both of them utilise CityGML principles.

One of the similarities is the 3D *modelling* pipeline. This is evident in the heavy reliance of both of them on ordnance survey data to extract physical parameters such as building heights, footprints, buildings orientation, and others. Furthermore, it is also clear in the adoption of historical energy consumption data from energy companies for monitoring and validation purposes. Like some of the recent 3D-GIS approaches (e.g. (Jokela, 2016)), EvoEnergy also considered the integration of advanced acquisition and surveying technologies, including 3D laser scanning, and Lidar, as previously discussed in chapter two. However, these were not implemented in the current research due to the lack of time and resources.

In contrast to the above similarities, there are also some differences between both approaches. First, the level of details (LOD) employed for predicting heat energy demand in residential buildings represents a significant difference. As addressed in chapter two, the majority of existing 3D-GIS approaches use CityGML models with a level of detail (LOD1 or LOD2). This means that there is a non-consideration of openings and protruding building elements (e.g. balconies). EvoEnergy overcomes this limitation by utilising a superior level of detail, CityGML LOD3, more precisely. As a result, EvoEnergy is expected to provide a higher prediction accuracy of heating demand consumption than the existing 3D-GIS models. Indeed, a recent study conducted by Monien et al. (2017) suggested that the heating demand figures predicted using LOD3 models were approximately 6% more accurate than those employing LOD2 models.

In addition to the above, another major difference is the degree of *independence of calculation engines* from the 3D model itself. As discussed in chapter two, the majority of existing CityGML urban energy models such as the energy Atlas of Berlin (TUBerlin, 2017), are entirely independent of energy calculation engines. This means that most of the energy predictions occur in Excel-based simulation tools incorporating steady-state algorithms (e.g. Fraunhofer-IBP). In this respect, it was argued that the potentials of CityGML were not fully explored and utilised. For those reasons, EvoEnergy was carefully designed to benefit the most from the CityGML model by integrating the prediction algorithm within the main user-interface. In this way, relevant geometric information to every single residential building (e.g. volume) is extracted from its relative 3D model. Moreover, saved to its physical module. These parameters will, in turn, be used as inputs for predicting the energy use of the envisaged dwelling(s). This is believed to speed up the prediction process, reduce costs, prevent technical issues, and most importantly leads to a smarter exploitation of CityGML models.

The *operating environment* also represents a major difference and a core contribution of this research. Most of the 3D-GIS energy forecasting models operate on well-known GIS software such as CityEngine and ArcGIS. The work conducted by Nouvel et al. (2015), Chen et al. (2017), and Kaden and Kolbe (2013) reinforces this claim. This is considered absolutely fine for many urban energy planners who should be familiar with these systems. However, considering that public engagement plays a prominent role in meeting the CO2 emission targets, extending these tools to energy consumers necessitates the adoption of other non-technical, flexible, interactive, and user-friendly platforms. One of these platforms is game engines, which in turn have received growing attention in urban planning in recent years. In this context, it is worth mentioning the notable urban planning gaming applications of "***Betaville***" by Skelton (2013) and "ASPECT Studios 3D Viewer" by Greenwood et al. (2009). However, despite the invaluable contribution of many scholars, there **is No** gaming application tailored to suit the needs of urban energy planning. For those reasons and for others related to the intention to extend the capability of EvoEnergy to the general public, Unity3D, an open source game engine, was utilised as an operating platform.

In addition to the above, *the application in the building life-cycle* depicts a prominent difference between EvoEnergy and 3D-GIS approaches. The latter has been mainly dedicated to the *operational stages* of building life-cycle. More specifically, they have been heavily linked with targeting buildings with retrofit potentials through assessing their heat demand (Nouvel et al., 2015). Furthermore, their capability was extended in the last 2 years to cover the evaluation of renewable energy potentials of different building areas (e.g. walls, roof, etc...) (Saran et al., 2015). However, the problem with 3D-GIS approaches is that the effect of socio-economic and demographic factors is ignored. In other words, predictions are based on standard household profiles and occupancy schedules. EvoEnergy acknowledged this limitation and tried to overcome it by incorporating a socio-economic module in its prediction algorithm. For those reasons and other related to the higher level of detail of EvoEnergy, it is argued that predictions are more accurate and reliable. It should be noted that the accuracy and reliability of EvoEnergy will be assessed in the following sections. As indicated at an earlier stage, prediction generated by EvoEnergy should be employed at any stage of the building life-cycle. This is because they are based on statistical models developed based on a UK government panel data surveys that are representative of the whole population and dwelling stock.

### 9.2.3 Discussion in relation to disaggregated techno-socio-economic approaches

Undoubtedly, the developed 3D urban energy prediction tool in this study (EvoEnergy) derives many of its principles from disaggregated techno-socio-economic approaches, discussed in chapter three. This is mainly because these approaches have a great ability to consider heterogeneity across individuals, which permits the identification of accompanying changes in their energy usage patterns. Thus, they are considered adequate tools for developing policies that target specific household types or segments (e.g. low-income high energy consumers).

In the light of the above, one of the similarities lies in the nature of *data source* and *data inputs* that constitute the forecasting model. Indeed, both EvoEnergy and disaggregated techno-socio-economic rely heavily on well-known panel data surveys such as the British household panel data survey. This is because this type of data has a powerful ability to capture complex dynamic relationships (e.g. Household mobility and change in income, education, etc...). The work conducted by Meier and Rehdanz (2010), Longhi (2014), Longhi (2015), and Jones et al. (2015) supports this claim. This type of data has, in turn, an implication on the nature of parameters involved in the prediction model. In other words, EvoEnergy shares a lot of socio-economic, demographic, and technological parameters with these approaches since they are widely available on panel data surveys such as BHPS and UKHLS. Moreover, because the literature has shown their prominent impact on the residential energy consumption patterns. Finally, another resemblance between EvoEnergy and these approaches is the *nature of prediction model*. Indeed, the majority of these approaches utilise well-known statistical methods (e.g. fixed-effects/ random effects models), because they capture heterogeneity within/ across groups and overtime.

As much as there are many similarities between EvoEnergy and disaggregated techno-socio-economic approaches, there are also some differences. First, despite they can account for heterogeneity across different or similar households and predict energy patterns accordingly, disaggregated techno-socio-economic approaches lack appropriate communication tools. More precisely, their empirical outputs are usually shared in their row format in the form of reports, where at most cases some illustrations are utilised (e.g. tables and charts). This makes the decision-making process generic and ineffective in certain situations where location and building 3D geometry are crucial. There is an exception for the work of Druckman and Jackson (2008), who tried to incorporate 2D-GIS and well-known geodemographic segmentation systems to improve the communication aspect of these approaches. However, 2D-GIS is less comprehensive, offers less understanding and limited ways of solving issues related to urban energy planning than 3D approaches (Biljecki et al., 2015).

In addition to the above, the lack of 3D environments in techno-socio-economic approaches also limits their applicability to the operational and maintenance stages of building life-cycle only. This is because early design stages require geometric and physical inputs such as latitude, building orientation, building volume, and aspect ratio, which can only be derived from the 3D models of the dwellings themselves. For those reasons, the development of EvoEnergy was strongly based on CityGML (3D-GIS) modelling principles. This has not only permitted to compensate the lack of geometric and physical parameters in techno-socio-economic approaches but also extend their prediction ability to cover the early design stages.

Finally, the capacity to predict future energy consumption patterns based on the household life-cycle transition patterns represents the most important difference between "EvoEnergy" and disaggregated techno-socio-economic approaches. Moreover, the major contribution of this doctoral study. This will, in turn, enable a smarter monitoring of households across their life-cycle and develop proactive strategies. This could include the development of socio-economic policies that could empower certain household types (e.g. low-income lone parents) in the long-term. Moreover, help them reduce their energy usage without affecting their thermal comfort.

## 9.3 Strengths of EvoEnergy

The developed 3D urban energy prediction tool "EvoEnergy" possesses the subsequent strengths namely; accurate and reliable predictions, integrated vision, flexible prediction granularity, sensitivity to household life-cycle transitions, potential monitoring capabilities, and public engagement capabilities. These will be discussed in more depth in the following sub-sections.

### A- Validity and accuracy of electricity prediction model

As indicated in the methdology chapter, the developed electricity prediction model would be validated against the electricity usage figures of three consumers whose dwellings were previously inspected by Nottingham energy partnership assessors as part of the Rumourban project. Table 9-8Table 9-9, and

Table 9-10, depict the demographic and socio-economic characteristics of the 3 householders, respectively. Furthermore, highlight their HVAC as well as energy expenditure. Overall, all the householders were single non-elderly and held a never-married marital status. The reason behind this choice was influenced by the high proportion of single non-elderly in the Sneinton area. The first householder was 35, working part-time, has a vocational qualification, living in 1-bedroom end-terraced bungalow built in 1969, and uses gas as the main central heating fuel. Moreover, consumed 10 to £15 on electricity on average and around 3115 KWh per year. On the other hand, the second consumer was aged 22 with a lower income (0-£10,000), lived in a 2-bedroom terraced house, had electricity central heating, spent £15 weekly on average, and had an estimated yearly consumption of 4856 KWh. Finally, the third householder was 30, lived in a 2-bedroom flat, had a low income (0-£10,000), had a gas central heating, and consumed roughly 2800 KWh of electricity per annum.

All these demographic, socio-economic, and HVAC characteristics, were input to the EvoEnergy to predict their electricity usage. Table 9-7 compares the predicted figures with the ones calculated using the energy performance certificate (EPC). It is evident that there were minor discrepancies between both annual electricity figures across all householders. More specifically, for the first householder, the annual electricity predicted by EvoEnergy was 3023.522 KWh (Figure 9-4), which in turn was 2.94% lower than the figure estimated in the EPC report. Similarly, for the second householder, the yearly electricity consumption of predicted by EvoEnergy was 4.56% lower than the expected yearly usage in the energy performance report (EPC) (Figure 9-5). Finally, the EvoEnergy estimated usage of householder 3 was 4.61% higher than the EPC expected consumption (Figure 9-6).

In order to determine the reason behind the above inconsistencies, the three EPC reports were carefully analysed. It is believed the main reason behind the high electricity consumption of householder number 1 is due to the inefficiency of the appliances, especially cold appliances that run 24 hours. Indeed, the age of fridge freezer, washing machine, television, vacuum cleaner, and cooker possessed by this consumer was 10 years. According to SUST-it (2016), a 10-year old fridge freezer consumes approximately 28% more electricity than the new one. As for the higher annual electricity figure of consumer number 2, this could be attributed to the fact that their heating duration in winter was 1 hour above the typical heating pattern for couple without children, which is 8 hours according to DECC (2014b). Finally, the lower yearly electricity usage of the householder number 3 could be related to the fact that this consumer does not possess major appliances such as fridge freezer, cooker, and washing machine, due to their low-income.

After analysing the above fact, it is clear that the electricity prediction model had a good accuracy.

Table 9-7. Comparison between the actual and estimated annual electricity usage of the three households

| | EPC yearly estimated electricity (KWh) | EvoEnergy predicted annual electricity (KWh) | Difference |
|---|---|---|---|
| **Householder 1** | 3115 | 3023.522 | -2.94% |
| **Householder 2** | 4856 | 4634.717 | -4.56% |
| **Householder 3** | 2800 | 2935.418 | +4.61% |

Table 9-8.Demographic, socio-economic, and energy related characteristics of household number 1

| | | Input | Value |
|---|---|---|---|
| **Householder 1** | **Demographic characteristics** | Household type | Single non-elderly |
| | | Householder age | 35 |
| | | Household size | 1 |
| | | Marital status | Never-married |
| | **Socio-economic characteristics** | Employment mode | Part-time |
| | | Goldthorpe class | Skilled manual worker |
| | | Income band | 20,000 - £25,000 |
| | | Benefit income band | 0 – £5,000 |
| | | Level of education | Vocational college |
| | **Dwelling characteristics** | Dwelling type | End-terraced bungalow |
| | | Year of built | 1969 |
| | | Number of bedrooms | 1 |
| | | Number of stories | 1 |
| | | Floor area | 50 $m^2$ |
| | | Tenure type | Rented from local authorities |
| | | Monthly rent | 290 |
| | **HVAC and energy expenditure** | Has central heating | yes |
| | | Main heating fuel | Gas |
| | | Bills included in the rent | No |
| | | Type of electricity tariff | Pre-Payment |
| | | Reported energy figure | 10-£15 per week |
| | | EPC Estimated yearly electricity usage | 3115 KWh |

Table 9-9.Demographic, socio-economic, and energy related characteristics of household number 2

| | | Input | Value |
|---|---|---|---|
| Householder 2 | **Demographic characteristics** | Household type | Single non-elderly |
| | | Householder age | 22 |
| | | Household size | 1 |
| | | Marital status | Never-married |
| | **Socio-economic characteristics** | Employment mode | Part-time |
| | | Goldthorpe class | Routine non-manual |
| | | Income band | 0 – £10,000 |
| | | Benefit income band | 0 – £5,000 |
| | | Level of education | Higher education |
| | **Dwelling characteristics** | Dwelling type | Terraced |
| | | Year of built | 1969 |
| | | Number of bedrooms | 2 |
| | | Number of stories | 1 |
| | | Floor area | 70 $m^2$ |
| | | Tenure type | Rented from private landlord |
| | | Monthly rent | £345 |
| | **HVAC and energy expenditure** | Has central heating | yes |
| | | Main heating fuel | Electricity |
| | | Bills included in the rent | No |
| | | Type of electricity tariff | Direct-debit |
| | | Reported energy figure | £15 per week |
| | | EPC estimated yearly electricity usage | 4856 KWh |

Table 9-10.Demographic, socio-economic, and energy related characteristics of household number 3

| | | Input | Value |
|---|---|---|---|
| **Householder 3** | **Demographic characteristics** | Household type | Single non-elderly |
| | | Householder age | 30 |
| | | Household size | 2 |
| | | Marital status | Never-married |
| | **Socio-economic characteristics** | Employment mode | Part-time |
| | | Goldthorpe class | Semi-unskilled manual worker |
| | | Income band | 0 – £10,000 |
| | | Benefit income band | 0 – £5,000 |
| | | Level of education | |
| | **Dwelling characteristics** | Dwelling type | Converted flat |
| | | Year of built | 1969 |
| | | Number of bedrooms | 2 |
| | | Number of stories | 1 |
| | | Floor area | 70 $m^2$ |
| | | Tenure type | Rented from local authorities |
| | | Monthly rent | £350 |
| | **HVAC and energy expenditure** | Has central heating | yes |
| | | Main heating fuel | Gas |
| | | Bills included in the rent | No |
| | | Type of electricity tariff | Direct-debit |
| | | Reported electricity figure | £8.5 per week |
| | | EPC estimated yearly electricity usage | 2800 KWh |

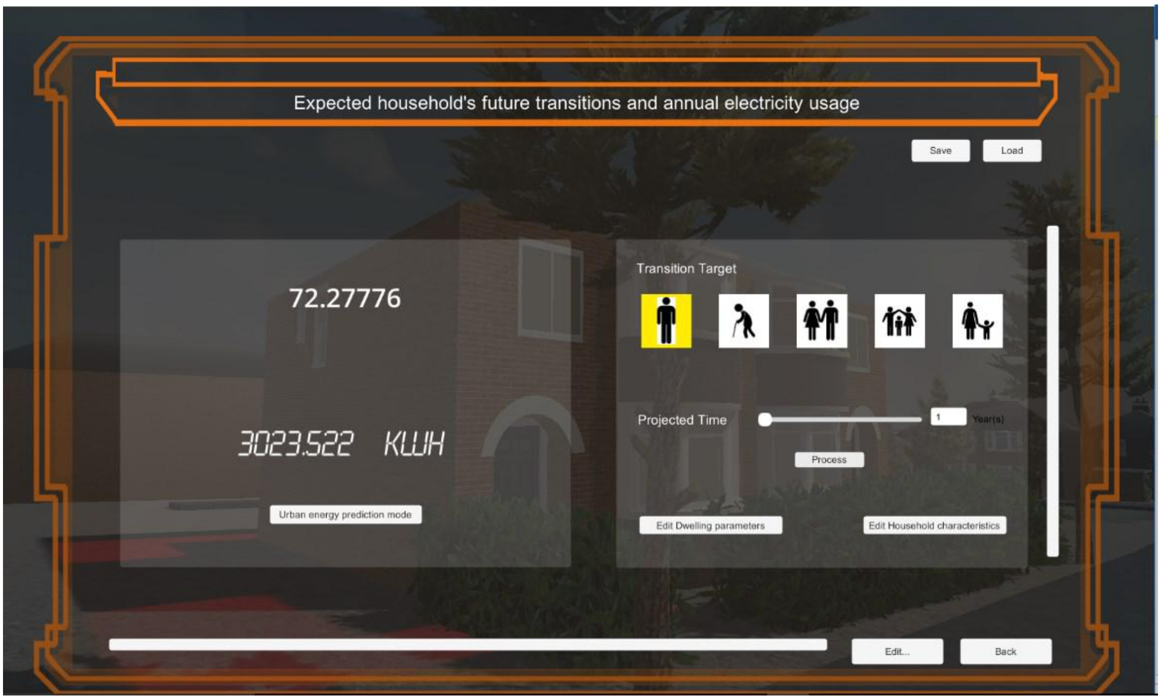



Figure 9-4. The estimated transition and annual electricity usage of household one by EvoEnergy

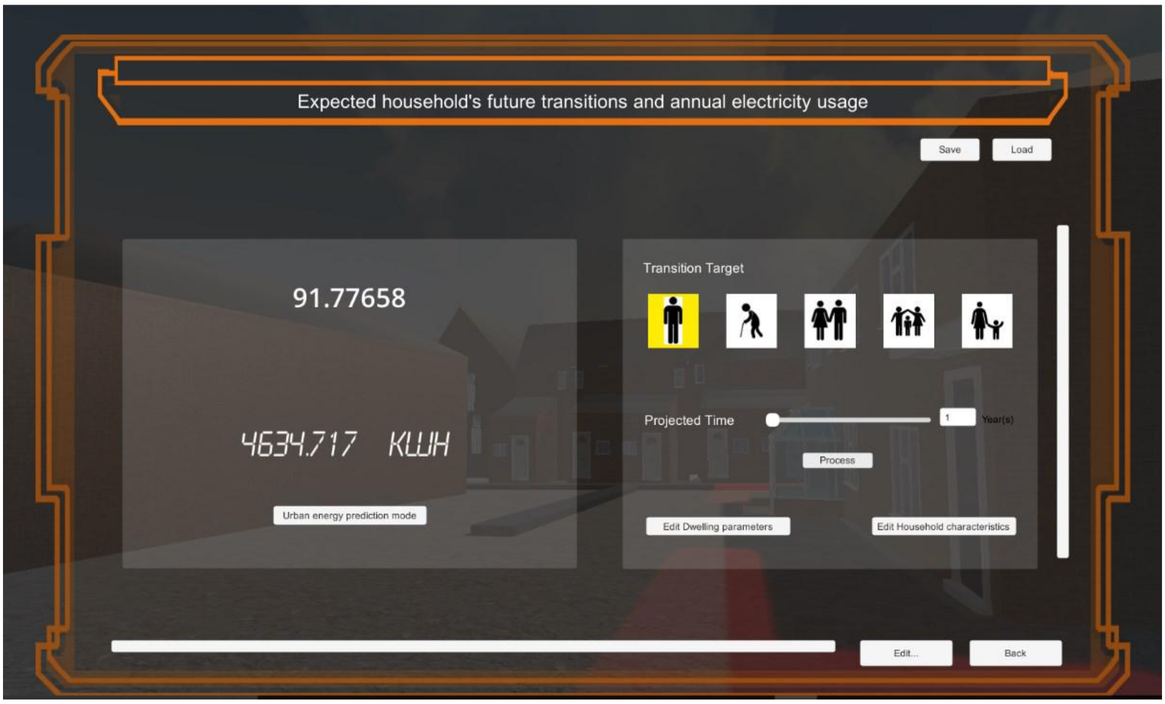



Figure 9-5.The estimated transition and annual electricity usage of household two by EvoEnergy

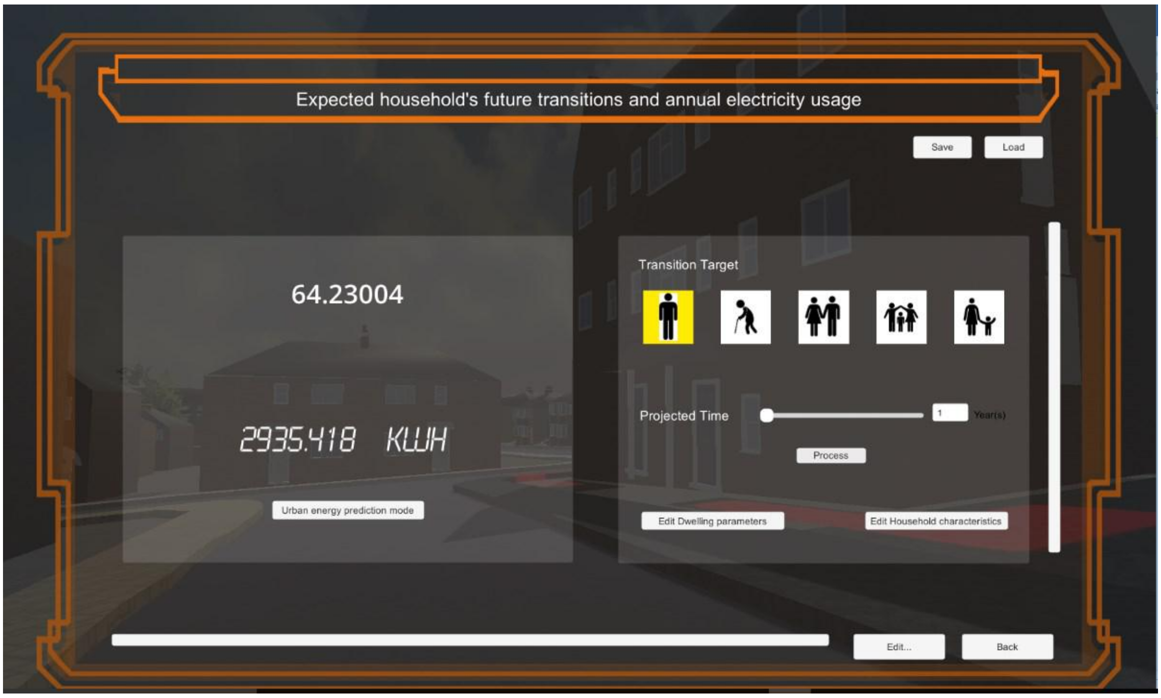



Figure 9-6.The estimated transition and annual electricity usage of household three by EvoEnergy

## B-Integral vision and principles

Initially, the departing point and the main intention behind conducting this doctoral research was to develop a 3D urban energy prediction tool which combines two distinct approaches of urban energy planning; physical and socio-economic. This was because the analysis of the relevant literature, showed that both approaches do not account for each other. Since the problem of energy sustainability in the residential sector is complex and depend on a variety of interwoven factors (e.g. socio-economic, technological, climatic, cultural, etc...), this fragmentation can lead to the generation of policies and measures that are ineffective in coping with future challenges. Therefore, the integral vision and principles of EvoEnergy represents a major strength of EvoEnergy. Primarily, this was evident in the developed framework outlining the structure and major principles of envisaged 3D urban energy prediction tool "EvoEnergy" in chapters 2, 7, and 8. Moreover, in the implementation of CityGML modelling approaches, the adoption of certain principles of techno-socio-economic approaches (e.g. use of panel data, statistical analysis, etc...). Again, EvoEnergy, in current state, is far from being a fully holistic 3D urban energy prediction tool as this necessitates more time and resources. However, since it is based on CityGML, its future development framework could host new modules such as transportation. In this way, it is possible for urban planners, for instance, to predict the CO2 emissions associated with the residential sector and transportation with the help of smart simulation tools. This allows to develop strategies and measures that could help reduce CO2 emissions in new developments such as increasing the density of the urban fabrics to reduce travel.

### C-Flexible granularity

The development of EvoEnergy energy followed an integrated bottom-up approach. For those reasons, it is equipped with a flexible prediction granularity. This implies that it is possible to predict residential energy consumption from building to the city level. This entails the building, parcel, block, district and city levels. At the building level, physical parameters are retrieved from the 3D model of the dwelling and fed to the physical module. Similarly, the socio-economic factors are obtained through a survey and fed directly into the socio-economic module. The predictions for this corresponding household are in turn computed based on the physical and socio-economic inputs. Of course, the predictions at the city level in one go can be less accurate and reliable. However, this could be improved, if the city is divided into different zones, where each one is composed of specific archetypes.

### D-Sensitivity to household life-cycle transitions and potential monitoring capabilities

Undoubtedly, household life-cycle transitions represent the backbone of this doctoral research. Indeed, EvoEnergy, in its current state, can estimate the annual electricity consumption of households based on their evolution in the next 10 years. These estimations are sensitive to the nature of the envisaged transition pattern (e.g.

couples without children) and the projected time (e.g. next 3 years). The likelihood of these transitions are in turn influenced by the demographic and socio-economic characteristics of the households themselves. This enables urban energy planners to generate policies that cope with future development of urban neighbourhoods based on their demographic and socio-economic profile.

Another strength of EvoEnergy is the possibility to monitor the energy consumption of different households accompanied with their future life-cycle transition patterns. For example, comparing the transitions and energy usage of lower-grade professional single non-elderly with routine non-manual single non-elderly in 10 years. In this respect, an interesting application would be to monitor and compare the energy figures of households who remain in the same household type. Furthermore, whose demographic as well as socio-economic characteristics do no change for a period of time. In this way, it is possible to determine the effect of other factors such as behaviour and efficiency of appliances.

**E-Public engagement capabilities**

Although this version of EvoEnergy was developed to cater the needs of urban energy planners, it can be easily improved and adapted to be used by energy consumers in the near future. This is mainly because it was initially developed within a game engine environment (Unity3D), which allows to deploy any digital content across a wide range of platforms. These include; TV, smartphones, tablets, VR, and the web. Premised on that and with the help of geo-fencing, households of a particular neighbourhood (e.g. with high energy figures) are suggested to install EvoEnergy on their smartphones. Once achieved, they are required to fill a short survey on their demographic and socio-economic characteristics. The application in turn estimates their energy consumption based on their expected family transitions patterns. Moreover, maps them onto their dwellings. The households could also navigate in the 3D city model with the possibility to employ VR technology (e.g. Gear VR).

Furthermore, they could compare their figures with similar households but with lower energy figures and obtain recommendations on how to reduce their carbon foot print. This will encourage them not only to reduce their CO2 emissions but also develop pro-environmental attitude and awareness.

## 9.4 Limitations of EvoEnergy

### A-Dependency on government data and replicability across countries

The socio-economic module of EvoEnergy relies heavily on inputs from household panel data surveys. These are collected, managed, and maintained by national agencies funded by the UK government such as understanding society. Considering that the prediction models need to be frequently updated to keep pace with a rapidly evolving society, the unavailability or access restriction to these data can be problematic. Indeed, this signifies that the only alternative route to overcome this issue would be to conduct a national panel data survey, which is impractical given the limited of time and resources for such a purpose.

In addition to the above, another limitation of EvoEnergy is that it is not replicable to other countries. In other words, EvoEnergy can only predict and monitor the energy consumption figures of British households, since it was partially developed using the British household panel data survey (BHPS). This means that its successful implementation in another country requires a new statistical analysis of its panel data household survey, which leads to the development of new prediction models. However, this is not always possible, since the availability and access to such data in certain countries, especially developing countries, can be problematic.

### B-Limited technological module

The limited time and resources present in this doctoral study prevented the development of a sophisticated technological module which takes into account the number, type, and efficiency of appliances in the household. Although EvoEnergy predicts the yearly electricity consumption with a good accuracy, the effect of certain prominent technological factors (e.g. age of appliances) is implicit. This is because the only technological parameters available in EvoEnergy are; existence of central heating, type of boiler, existence of renewable energy sources, and type of heating controls. However, the collection of such detailed and accurate inputs at the household level requires specific procedures such as energy audits (Ndiaye and Gabriel, 2011) and sub-metering (Greene et al., 1989), which might be costly to undertake at the city level. In this respect, another solution could be to incorporate the average energy figures of certain appliances from the household electricity survey, which monitored the electricity demand and consumption of 251 households in England from May 2010 to July 2011. The only drawback of this approach is that the impact of appliances' age and efficiency is not captured.

### C-Prediction module for socio-economic factors

As demonstrated in chapters five and six, EvoEnergy predicts the electricity consumption of households based on their future expected transition patterns over their life-cycle. Furthermore, the change in their socio-economic and demographic characteristics. However, in order to see the change in electricity usage in function of socio-economic characteristics, the user has to change envisaged inputs in the socio-economic module and re-run the predictions. This is because the module was not based on the **predicted** change in the households' socio-economic characteristics in 10 years, due to feasibility issues. Indeed, the accomplishment of such objective requires the development of 10 respective prediction models for **each** socio-economic variable, where each model represents a particular time in the future (e.g. next 2 years).

## 9.5 Recommendations for future work

Throughout the course of this doctoral study, certain ideas have been rejected, since they either fall beyond the scope of the dissertation or necessitate considerable time as well as resources. Other ideas were also excluded due to the lack of data availability. Premised on that and on the limitations of the developed urban energy model "EvoEnergy", this sub-section sheds the light on the most important ideas as recommendation for future work.

- **Development of simulation based-physical module**

EvoEnergy can retrieve basic physical parameters from the CityGML model itself (e.g. volume. Area, glazing ratio) to calculate the heating demand/ consumption of the household. However, this is not sufficient, since the incorporation of intelligent simulation tools within its physical module (e.g. EnergyPlus) is indispensable for assisting, monitoring, and evaluating physical improvement strategies. This does not only apply to existing residential dwellings but also to new developments. In the case of the latter, energy simulation plugins could offer a great assist in the early design stages, where urban planners can adjust certain parameters such aspect ratio, building heights, and orientation, which could lead to optimal energy efficiency.

- **Building an international socio-economic module**

Based on the fact that EvoEnergy is only confined to the UK, one of the future possibilities is to build an international socio-economic module covering different countries such as Germany, Italy, and Spain. Although this can be problematic due to data unavailability or access restriction, as indicated previously, the achievement of this module simply requires the development of new prediction models by applying some statistical techniques to these data. Furthermore, this also necessitates a new CityGML modelling of the envisaged urban areas. This will in turn enable invaluable comparisons between similar households over different countries. However, the physical module also has to adapt to changes in the climatic conditions of these countries. So that a fair decision-making is made.

- **Integrating other modules for optimal urban planning decision-making**

Making optimal or near-optimal decisions require a system that considers different aspects of urban planning, including land use, transportation and energy sustainability (Ghauche, 2010). From this premise and considering the capabilities of CityGML, EvoEnergy could be further developed to host other modules. However, this should be gradually achieved following several stages. For example, in stage 1, the requirements of urban transportation planning decision-making should be first studied. On the other hand, in phase 2, a transportation module should be developed based on these requirements and while considering interoperability mechanisms between the energy and transportation modules. In stage 3, the incorporation of the transportation module into EvoEnergy, should occur.

## 9.6 Conclusion

The aim of this doctoral study was to explore the potential of digital technologies to develop an integrated 3D urban energy prediction tool which could support sustainable energy planning decision-making in the residential sector. In order to achieve this overall aim, seven objectives were set (please see chapter 1). The research has met these objectives to the extent that further work was devoted to developing the user-interface of EvoEnergy in response to feedback from urban planners who attended the participatory workshop. The subsequent paragraphs will discuss how these objectives have been met.

The first step towards achieving the research aim was to conduct an extended literature review which explored existing urban energy planning approaches. Furthermore, evaluated their strengths and weaknesses. Based on that, two main distinct approaches were identified; Physical and techno-socio-economic. The analysis of these approaches permitted to reinforce the claims made in the introduction chapter with regards their fragmentation. Moreover, provided a strong evidence on their non-consideration of the concept of household life-cycle transitions. In addition to that, the examination of approaches supporting physical improvement strategies in chapter 2 allowed the elicitation of the modelling principles of the 3D urban energy tool "EvoEnergy". These include, overall structure, data collection procedures, 3D modelling pipeline, type of modules, workflow, and validation process.

To answer the research questions that arose from the analysis of the literature review, different methodological options were first explored and evaluated. This discussion followed a hierarchical structure (inverted pyramid) starting from philosophical position of the research to the research design. To analyse and statistically model the demographic evolution of single non-elderly in 10-15 years, secondary data analysis of the British household panel data survey (BHPS) was utilised. In particular, binary logistic regression models were developed to investigate the demographic and socio-economic factors affecting occurring transitions; consequently, predict future transitions. Although the nature of factors and their degree of influence were not consistent across models, this analysis presented some interesting findings to report. For example, the average single non-elderly household has a 20% chance of moving to a couple without children household and 13% of becoming couple with children in 5 years. However, these percentages can also increase with the householder age. More precisely, for one additional year in the householder age, the odds ratio for remaining single non-elderly in the next 5 years is lower by 44%.

After achieving the above objective, the impact of the predicted household transitions on their annual gas and electricity consumption figures was investigated. This was attained by utilising point bi-serial correlation analysis. Overall, the results indicated that households transition patterns do mostly have a significant positive effect on their energy consumption figures. However, this impact is mostly weak. Moreover, negative on the future energy usage figures of lone parent households. This provided a clear indication on the possibility of predicting residential energy consumption figurers in function of household transitions. For those reasons, a linear regression model, which comprised transitions, demographic, and socio-economic variables, was developed for the predicting future electricity usage. Continuing the previous example, approximately 56% of those who are expected to move to couple without children household in 5 years should consume between 2000-4000 KWh electricity per annum, whereas 26% of them, could use more than 4000 KWh annually.

To develop the 3D urban energy prediction model (EvoEnergy) using the relevant standards and technologies, a 3D semnatic model with a CityGML LOD3 level of detail(s) had first to be created. For those reasons, a pilot area within the city of Nottingham, Sneinton, more precisely, was selected. This choice was mainly influenced by its proximity from the researcher and the availability of secondary data and historical energy consumption data of some of its residents. This was followed by a demographic and socio-economic analysis of the pilot area in relation to Nottingham city. This analysis, in turn, suggested a dominance of single non-elderly households in both locations (Sneinton and Nottingham city). Furthermore, advised that the demographic and socioeconomic profile of the Sneinton area is nearly-representative of the Nottingham city one, in spite of few discrepancies. This dictated the particular focus placed on single non-elderly households in this research.

Following the accomplishion of the 3D semantic model, which represents the 3D virtual environment of EvoEnergy, the final development of EvoEnergy occurred. This entailed the enhancement of the initial CityGML model of the Sneinton area in a game engine environment (Unity 3D) by building a user-friendly interface which convolves the necessary modules, including physical and techno-socio-economic.

To convey the research findings to urban planners in a meaningful and useful way to support their decision-making process, they have been collaboratively working with the researcher not only to benchmark but also improve the developed 3D urban energy prediction tool. Areas that were evaluated by the group of urban planners and energy experts during the participatory workshop held At Nottingham Trent University included; the navigation in the 3D virtual environment, user-interface, and the household transition module and prediction modules. However, the participants' suggestions revolved around three main points namely; the link between the user-interface and the CityGML model, semantic querying tools, and 2D-GIS thematic mapping.

To conclude, this thesis involves a scalable and generic dimension that should be discussed. More precisely, the bottom-up nature of EvoEnergy does not only permit to predict domestic energy usage at multiple urban scales via extrapolation but also potentially allow its adoption in other countries than the UK. Of Course, this is not a straightforward task as it requires tedious pre-processing of longitudinal datasets and the development of prediction models relative to each country during a given period. Despite being possible, the scalability of EvoEnergy relies heavily on the availability and quality of longitudinal data, as discussed at an earlier stage. This implies that developing countries, who do not possess central data repositories or restrict access to data, unfortunately cannot benefit from EvoEnergy. As a result, this opens the doors to considering alternative methods for developing socio-economic modules. One of the possibilities is to utilise datasets of a given country and apply its prediction models to another one with similar cultural, climatic, and socio-economic context. This is while adjusting the weighting factors of the models to account for minor differences and eliminate bias. Another possibility could be to use data integration techniques to build a single database from multiple data sources pertaining to a given developing country. However, this approach is not cost effective due to the changing nature of datasets over time.

# A

# Appendix

# Literature review related appendix

## A.1.1 Different approaches to the calibration of building energy simulation tools

Table A. 1.Shows the different types of calibration and provides some examples on each category.

| Calibration Category | Example of case studies |
|---|---|
| **Manual, iterative and pragmatic intervention** | (Kaplan, McFerran et al. 1990); (Diamond, Hunn 1981) (Hunn, Banks et al. 1992) |
| **A suite of informative graphical comparative display** | (Haberl, Bou-Saada 1998) (Bronson, Hinchey et al. 1992) (Haberl, Sparks et al. 1996) |
| **Special tests and analytical procedures** | (Manke, Hittle et al. 1996) (Liu, Claridge 1998) |
| **Analytical/mathematical methods of calibration** | (Sun, Reddy 2006) (Westphal, Lamberts 2005) |

Table A. 1 outlines various calibration techniques which are employed in adjusting the inputs of building energy simulation tools. The classification included in Table A. 1, which consists of four distinct techniques, is the most common in the literature and was initially developed by Agami Reddy (2006).

Raftery et al. (2014) indicate that iterative pragmatic interventions rely on manual adjustment of inputs by the user. For those reasons, this process relies heavily on user experience as well as the employment of trial-and-error technique to adjust the inputs until matching the measured data. In addition to that, studies within this category complement energy bills with walk-through audits or/and short-term monitoring or even with some graphical plots (Sun, Reddy 2006). A good example of this category is the work conducted by McFerran et al. (1990) in which short-term energy monitoring (STEM) of 1month was suggested to calibrate models instead of full-year data. Once this data had been obtained, statistical analysis was applied to generate internal load schedule for DOE-2. This study achieved about a 10% difference between the calibrated and measured data on an annual basis.

Concerning the calibration based on a suite of informative graphical comparative display, series graphics (e.g. 3D plots) and graphical indices (max, min, mean, median, percentiles etc...), are employed to help determine where errors occur between the measured and simulated data. This approach often serves as a decision-making support tool for the users of the manual and iterative intervention with regards the type of parameters that should be calibrated. However, each visualisation technique is designed to serve a specific purpose. For example, the 3D plot, developed by Bou-Saada (1998) to investigate the hourly difference between the computed and actual data, is adequate for adjusting time-dependent parameters such as heating loads. The power of this representation lies in the ease of detecting minor differences. The carpet-contour plot addressed by Raftery (2011); on the other hand, visualises the mean energy consumption against two uncorrelated variables as an attempt to speed up the detection of discrepancies between measured and computed data (Figure A. 1). Although highlighted as an advantage by the authors, it is believed that this technique is unsuitable for partially correlated variables such as dry bulb temperature and direct solar irradiance, as such some users can be easily misled by erroneous conclusions. Secondly, anomalies in the building performance can be detected but only if the right or typical building operational performance is known.

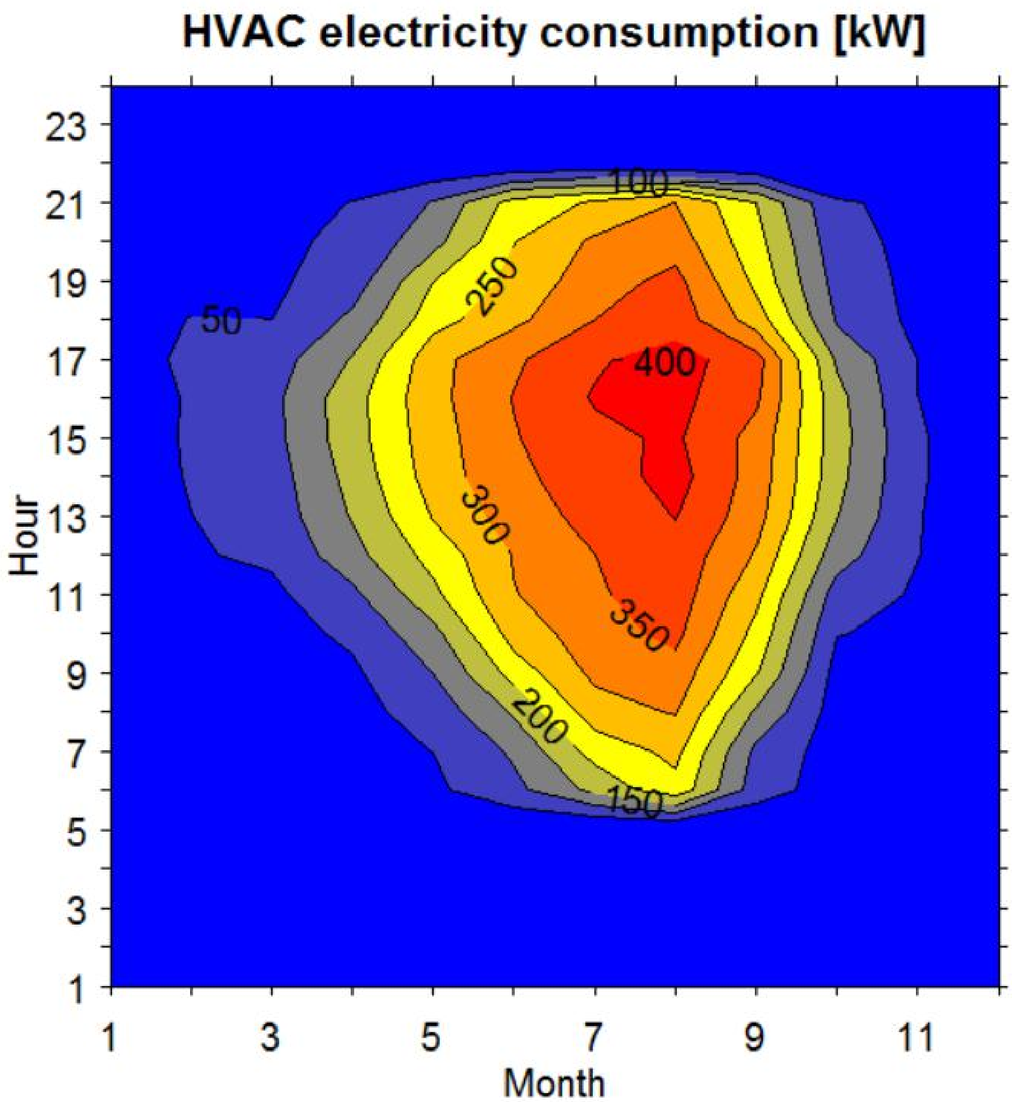



Figure A. 1.Carpet-Contour plot showing mean value of HVAC consumption against hour of day and month of year adopted by Raftery (2011).

Calibration based on special tests and analytical procedures usually uses one or the combination of the following procedures namely; Non-intrusive monitoring (e.g. Blink test), short-term energy monitoring (STEM), and energy audits. Although these measures can be also adopted in other categories (e.g. manual and iterative intervention) as discussed previously, the absence of mathematical or statistical procedures throughout the calibration process is what distinguishes it from others. A good example of calibration under this category is the work conducted by Soebarto (1997) who utilised two to four weeks (STEM) of hourly and monthly energy consumption in the calibration of two buildings models. Lighting and plug loads were non-intrusively monitored by turning on and off electrical loads for a short period (Blink Tests). Although the hourly calibration accuracy for this method was +6.7 CV RMSE (Root-mean-square error statistics obtained after cross-validation), it has been criticised for easily missing significant errors.

Studies within the last category (Analytical/mathematical methods of calibration or automatic) consider calibration similar to an optimisation problem. Therefore, it is tackled through the use of optimisation algorithms to identify automatically the most significant parameters to adjust and by how much. Westphal and Lamberts (2005), for instance, suggested a six steps calibration method for building energy simulation in which sensitivity analysis was used to detect the most influential parameters on electricity consumption. The authors found a very close match of 1% between the predicted and measured annual electricity consumption. However, calibration under this category is only efficient if fewer parameters are optimised, usually equal or less than 25 as recommended by Reddy (2006).

After analysing the above calibration approaches, the following issues have been concluded. First, apart from the lack of a universal methodology or guidance on calibration, there is no clear evidence on to what extent each calibration approach would improve accuracy. Indeed, this is believed to be highly dependent on various factors including the calibration pipeline itself, type and length of monitoring, the degree of influence of the calibrated inputs, and even the user experience as well as the interpretation of results (Agami Reddy, 2006). The nature and accuracy of site investigation seem also to be a determinant factor. This is quite evident after analysing the study carried out by Huang et al. (2007) in which the energy consumption of two office buildings in Shanghai was compared before and after calibration with monthly energy bills. After using the same calibration pipelines, the mean bias error index of the first building decreased dramatically by 31.9% for electricity, whereas for the second, this value decreased by approximately 12% which is the third of the first one. It could be argued that this difference is due to the fact that the second site investigation did not include as many details as the first one.

Secondly, it is clear that the majority of calibration approaches are not only complex but also require sufficient statistical expertise and engineering judgement (Macdonald 2002). Furthermore, since achieved iteratively, relying on site measurements, they are costly as well as time-consuming. Thirdly, calibration can be ineffective in case on-site data are unattainable or might not be practically measured. Finally, there is a paucity of approved integrated tools or automated processes that aid calibration (Coakley, Raftery et al. 2014).

### A.1.2. Comparative study between two important 2D-GIS energy estimation models

Figure A. 2 below depicts the energy prediction process followed by Jones et al. (2001) and Heiple and Sailor (2008), respectively. For reasons related to the lack of time and resources, Jones et al. (2001) decided to employ a simplified baseline model (BREDEM-9) and cost-effective data collection methods comprising quick site inspections (drive-pass) as well as analysis of historical sources. First, quick inspections were adopted to acquire data related to dwellings' geometry such as the number of floors, and dwelling type, whereas historical data analysis, was utilised to obtain dwellings age. This kind of data input along with the geometric ones were used to generate 100 archetypes. After some assumptions covering thermal characteristics had been made (e.g. U-values), the annual energy consumption and CO2 emission of each archetype were estimated using an energy simulation tool which supports BREDEM-9. More precisely, the authors adopted the domestic energy assessment procedure (DEAP). Once the calculation comprised all the 100 archetypes, certain physical measures included in the UK home energy conservation act (HECA) were applied to the estimated consumption values. Finally, the archetypes were extrapolated in GIS to thematically visualise the energy use and CO2 emission of Neath Port Talbot residential sector before and after physical improvements as shown in Figure A. 3.

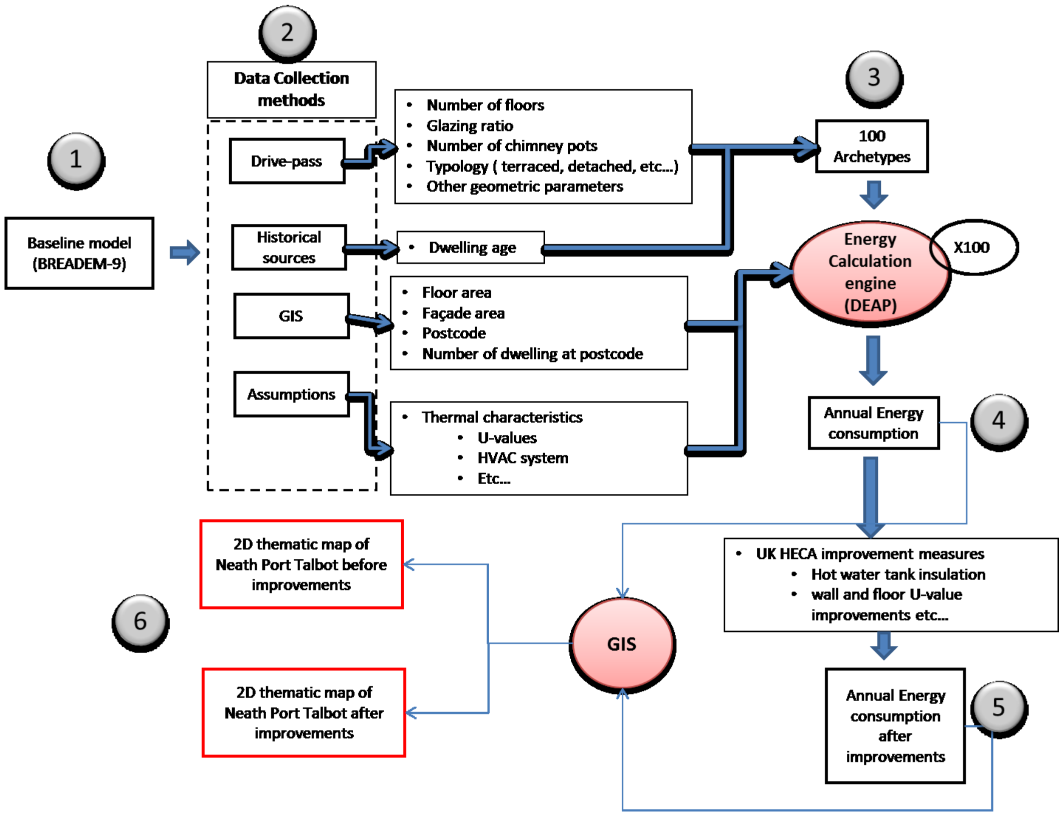



Figure A. 2.The energy estimation process adopted by (Jones, Lannon et al. 2001)

Heiple and Sailor (2008) followed a different strategy as shown in Figure A. 4, although certain methodological steps such as Archetyping, simulation, and outputs' visualisation, are common with Jones et al. (2001). The strategy includes using a more robust and profitable secondary data collection method, more precisely national residential energy consumption survey (RECS). Furthermore, a two steps validation process. The first one, which occurred after the simulation of the archetypes' energy usage intensity (EUI), consisted of comparing (EUI) values of the archetypes and existing dwellings. The inputs were adjusted until the discrepancy between the values is less than 10%. The second validation, on the other hand, encompassed the weighting of resulting annual energy consumption values against the ones estimated with top-down models.

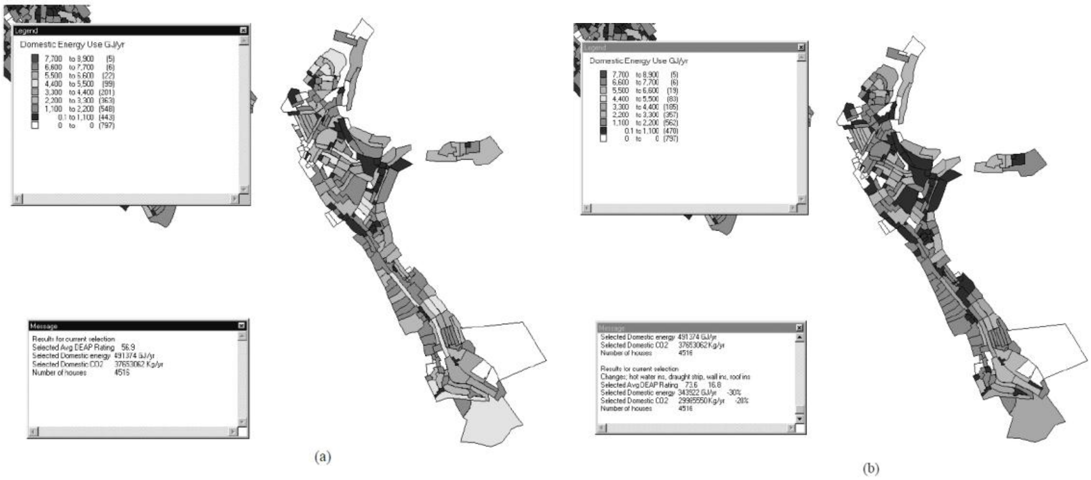



Figure A. 3.Expected energy consumption and CO2 emission of Neath Talbot residential sector before (a) and after (b) physical measures have been applied.

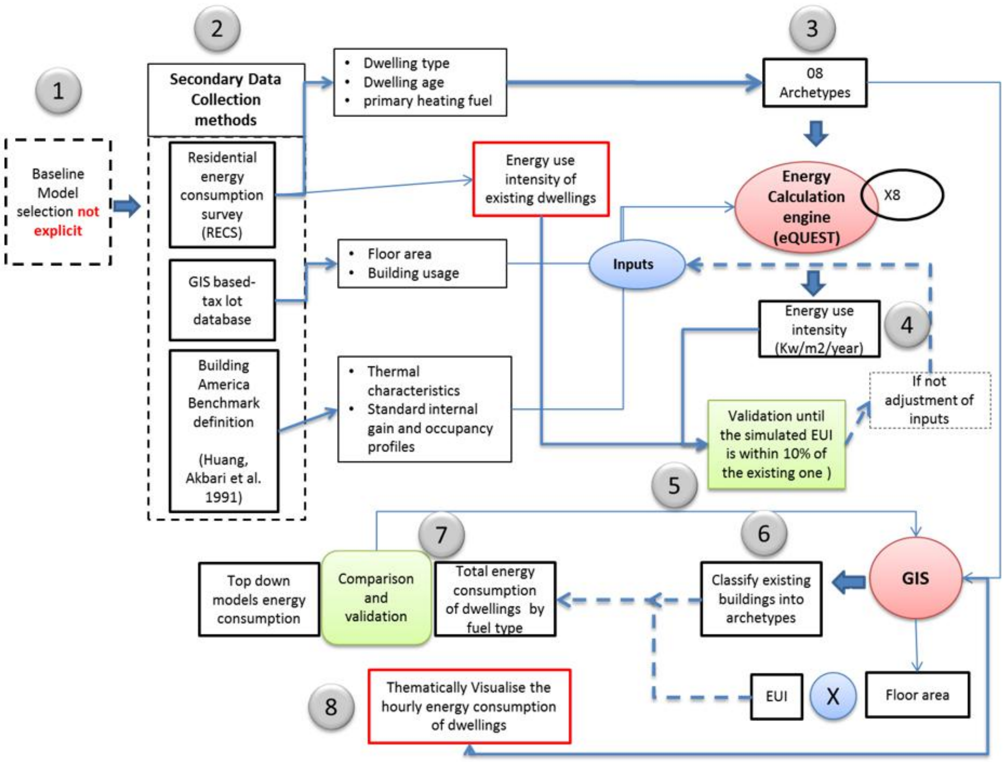



Figure A. 4.The energy estimation process adopted by Heiple and Sailor (2008)

### A.1.1.1 Reliability and consideration of occupants' behaviour

In the light of the below Table A. 2, it is clear that energy forecasting using the models of Jones et al. (2001) and Heiple and Sailor (2008) is time-consuming. Indeed, this appears to be not only associated with the analysed models but also with this approach in general since it relies heavily on building energy simulation tools.

Table A. 2.Reliability and consideration of occupancy factors of (Jones et al. 2001; Heiple, and Sailor 2008).

| Researchers | Reliability | Time - effectiveness | Occupancy schedule | Demographic Socioeconomic Variation | Drawbacks |
|---|---|---|---|---|---|
| **(Jones et al. 2001)** | Highly questionable | low | Not acknowledged | Not considered | -Archetypes not generated according to social aspects (e.g. occupancy)<br><br>-Lack of validation process<br><br>-Inaccurate data collection methods |
| **(Heiple and Sailor 2008)** | Fairly Reliable | low | Standard (typical) | Not considered | Archetypes not generated according to social aspects (e.g. occupancy)<br><br>-EUI values of different household sizes were assumed to be identical |

These tools, as described at an early stage, require many specific inputs which necessitate the adoption of elaborate data collection methods such as energy audits. Therefore, assumptions are used when these parameters are inaccessible. Another major drawback of the approach of Jones et al. (2001) is the highly questionable reliability of their model due to the lack of validation process, although the researchers claimed that BREDEM-9 is a well-recognised model and does not need any validation. On the contrary, it is believed that it is indispensable since the overall performance of the energy prediction model can be affected by uncertainties in the data inputs and not necessarily by the baseline model itself. Indeed, validation is not only crucial for the success of the developed model as claimed by (Mathews, Etzion et al. 1997) but also it could offer information on how much occupants' behaviour differs from that of the typical household (Strzalka et al. 2011). Apart from the prominence of validation, it seems that both models at most applied standard internal gain and occupancy schedule profiles. In other words, this implies that households living in similar typologies regardless of their socioeconomic differences consume energy identically, which is unrealistic. This is due to the non-consideration of occupancy factors when building archetypes, which could be in turn caused by the limitation of these models in capturing of demographic and socio-economic variation of households as explained by (Hausfather et al. 2010) in the following statement;

*"…engineering-based approaches and their limitations in capturing variations due to demographic and socioeconomic-driven behavioural characteristics somewhat restrict their usefulness in large-scale high-resolution analysis"*.

Other researchers including Larsen and Nesbakken (2004); and Swan and Ugursal (2009) also share similar views on this issue.

### A.1.3 CityGML building module and its levels of detail(s) (LOD).

#### LOD0

It is the coarsest level of detail as buildings are visualised in function of their roof outline (or footprint) in a 2.5D fashion (Figure A. 5). The displayed semantic information includes, *class, function, year of construction, type of representation,* as depicted in Figure A. 5. Although LOD0 is rarely employed in 3D applications, it is still used for other purposes such as hydraulic modelling (water bodies) and route planning (Gröger and Plümer 2012).

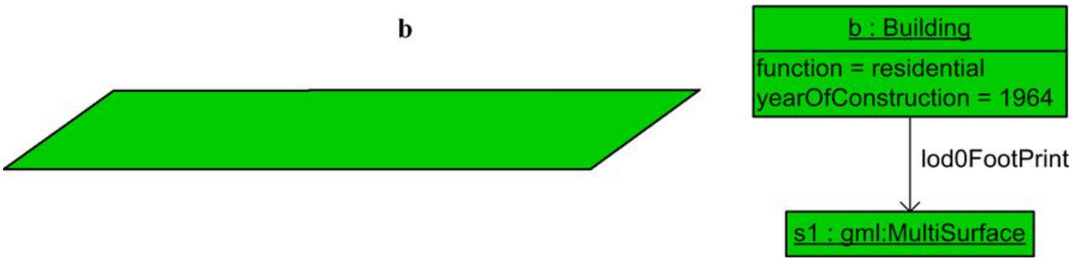



Figure A. 5.LOD0 structure (Gröger and Plümer 2012).

## LOD1

It is simply the extrusion of the building footprint horizontal surface in LOD0 into a block. As illustrated in Figure A. 6, buildings in LOD1 can be subdivided into different entities while sharing the same class, function, and construction year. However, external buildings components such as dormers, balconies, and cantilevers, are not accounted for. Another prominent feature in LOD1 is that terrain intersection with buildings can be identified and displayed as a curve. Each building subpart contains an ID, a representation type, and a roof type. However, the roof visualisation does not necessarily match the roof type in reality (e.g. gabled roof) (Krüger, Kolbe 2012). Therefore, it is not suitable for accurate energy consumption prediction applications since volumes are not accurate. Furthermore, the estimation of other parameters such as solar gain is unrealistic.

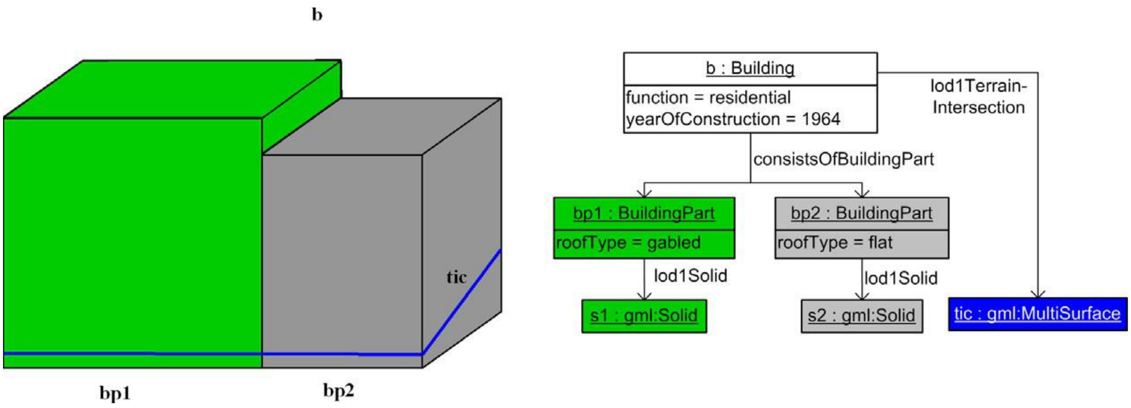



Figure A. 6.LOD1 geometric and sematic representation (Gröger and Plümer 2012).

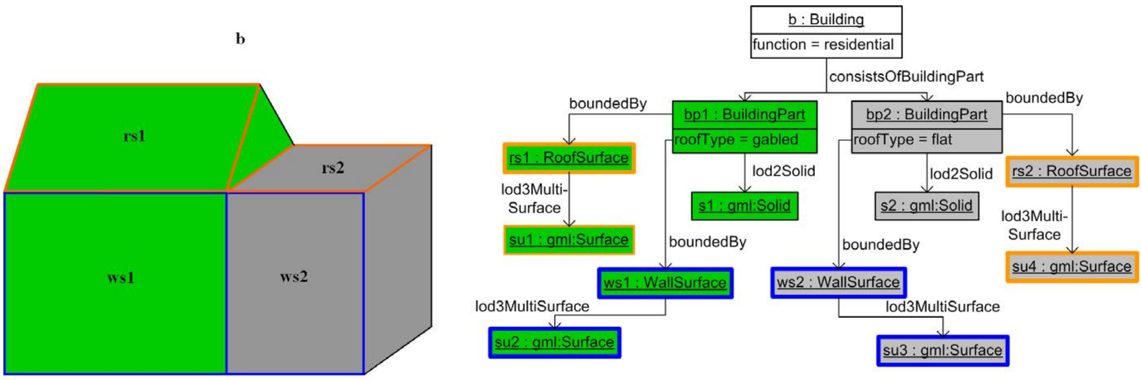



Figure A. 7.LOD2 geometric and sematic representation (Gröger and Plümer 2012).

## LOD2

The main difference between LOD1 and LOD2 is that the latter contains a rough roof shape structure. Moreover, a thematically distinguished bounding surfaces such as wall, roof, and ground surfaces, as shown in Figure A. 7. In addition to the above issues, this LOD is not suitable for accurate urban energy prediction application due to the absence of important building components such as shading devices, and openings.

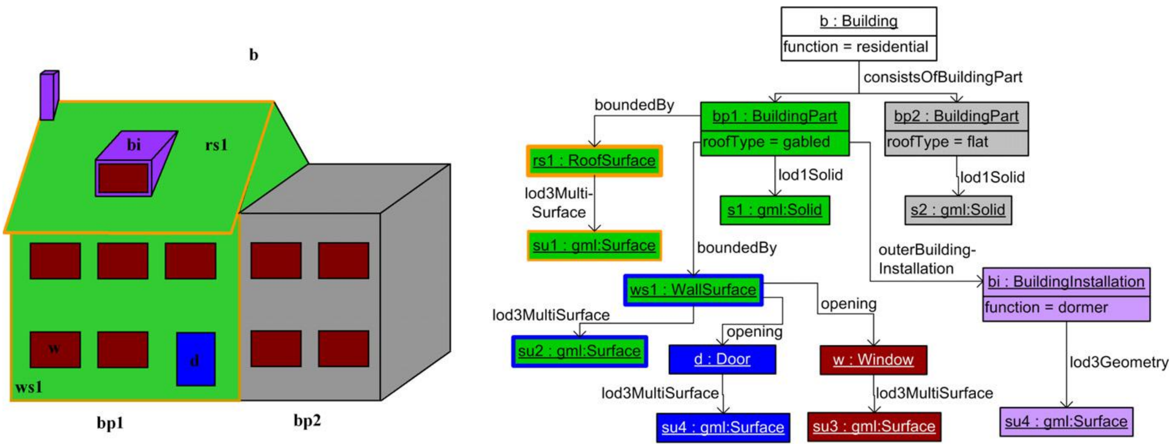



Figure A. 8.geometric and sematic representation of LOD3 (Gröger and Plümer 2012)

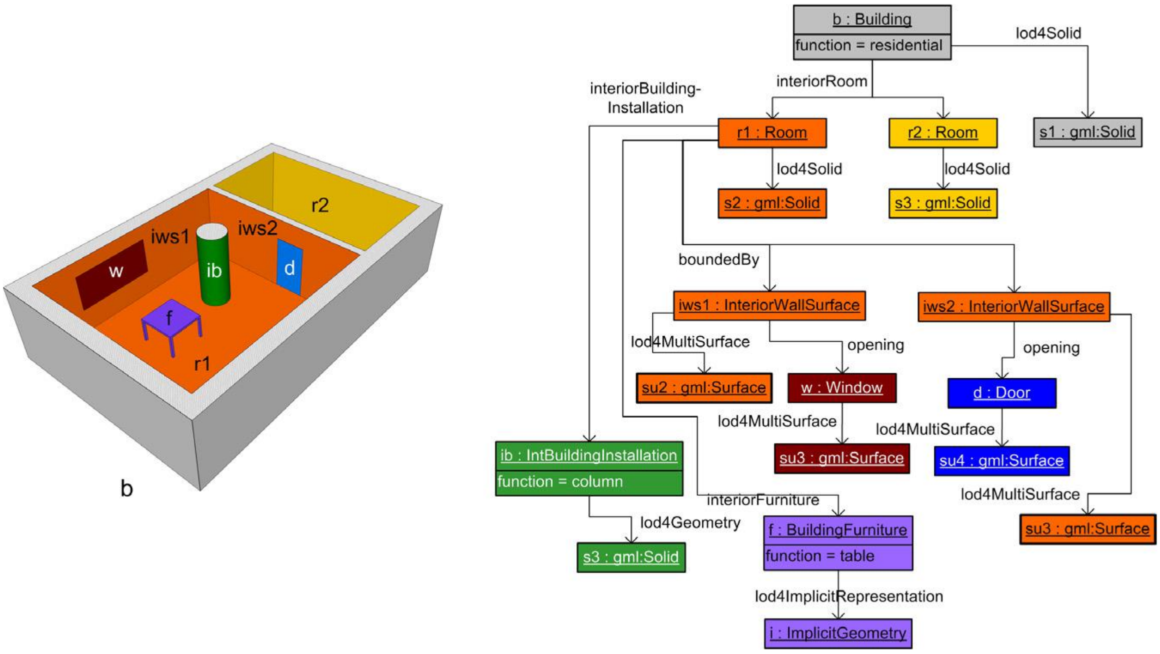



Figure A. 9.Geometric and sematic representation of LOD4 (Gröger and Plümer 2012)

## LOD3

It is achieved by adding a detailed roof structure, facades components including openings, and outer components (dormers, balconies, etc…) to LOD2 models. Similar to BIM, elements such as openings, dormers, and chimneys belong to the initial object/surface they were assigned to, although can be presented with separate attributes as demonstrated in the Figure A. 8 above.

## LOD4

It complements LOD3 by considering the buildings' interior structure. The representation of interior elements is performed hierarchically. For example, as illustrated in Figure A. 9, all the objects located within room 2 space including interior walls, doors and windows have to be linked to that room which is in turn connected to the main building.

### A.1.4 Semantic mapping of IFC and CityGML LOD4 building module

There exist around 900 classes in IFC. However, the research conducted by Van Berlo (2009) has shown that only 60-70 can be transformed to CityGML in which only 17 representing the building elements as shown in Table A. 3. Due to their importance for urban energy planning, the following classes namely; *building structure, building space, building windows and doors, building slabs and roofs, building walls, and building installations*, will be compared across both standards; CityGML and IFC.

Table A. 3.Relevant IFC classes that can be transformed to CityGML

| IFC class | CityGML Type |
|---|---|
| **Ifc Building** | Building |
| **Building address** | Address |
| **IfcWall** | Wallsurface/ Interior Wallsurface |
| **IfcWindow** | Window |
| **IfcDoor** | Door |
| **IfcSlab** | RoofSurface or FloorSurface( depending on the storey location) |
| **IfcRoof** | RoofSurface |
| **IfcColumn** | Column |
| **IfcFurnishingElement** | BuildingFurniture |
| **IfcFlowTerminal** | FlowTerminal |
| **IfcBeam** | Beam |
| **IfcSpace** | Room |
| **IfcStair** | Stair |
| **IfcRailing** | Railing |
| **IfcAnnotation** | Annotation |

## Building structure

From analysing Figure A. 10 , it is clear that a building in IFC is clearly divided into storeys and spaces defining them, whereas, in CityGML LOD4, it comprises rooms where each one is encircled by boundary surfaces. In other words, CityGML building module lacks an explicit definition of storeys unless defined by users as the aggregation of all building components at a given height, which makes it less smoothly structured than IFC (Donkers, 2013). To conclude, this comparison has shown that both standards structures are completely different from each other.

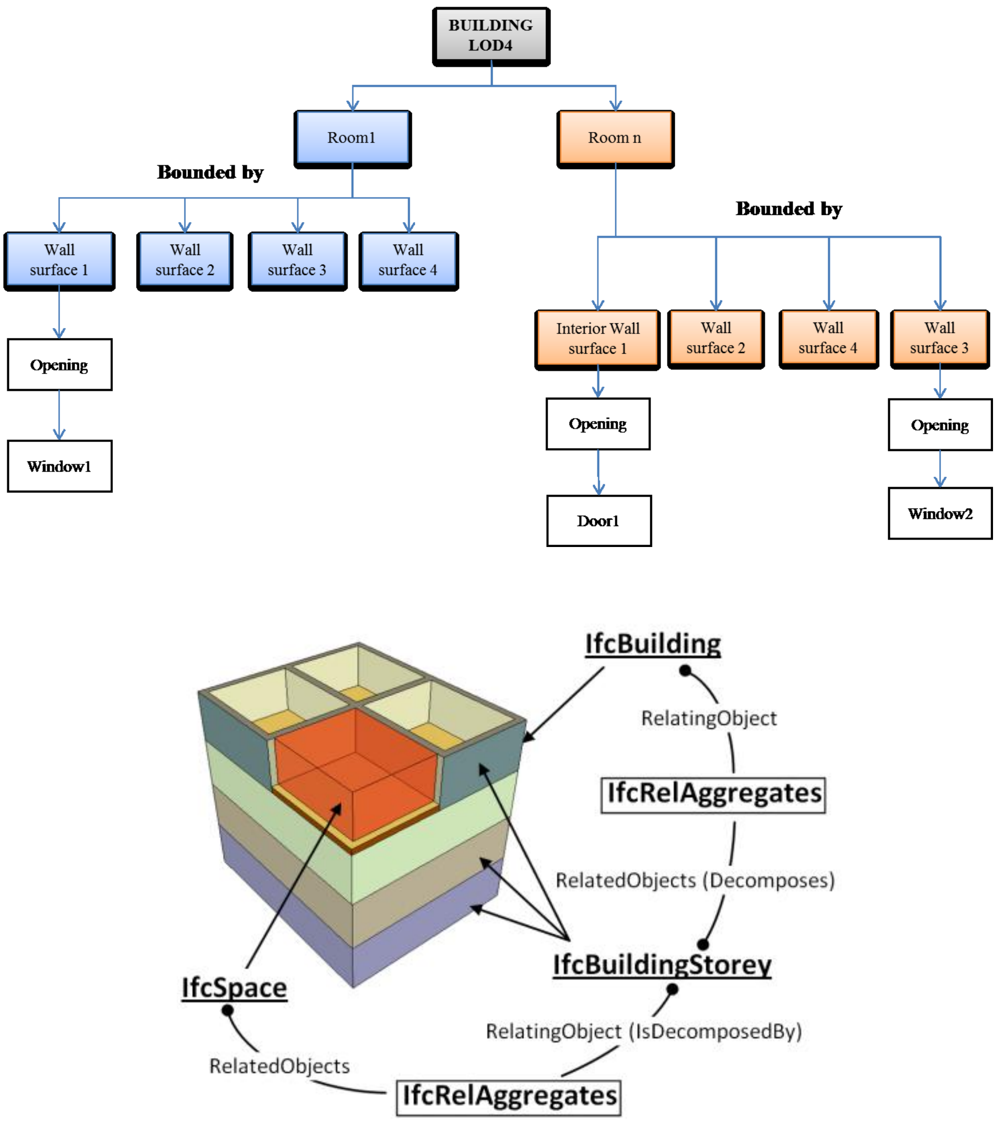



Figure A. 10. Structure and geometric representation of CityGML LOD4 and IfcBuilding

## Building space

Table A. 4.semantic mapping and transformation steps from IFC to CityGML adopted from (El-Mekawy et al., 2011).

| IFC Elements | Transformation details | CityGML Feature types |
|---|---|---|
| **IFCOpening Element** | Checking the relation IFC_RelFillsElement of IFC0peningElement with the IFCDoor or IFCWindow element: then the properties of IFCDoor or IFCWindow are attached to the respective IFCOpeningElement. IFC0peningElement is converted into Door or Window MultiSurface geometries in CityGML | Window MultiSur-faces<br>Door MultiSurfaces |
| **IFCSpace** | IFCSpace geometry, which often is a parametric geometry in IFC is converted into boundary representation geometry and translated into a Room feature (LoD4Solid) in CityGML. | ROOM |
| **IFCSpace** | IFCSpace is converted into multiSurfaces. Based on the height and relative altitude of IFCSpace the decision about each surface is taken, whether it is a CeilingSurface or a FloorSurface. If the height is between specific thresholds, then it is tagged as InteriorWallSurface. Furthermore, Window and Door surfaces are deducted from InteriorWallSurfaces. | FloorSurface<br>CeilingSurface<br>InteriorWallSurface |
| **IFCWall** | IFC-Wall is converted into multisuifaces. The multiswfaces are translated to WatiSurfaces in CitvGML, which represent the exterior shell of the building and have no connection to the Room feature type | Wall-Surfaces |
| **IFCStairs**<br>**IFCBeam**<br>**IFCColumn** | The *IFCStairs, IFCBeam,* and *IFCColumn,* are translated into multisurface boundary geometries in CityGML Moreover, IFC elements, which are within a specific room are transformed into *IntBuildingInstallation,* | IntBuildingInstallation |

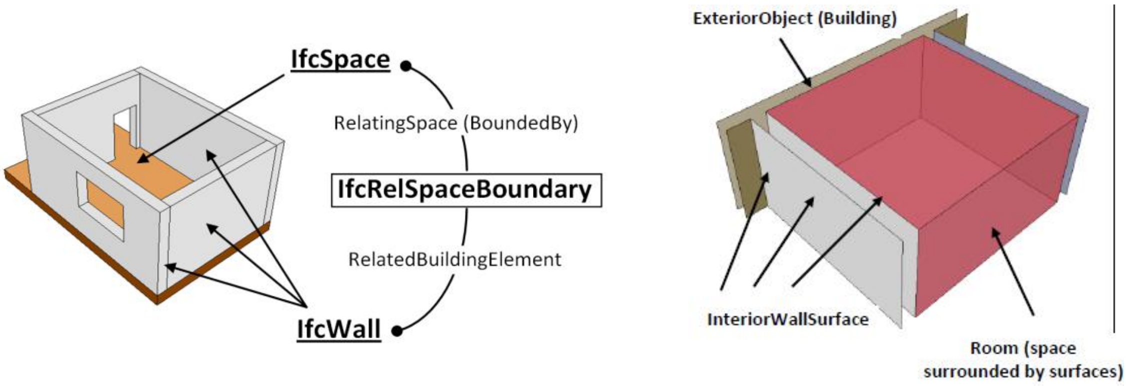



Figure A. 11.Building space definition in IFC and CityGML

Since IfcSpace is linked to IfcBuildingStorey, it does not only define areas but also volumes bounded by related building elements (IfcSpaceBoundary). This comprises both building rooms and circulation areas. Although CityGML Rooms does have the same concept as IfcSpace, CityGML rooms are surrounded by bounding surfaces instead of solid elements as illustrated in Figure A. 11 and Table A. 4. Furthermore, CityGML rooms are usually represented as a GML Solid or multisurface, whereas in IfcSpace, it is a combination of Constructive solid geometry (CSV), sweeping, and infrequently Brep (bounding representation). For these reasons, the conversion of IfcSpace to CityGML room is achieved by means of Brep geometry model transformation of IfcSpace (Khan, Donaubauer et al. 2014). It should be also noted that is possible to derive floorSurface/ CeillingSurface/ and InteriorWallSurface from IfcSpace through its conversion to multisurface involving users' judgement with regards the nature of elements (e.g. roof or ceiling).

### Doors and windows

In CityGML, windows and doors are sub-classes of the Opening class (Figure A. 13). This does also apply to IfcWindows and IfcDoors which are both linked to IfcOpeningElement as shown in Figure A. 12. The role of this latter is to geometrically and semantically describe openings (Donkers 2013). However, the mechanism that defines doors and windows across both standards is slightly different. For example, a two-step process is employed to create IfcWindow or IfcDoor in IFCBuilding model as depicted in Figure A. 12. First, by generating an IfcOpeningElement through creating a voiding element (IfcRelVoidsElement) in the IfcWall. Secondly, by using IfcDoor or IfcWindow to fill the voiding element. On the other hand, the generation of CityGML doors or windows is achieved through a combination of surfaces such as Exterior /InteriorWallSurface, GoundSurface or FloorSurface (Figure A. 13). The conversion of IfcDoor/IfcWindow to CityGML Door/ window is achieved following the subsequent steps. To begin with, the relationship between IFCRellFills and IfcDoor/ IfcWindow is first investigated. Afterwards, the characteristics of IfcDoor / IfcWindow are embedded in IfcOpeningElement. Finally, this latter will be converted into multisurface door or window (El-Mekawy et al., 2011).

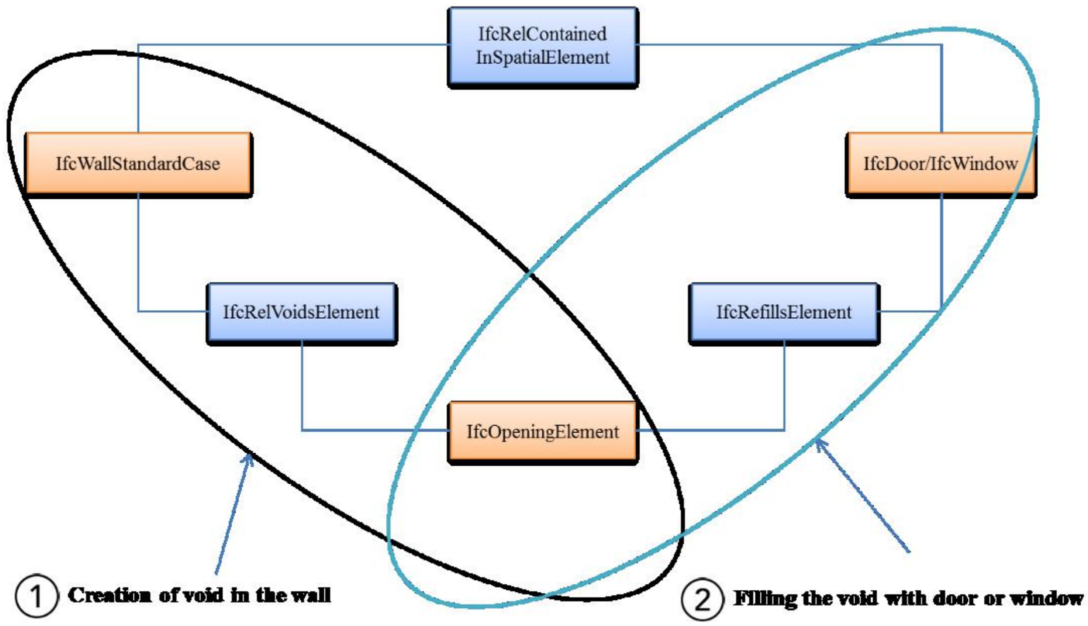



Figure A. 12.Doors and windows in IFC

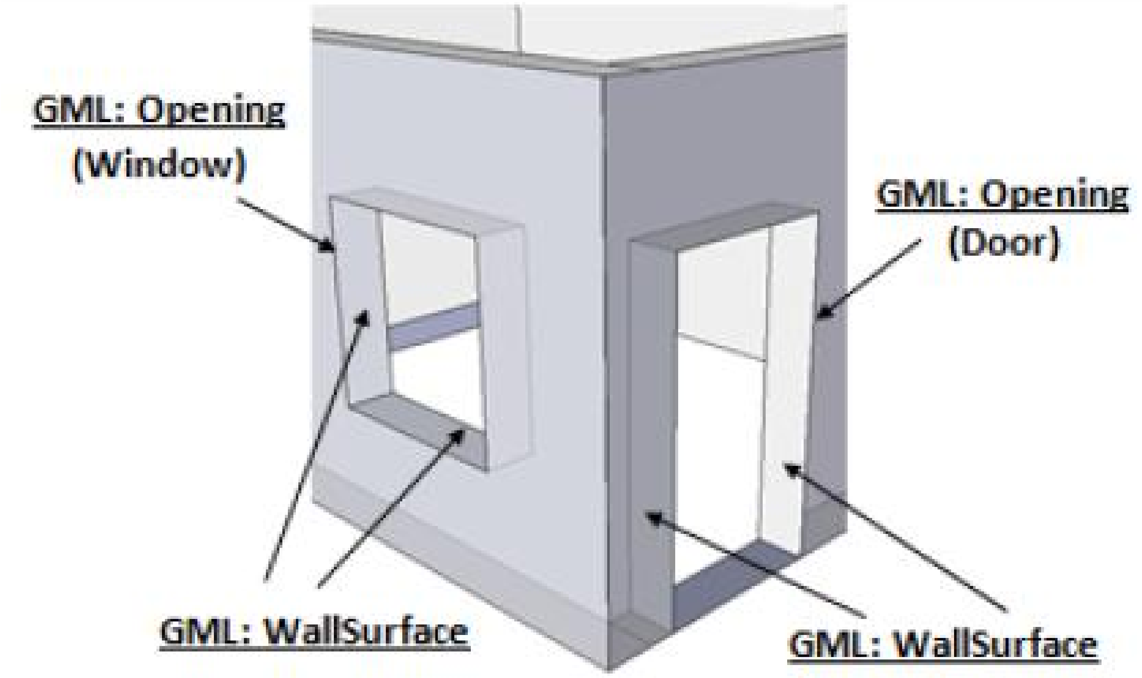
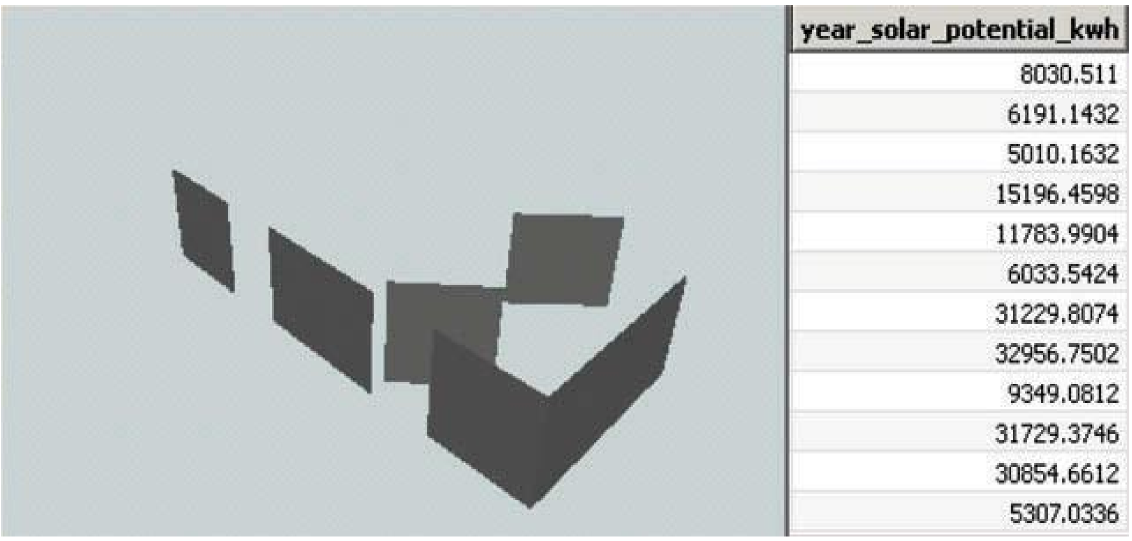



Figure A. 13.Door and windows in CityGML

## Slabs and roofs



| year_solar_potential_kwh |
|---|
| 8030.511 |
| 6191.1432 |
| 5010.1632 |
| 15196.4598 |
| 11783.9904 |
| 6033.5424 |
| 31229.8074 |
| 32956.7502 |
| 9349.0812 |
| 31729.3746 |
| 30854.6612 |
| 5307.0336 |

Figure A. 14.Roof and slabs in IFC and CityGML

Any roof or slab, including its components in IFC, is represented as one solid object as IfcRoof or IfcSlab, respectively. On the other hand, in CityGML, each one is composed of two surfaces only without considering other components (e.g. roof structure). For example, in the case of a CityGML Slab, its upper surface is considered as FloorSurface for the upper building level, whereas its lower one represents a CeillingSurface or GroundSurface for the lower level as shown in Figure A. 14. As for CityGML Roof, the upper surface is a RoofSurface, whereas the lower one is a ceiling surface for the below storey.

## Walls

Like slabs and roofs, a wall in IFC (IfcWall) is geometrically represented by its volume as a solid object which also embodies all its constituting elements and characteristics including material(s), dimensions, colour, thermal characteristics and others. However, this is different in CityGML as a wall is delineated by only its visible surfaces; GML: InteriorWallSurface and GML: WallSurface as depicted in Figure A. 15.

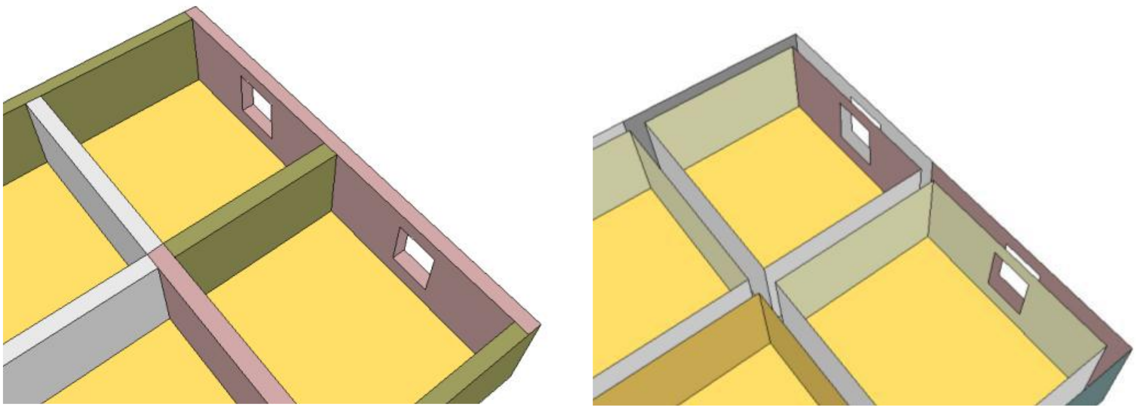


Figure A. 15.Walls representation in IFC(left) and CityGML(right)

# Part II

## Methodology appendix

### A.2.1 The Embedded design

This design depends on integrating both quantitative and qualitative data. However, one dataset has a secondary role as it would serve as a support to the other dataset which has a primary role in the study, as depicted in Figure A. 16. This design is used when one dataset is insufficient to answer the research question, or when there are different research questions to be answered in which each one requires different data types (Lieberman, 2005). An example of this research design includes experiment research, where the investigator encompasses qualitative data (e.g. from focus groups) to analyse the process of intervention, develop a treatment, or to act upon the experiment results. However, despite the potential of this design in terms of being less time-consuming and more logistically manageable especially by graduate students (Creswell, 2013), it is prone to the flowing issues. First, it is very challenging to integrate results when different research questions are present in the study. Indeed, this could lead to not addressing certain research questions adequately given the imbalanced integration of research methodologies. Secondly, this design requires very careful planning as the researcher has to decide on the primary/ secondary nature of the collected data, their purpose, and timing while considering that one data complements the other (Johnson et al., 2007).

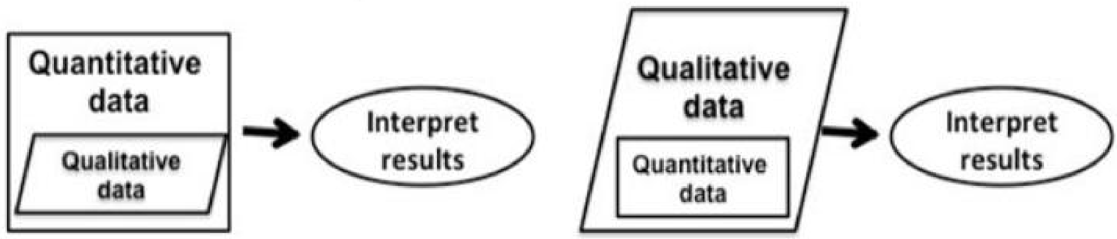



Figure A. 16.Mixed-methods embedded design, (Creswell, 2013).

### A.2.2 The Explanatory design

It is a dual-phase mixed methods design in which qualitative methodology serve as interpretative mean to existing quantitative results (Figure A. 17) (Creswell, 2013). Instances where this design in encompassed include; adopting qualitative data to explain results' significance or even surprising outcomes (Morse, 2003); supporting a purposeful sampling of groups or participants based on quantitative outcomes before proceeding to qualitative research in the second phase (Tashakkori and Teddlie, 2010). The main advantage of such design lies in straightforwardness of implementation as the investigator has the ability to carry out two separate investigations with different methods and gather one type of data individually. Subsequently, this will provide a clear delineation for the reader since the writing process of the final report is pretty straightforward. Another benefit is that it lends itself to other research designs including single mixed-methods and multi-levels inquiry, which will be addressed subsequently (Creswell and Clark, 2007). However, despite those benefits, this design requires a lot of time, especially for novice researchers who are not aware that qualitative phase takes more time than the quantitative one (Creswell, 2013). In addition to that, one major disadvantage of explanatory designs is their paucity of any specification about sampling in the qualitative phase because this primarily depends on the quantitative phase (Creswell and Clark, 2007). This hinders its approval by institutional review boards. For those reasons, the explanatory research design is excluded from this study.


Figure A. 17.The Mixed-methods explanatory design, (Creswell, 2013).

### A.2.3 The Exploratory design

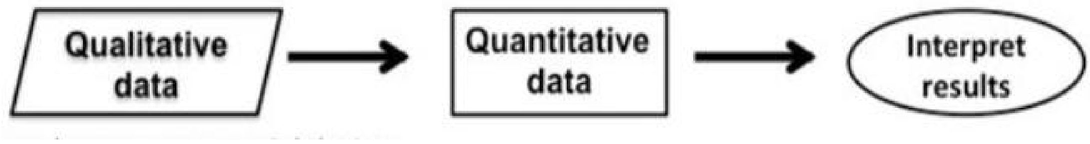



Figure A. 18.The Mixed-methods exploratory design, (Creswell, 2013).

Like the explanatory design, the exploratory one is also a dual-phase research design. However, the only difference is that the second research method (quantitative), is informed by the outcomes of the first method (qualitative) (Greene et al., 1989), as depicted in Figure A. 18. It is implemented when there is a lack of instruments or measures, uncertainty about the nature of the employed variables, or there is a paucity of clear guiding theories or frameworks on the subject. Moreover, due to the fact that an enquiry using this design begins qualitatively, it is adequate for exploring phenomena (Creswell et al., 2003). Another advantage of this design is that it could serve as a generalisation tool for the qualitative results in order to cover the wider population. In addition to that, it is also possible to evaluate various aspects of emerging concepts or categorisations as well as measuring a phenomenon prevalence after having explored it in depth (Morse, 2003; Morgan, 1998). Exploratory design shares many advantages and limitations with the explanatory one. Amongst these benefits, is the straightforwardness of preparing, implementing, and interpreting the research design as well as findings. Similarly, exploratory designs can be easily adapted in multi-level research studies (Creswell and Clark, 2007). On the other hand, besides being a time-consuming approach, it can be rejected by institutional review boards due to the lack of specification of the qualitative phase procedures. Therefore, the exploratory design is not considered relevant to this study.

### A.2.4 The triangulation design

Triangulation is considered the most common and well-documented mixed-methods research design with a solid theoretical foundation (Creswell et al., 2003). Its main goal is to acquire different, yet complementary, types of data on the subject for better addressing the research questions (Morse, 2003). In this way, the strengths of quantitative methodology (e.g. large sampling, generalisation, ability to deal with trends) are merged with the ones of the qualitative methodology (e.g. in-depth and details) to eliminate their weaknesses. It is employed in situations where the investigator intends to compare and contrast quantitative with qualitative findings. Furthermore, it is adequate when the main aim is to validate or further develop quantitative outcomes with qualitative data (Patton, 1990). Another advantage of the triangulation design is its flexibility since it is possible to collect and analyse each data type independently. For those reasons, it is regarded suitable for the undertaken research. However, the triangulation research design umbrella comprises four different variations namely; the convergence model, transformation model, the validating quantitative data model, and finally the multi-level model (Creswell, 2013). Based on that, the suitability of these models in the current study will be discussed.

### A- The convergence model

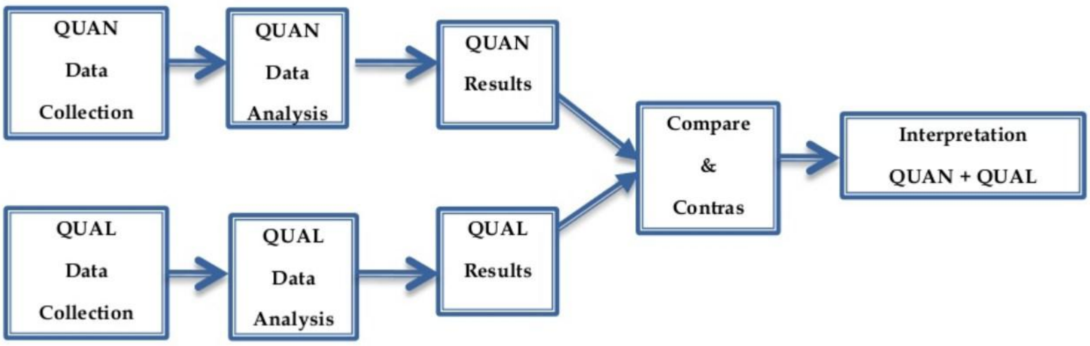



Figure A. 19.The convergence model (Tashakkori and Teddlie, 2010)

The convergence model, which depicts the traditional triangulation model, is based on collecting qualitative and quantitative data individually before converging their outcomes in the interpretation phase (Figure A. 19) for comparison or validation purposes. In this way, it is possible to attain well-substantiated and valid conclusions about the studied subject (Anderson et al., 1999). However, the main problem is that it is very challenging to converge two different types of data and outcomes meaningfully. Thus, the researcher has to utilise quantitative and qualitative data to address the same research aspects. Furthermore, has to consider the implications of different sampling methods and sample sizes. After analysing the above facts, it is clear that the convergence model is not seen relevant to this research.

## B-The transformation model

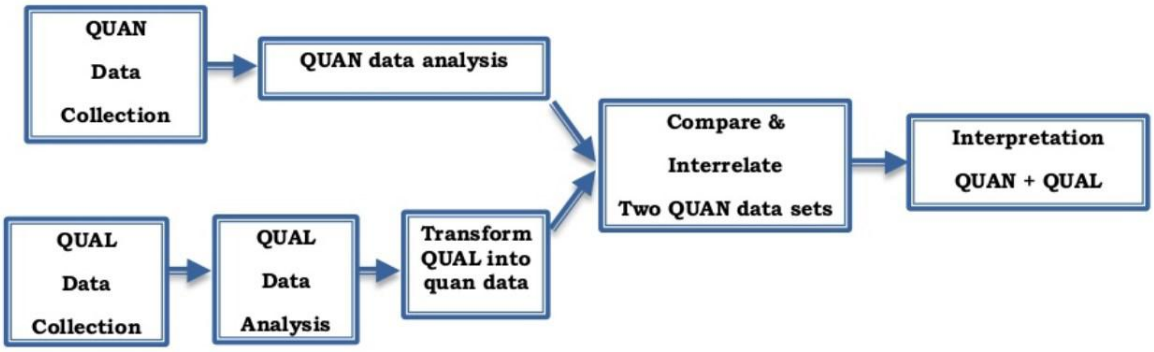



Figure A. 20.The transformation model (Tashakkori and Teddlie, 2010)

Like the convergence model, the transformative one also involves separate data collection procedures and analysis to quantitative and qualitative data (Figure A. 20). However, certain transformation procedures are applied to one data type (either qualitative or quantitative) after performing an initial analysis (Teddlie and Tashakkori, 2003). The aim of this transformation is to facilitate the comparison and interpretation of both datasets. A good example of a research adopting the transformative model is the work of Pagano et al. (2003), where qualitative data were employed to derive qualitative themes. Afterwards, these qualitative data were transformed into quantitative by coding the developed themes dichotomously, where 1 present and 0 not-present. This, in turn, enabled the researchers to apply some statistical procedures such as binary logistic regression and correlation analysis to identify possible associations between different themes as well as gender. However, the disadvantage of this approach is that it requires skilled investigators, since an awareness of different data transformation instruments prior to conducting the research, represents a necessity (Creswell, 2013). Considering the above facts, the transformation model is not adequate for the current research since the majority of envisaged qualitative data arise from participants' feedback on visual and functional aspects of the developed urban energy model, which would not be transformed into quantitative data.

### C-The validating quantitative data model

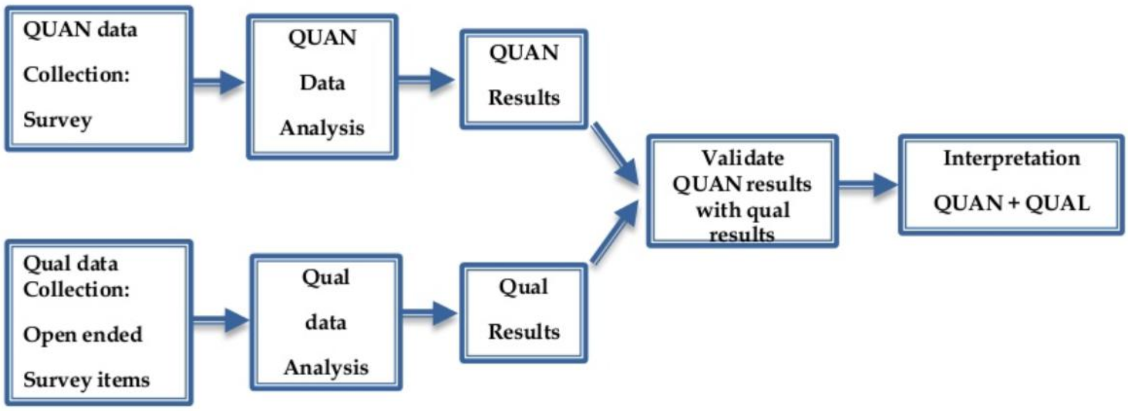



Figure A. 21.The validating quantitative (Tashakkori and Teddlie, 2010)

This model is used when the investigator intends to validate or expand on survey findings (quantitative) by the introduction of some open-ended questions, as shown in Figure A. 21. For those reasons, both data types are collected simultaneously when the survey is conducted (Creswell and Clark, 2007). For example, in their study of mass casualty incidents' psychological effect on forensic odonatologists, Webb et al., (2002) incorporated open-ended questions with their survey measures.

### D-The multi-level model

In this model, the research implementation process is divided into different complementary manageable phases in which each level uses a particular methodology and methods for data collection and analysis. Once an initial analysis is performed, outcomes from each level will be combined together as one unit in order to be interpreted (Figure A. 22), which provides a lot of flexibility and control (Teddlie and Tashakkori, 2003). The advantage of this design lies in the possibility of combining sequential design (exploratory and explanatory) and also embedded models, as addressed previously (Tashakkori and Teddlie, 2010). Therefore, it is very common for multi-year projects (e.g. PhD research) and large funded projects. Due to the fact that the undertaken research comprises complementary objectives and different types of research questions, it is more practical to divide the whole practical implementation process into different phases. Therefore, the multi-level model is seen as the most adequate triangulation design model for this study.

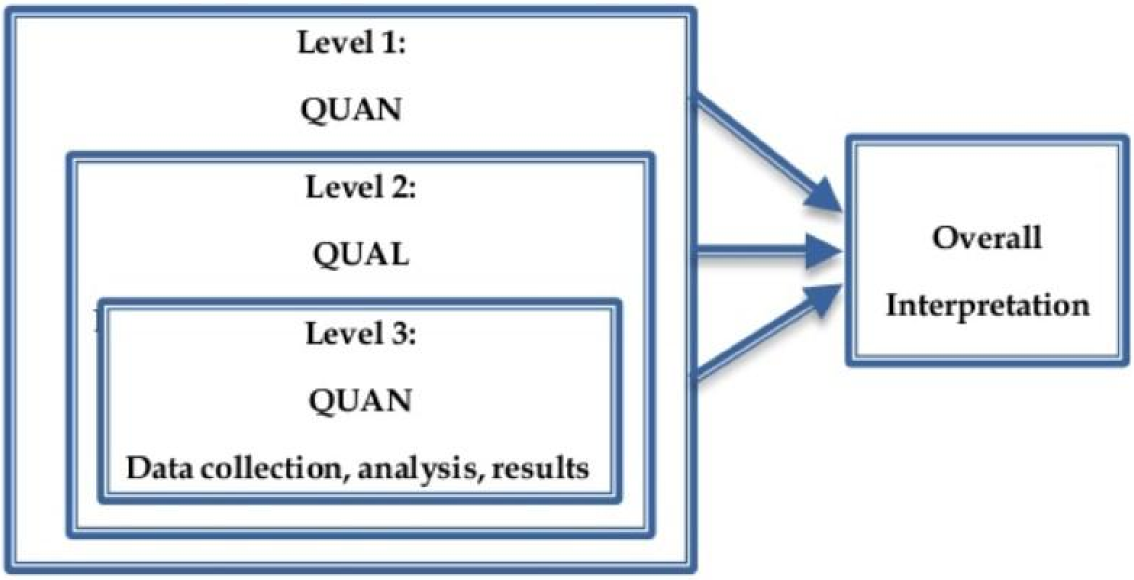



Figure A. 22.The multilevel model (Tashakkori and Teddlie, 2010)

# Part III

## Chapters 5 and 8 appendix

### A.3.1 . Data preparation and screening

Based on analysing key publications on panel data preparation and analysis including (Longhi and Nandi, 2014); Arellano, 2003), the below data preparation procedures have been performed;

#### A- Data merging, reshaping, and case filtering

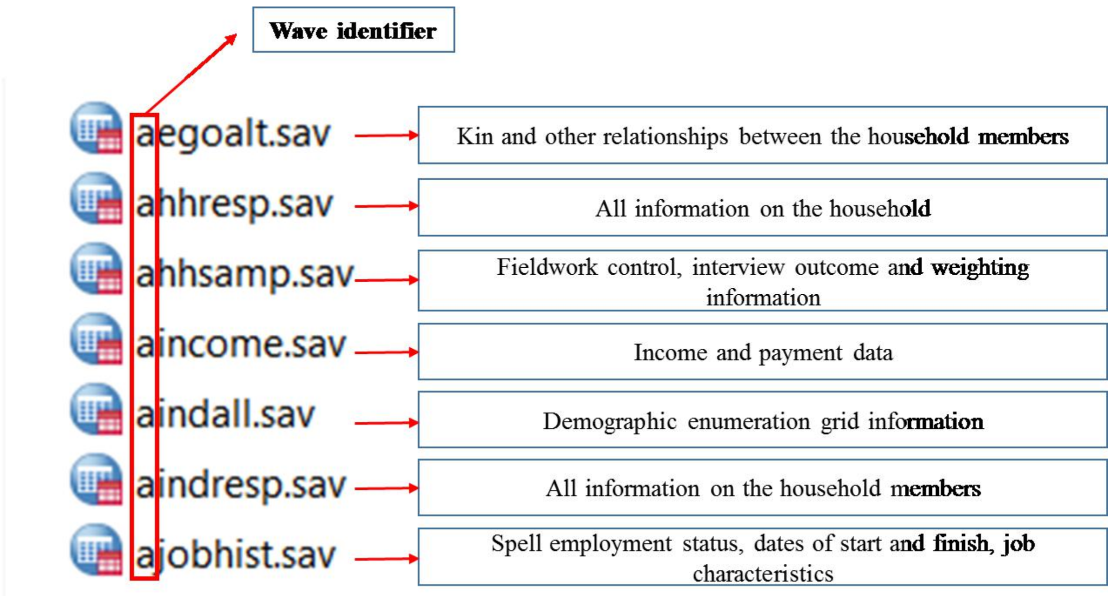



Figure A. 23.common BHPS wave's folder structure

As discussed in chapter 4, the downloaded BHPS dataset entails 18 folders in which each one represents a wave. Figure A. 23 above represents the common files included in each folder in which the first letter of each file represents the wave identifier. For instance, a*egoalt* file encompasses the nature of relationships between the household's individuals in wave1. However, only the *hhresp* files are considered in this research since they contain all socio-economic, demographic, and other information related to the household. On the other hand, *indresp* files were excluded because the analysis of each individual from the household is complex and beyond the scope of the undertaken research.

Based on the household unique identifier and the wave year (e.g.1991), all the *hhresp* files, except wave 6 which lacks energy expenditure variables, were merged into one file using SPSS. After that, with the help of case selection/filtering feature in SPSS, only households who were single non-elderly in 1991 (wave1) were left in the dataset, as shown in Figure A. 24. This resulted in initial number of 599 households making in it a total sample size of 6700 households across 17 years after deleting the ones with all bills included.

In order to permit the dataset analysis, the dataset was reshaped from a wide to a long format. It is possible to arrange panel data using the **restructuring data** option in SPSS. First, in the wide format, each subject is represented by only one row while each response by one column. For example, if the annual income of householder was observed in function of employment mode over 18 years, there will be only one row for this household. However, there will be 18 columns for income (income 1991, income 1992...., income 2008) and the same for the corresponding employment mode (Figure A. 25). On the other hand, in long format, multiple rows are employed to represent each subject while a new column for time variable will be added. This implies that each row depicts a subject at a given time as illustrated in Figure A. 25. Since most of statistical software packages, including Stata, handle panel data analysis only in the long-format, the BHPS data will reshaped from wide to long after being merged.

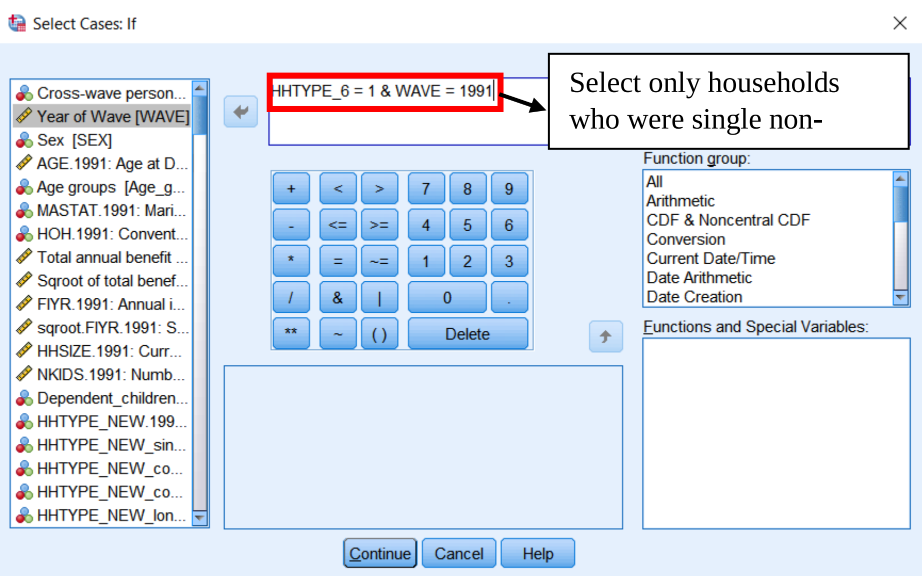



Figure A. 24.case filtering feature in SPSS

| ID | Variable1. 1991 | Variable n.1991 | Variable1. 1992 | Variable n.1992 |
|---|---|---|---|---|
| [redacted] | | | | |
| 2 | | | | |
| [redacted] | | | | |

| ID | WAVE | Variable 1 | Variable n |
|---|---|---|---|
| 1 | 1991 | | |
| 1 | 1992 | | |
| 2 | 1991 | | |
| 2 | 1992 | | |
| 3 | 1991 | | |
| 3 | 1992 | | |

Figure A. 25.The principle of reshaping panel data structure from wide to long format using SPSS

## B-Transforming energy expenditure to quantities

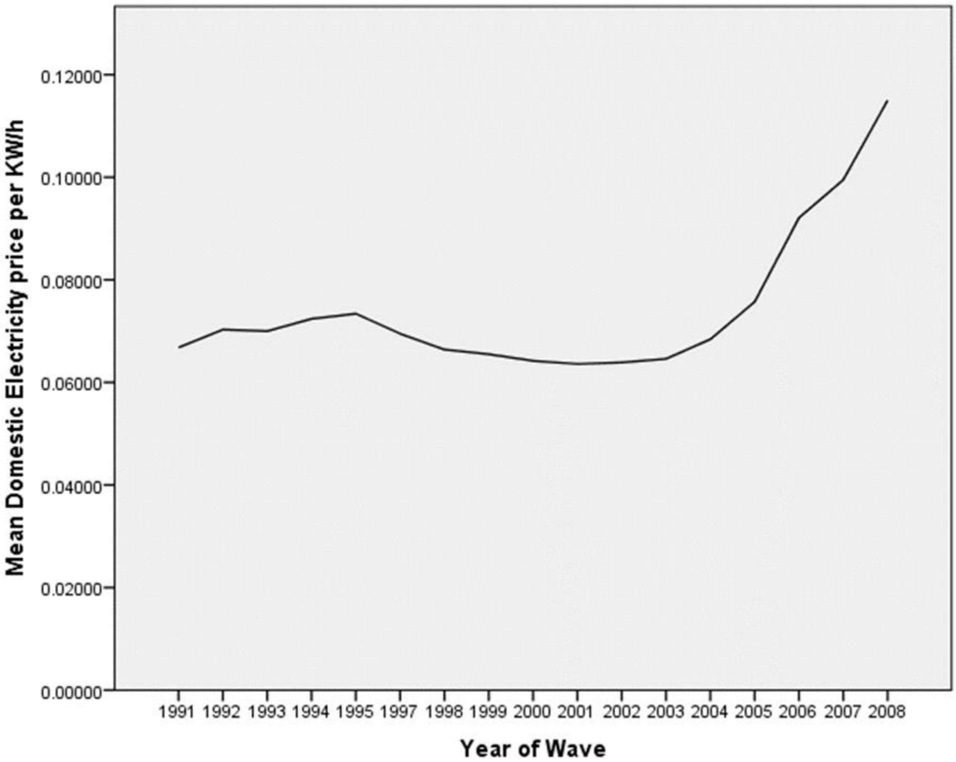



Figure A. 26.The UK average domestic electricity price per unit (KWh) in pounds from 1991 to 2008.

As addressed in chapter 3, fuel price has a significant effect on residential energy consumption patterns known as price elasticities. From this perspective and based on the fact that domestic gas and electricity prices in the UK variated significantly from 1991 to 2008, as shown in Figure A. 26 and Figure A. 27, their impact on the households' energy consumption will be isolated. This is achieved by converting energy expenditure variables to energy quantities through the following steps. First, from the premise that the first five waves encompass monthly energy expenditure and the rest are annual expenses, all the figures across the 17 waves were unified into the annual format. After that, UK official inflation indices for gas and electricity, retail price index RPI more precisely, were employed to derive the KWh unit prices and daily standing charges from 1991 to 2008 (DECC, 2015d). Finally, to get the yearly KWh energy consumption, the annual standing charge for each year was subtracted from the annual energy expenditure and the result was divided by the kwh price of the corresponding year.

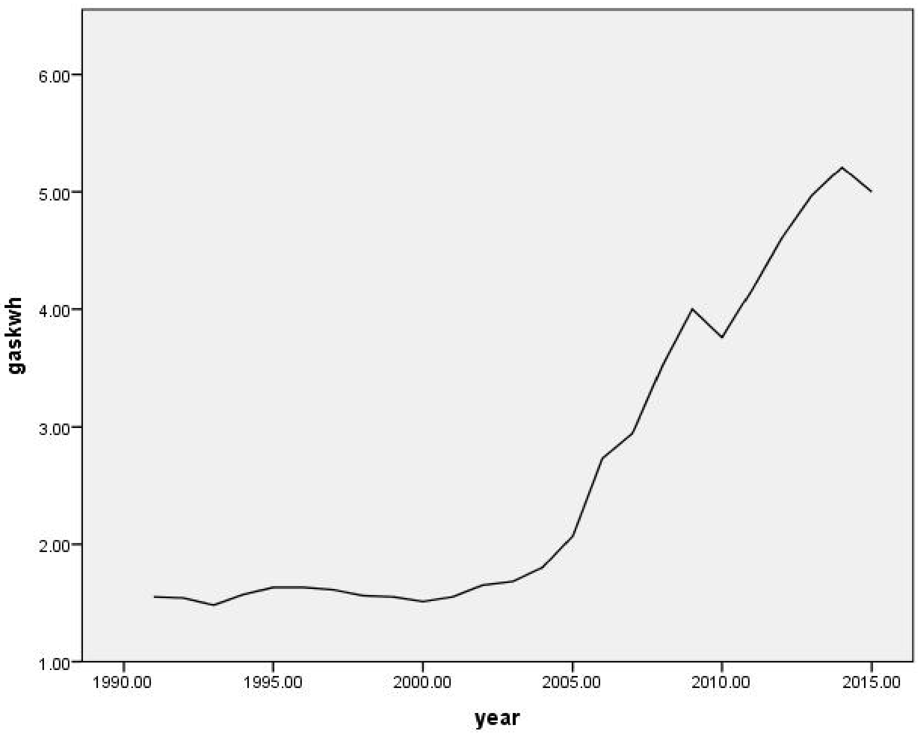



Figure A. 27.The UK average domestic gas prices per unit (KWh) in pence from 1991 to 2015

### C-Exclusion of households with bills included

Since one objective of this research is to investigate the variation in residential energy consumption, around 700 households with utility bills included in the rent could have a negative impact on the interpretation of the results. Therefore, we have decided to exclude them from the analysis.

Apart from the above issue, another problem, which consisted of non-logical annual energy expenditure figures such as 20 pounds per year, has been encountered. Based on the official report generated by BRE (2013) which suggested that the minimum annual electricity energy consumption figure in their sample of 1345 households was 400 KWh, any annual consumption value below 400 was removed and predicted using multiple imputation techniques. As for gas energy consumption, it was noticed that the majority of non-logical values were given by households who do not possess gas central heating but use gas instead for cooking purposes. Therefore, based on the UK energy demand breakdown chart (Figure A. 28) published by DECC (2012) which suggested that cooking represents 4% of the total energy usage, any value less than 4% (550 KWh) of the average household gas energy consumption (13800 KWh) will be not considered.

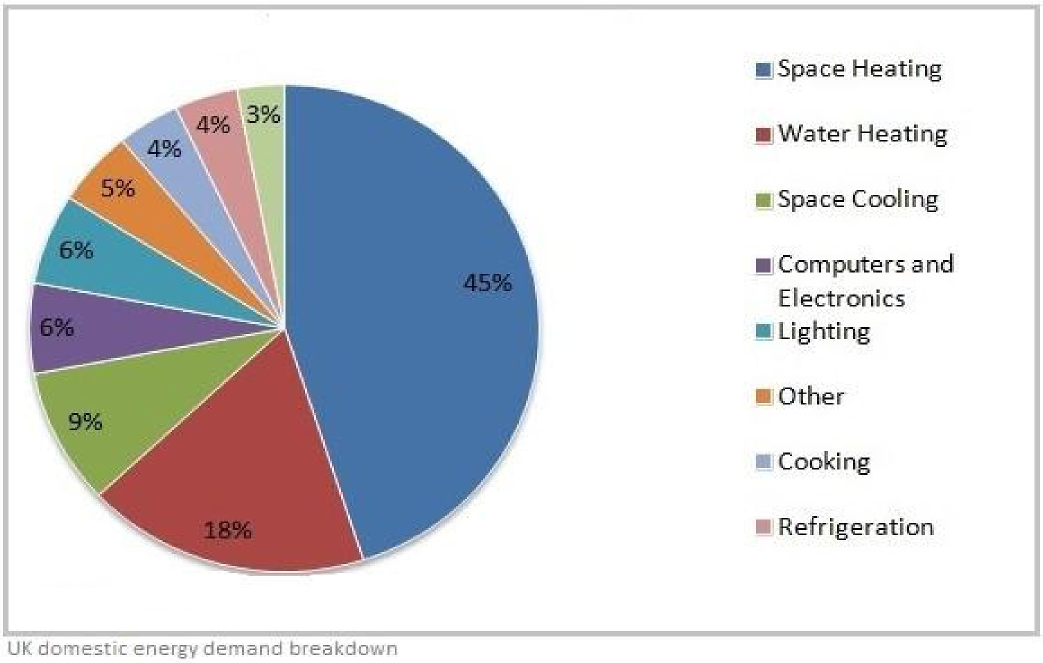



Figure A. 28.UK domestic energy demand breakdown (DECC, 2012b).

## D-Recoding variables

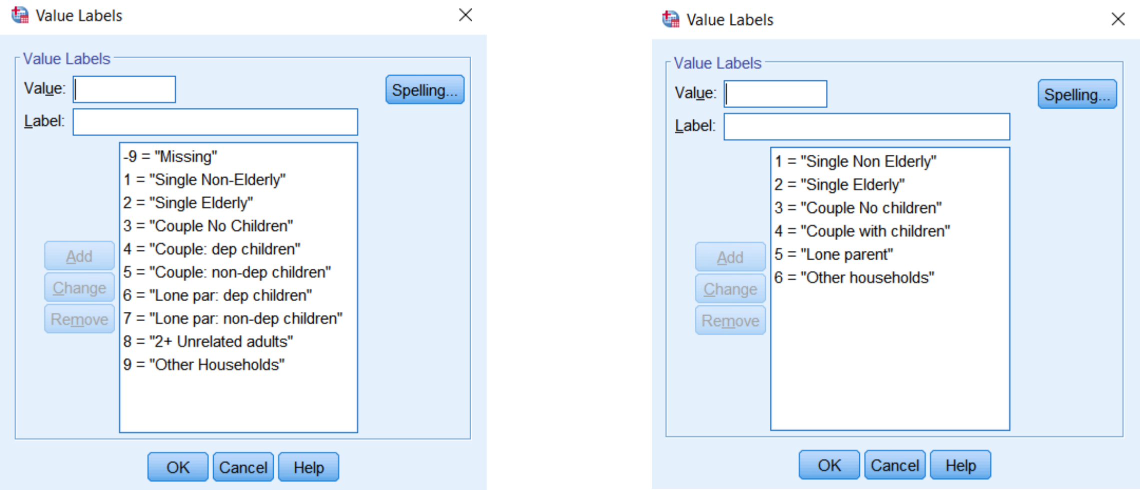



Figure A. 29.the categorical variable household type before and after being recoded

Table A. 5.The frequency table of the variable household type in the BHPS dataset

**Household Type**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | Single Non-Elderly | 3767 | 52.5 | 52.5 | 52.5 |
| | Single Elderly | 1136 | 15.8 | 15.8 | 68.3 |
| | Couple No Children | 1021 | 14.2 | 14.2 | 82.5 |
| | Couple: dep children | 820 | 11.4 | 11.4 | 93.9 |
| | Couple: non-dep children | 114 | 1.6 | 1.6 | 95.5 |
| | Lone par: dep children | 87 | 1.2 | 1.2 | 96.7 |
| | Lone par: non-dep children | 111 | 1.5 | 1.5 | 98.3 |
| | 2+ Unrelated adults | 85 | 1.2 | 1.2 | 99.5 |
| | Other Households | 38 | .5 | .5 | 100.0 |
| | Total | 7179 | 100.0 | 100.0 | |

Owing to the fact that categorical variables with many groups can inflate the dimension of the analysed dataset (Shmueli et al., 2016), the number of categories in the household type variable has been collapsed from 9 to 6 categories, as shown in Figure A. 29. This was achieved based on the frequency table of this variable (Table A. 5). In other words, groups with a very low frequency such as +2 unrelated adults (1.2%) and other households (0.5%), were combined in one category (other).

Conversely, children dependency was isolated as a standalone variable while combining couple dep children and non-dependent children in one category. The same applies to lone parents with and without dependent children. Following the recording of the dependent variable, it was transformed into a dichotomous variable using SPSS to enable further analysis. Similarly, age and income variables were recoded into categorical variables using SPSS to explore their interaction with other categorical variables using cross-tabulation.

### E-Transformation and cleaning of expenditure variables

As expected, the inspection of expenditure variables histograms suggested a non-normal distribution. For example, as illustrated in Figure A. 30, it is clear that the original household total annual income is negatively (right) skewed with a long left tail. This can be problematic since a large number of parametric statistical tests such as T-test, ANOVA, and regressions, assume the normality of variables. Therefore, not improving the normality of such variables implies violating this assumption and leads to misleading or erroneous results. However, since there are plenty of existing data transformation techniques, choosing the right transformation method is crucial to successful data analysis. McDonald (2009) advised on the use of the most frequent transformation in the studied field. Osborne (2010) suggests that square root, log, and inverse transformations are the most frequent in most fields including social sciences. Furthermore, argued their use should be gradually depending on the skewness and keratosis of the variable. For example, for variables with minor skewness, a square root transformation is more appropriate, whereas logarithmic or inverse transformations are adequate for the ones with extreme skewness. For example, log10 transformation has been applied to normalise the annual household total annual income as depicted in Figure A. 31 and Figure A. 32. Conversely, square root transformation was applied to normalise the household annual gas expenditure variable (Figure A. 33, Figure A. 34, and Figure A. 35).

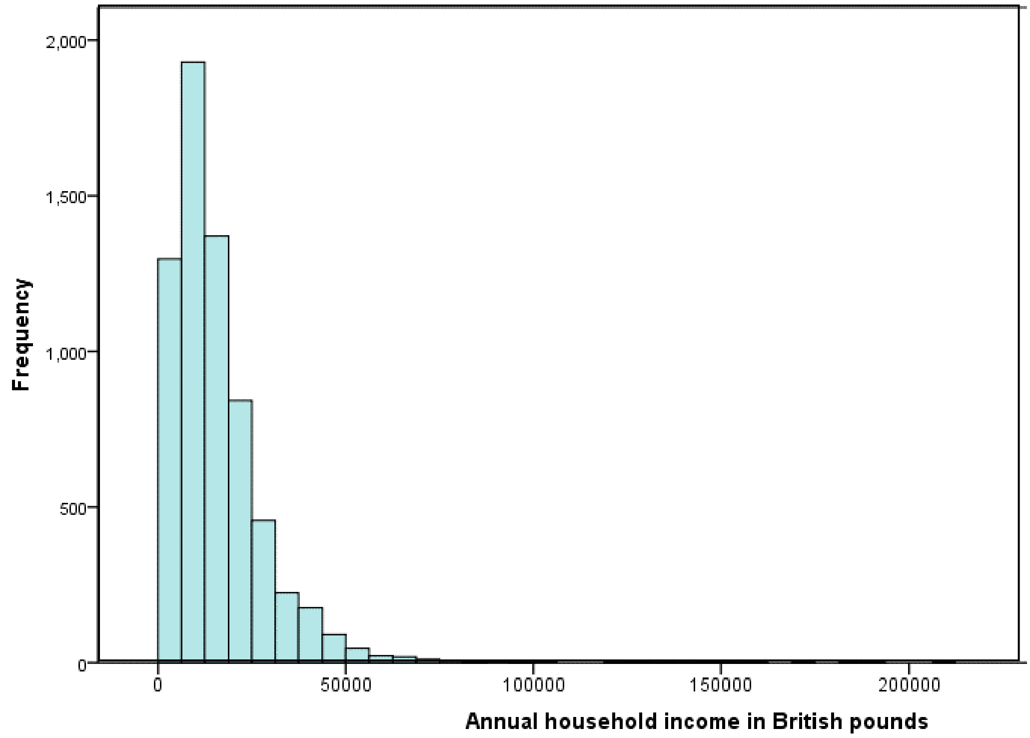



Figure A. 30.the histogram of annual household income before normalisation

Once the variables have been transformed, outliers will be first inspected for with the help of Box-and-whisker plots. The box-and-whisker plot employs the interquartile range, which is the difference between the upper and lower quartiles, as a dispersion measure. This, in turn, helps flag potential outliers in the dataset. However, visual inspection on its own is not enough. For those reasons, Hoaglin et al.(1986) labelling rule will be employed to accurately and objectively delete outliers. If any variable value exceeds the upper quartile plus the interquartile range times 1.5, it is an outlier as illustrated in (Equation A. 1). Similarly, if a value is less than the lower quartile minus the interquartile range times 1.5, it is an outlier ((Equation A. 2)).

$$F_U + 1.5(F_U - F_L) \qquad \text{(Equation A. 1)}$$

$$F_L - 1.5(F_U - F_L) \qquad \text{(Equation A. 2)}$$

An example of cleaning outliers is presented in Figure A. 36 and Figure A. 37 which show the box and whisker plots of log 10 household annual electricity consumption variable before and after deleting outliers, respectively.

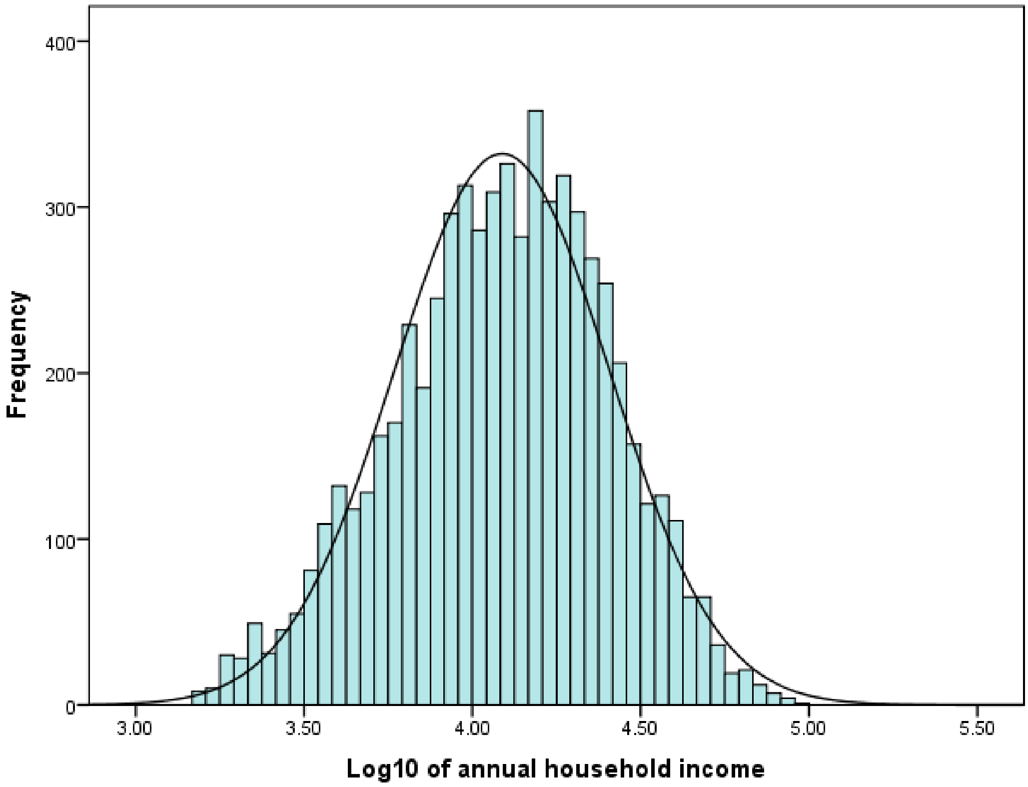



Figure A. 31.The histogram of annual household income after normalisation using log 10

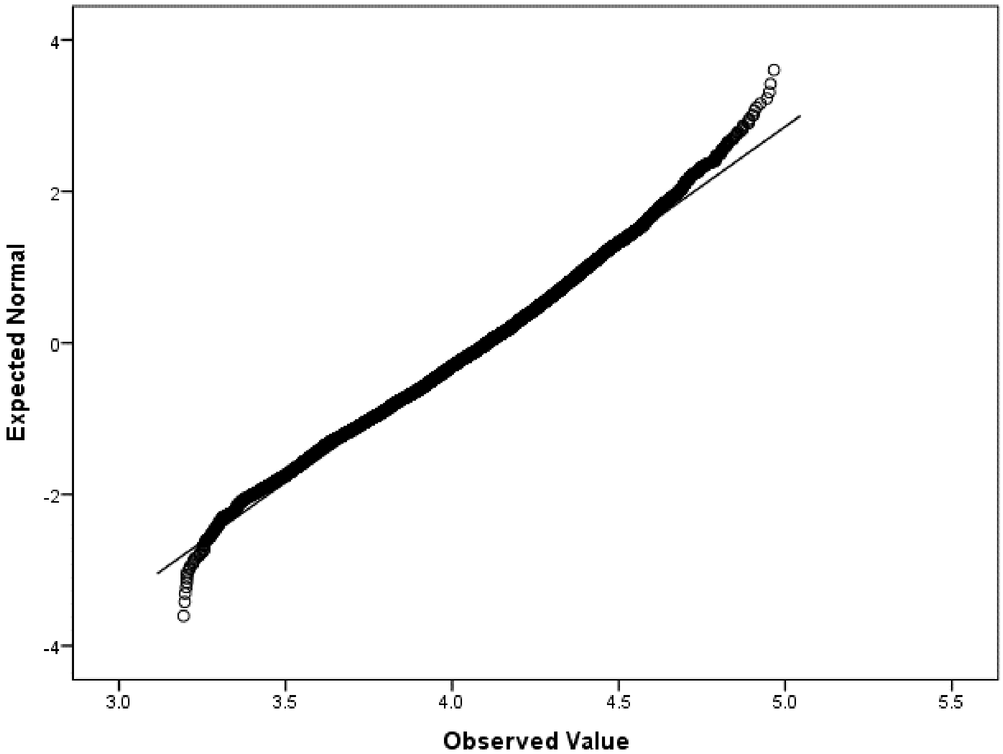



Figure A. 32.The Q-Q plot of the variable log 10 annual household total income

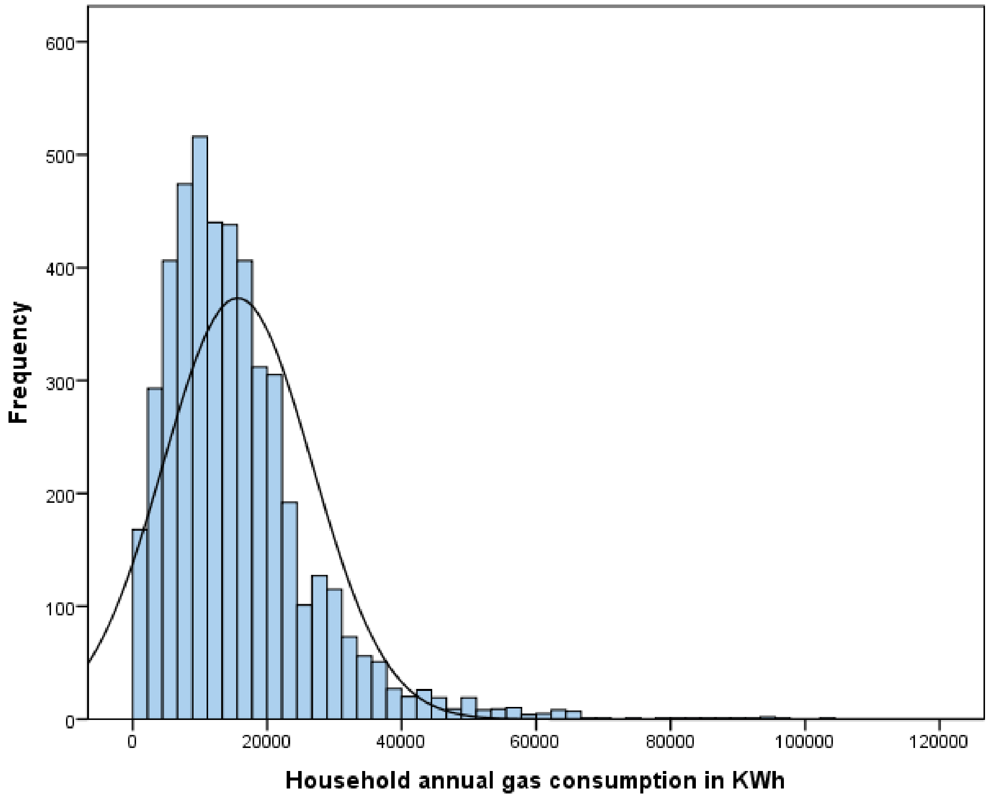



Figure A. 33.The variable household annual gas consumption before normalisation

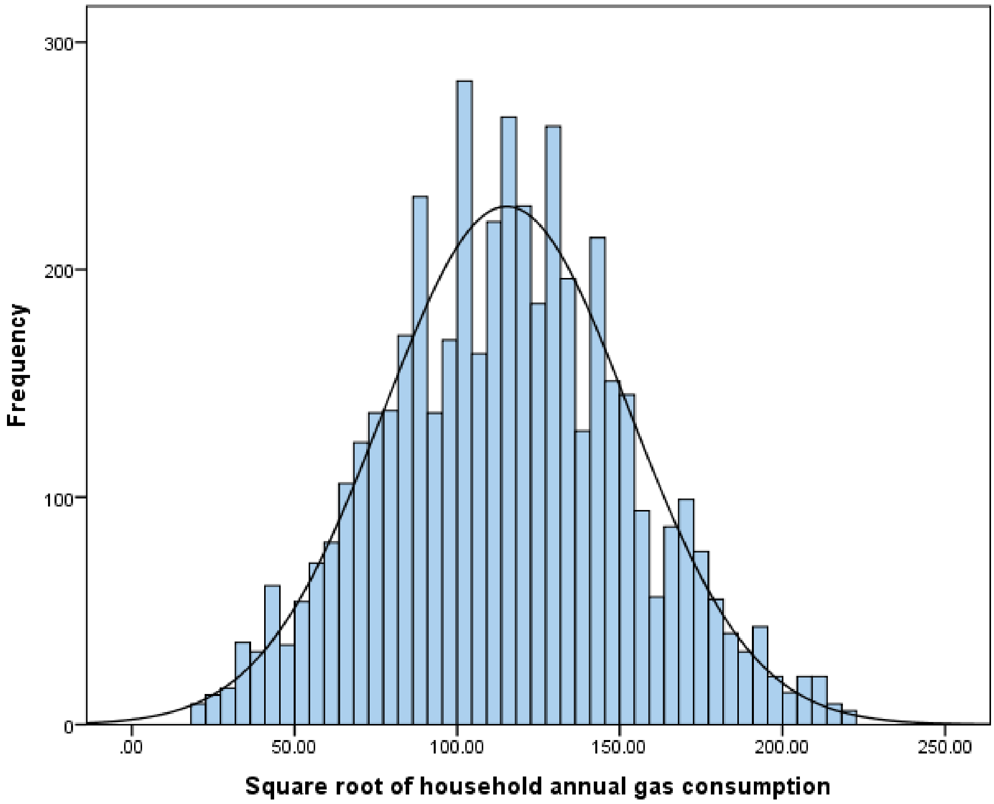



Figure A. 34.The variable household annual gas consumption before normalisation using square root

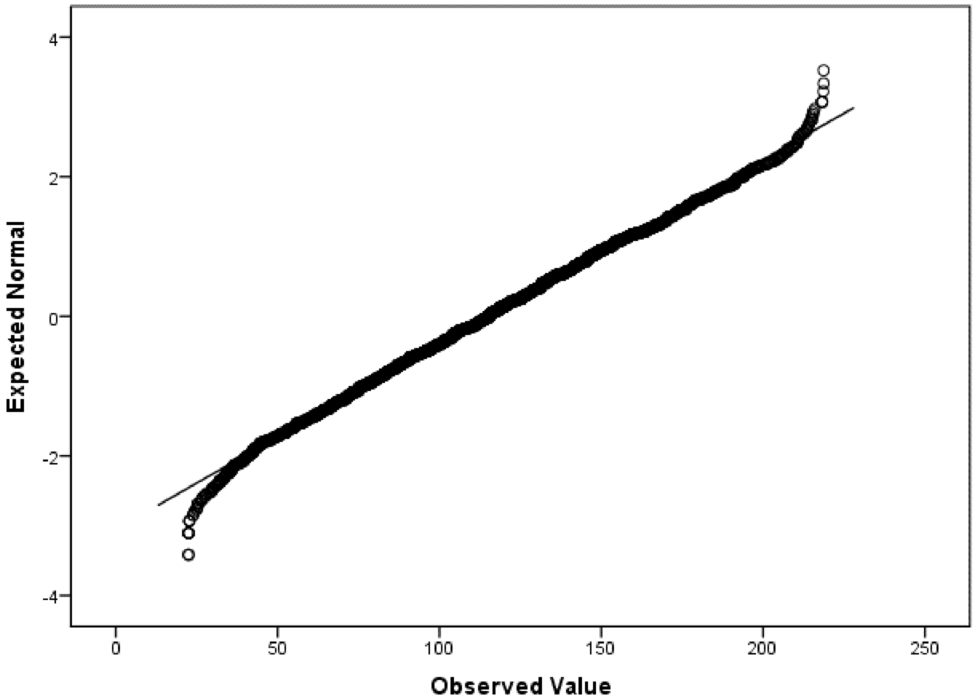



Figure A. 35.The Q-Q plot of the variable square root of annual gas consumption

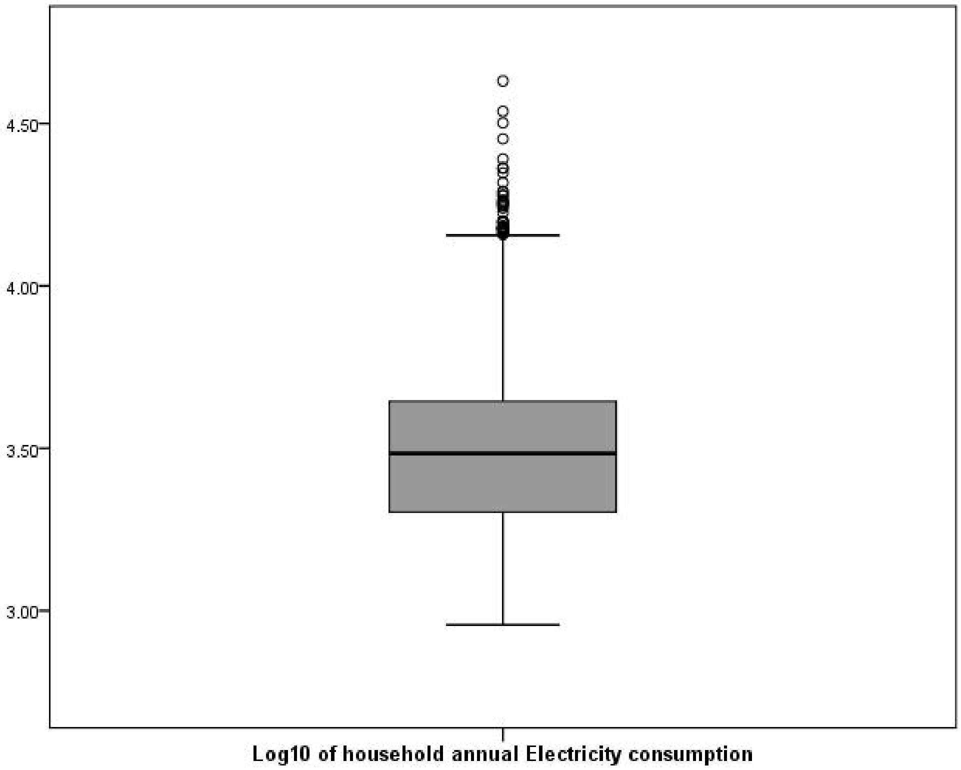



Figure A. 36.Whisker and box plot of the variable log 10 household annual electricity consumption before cleaning outliers

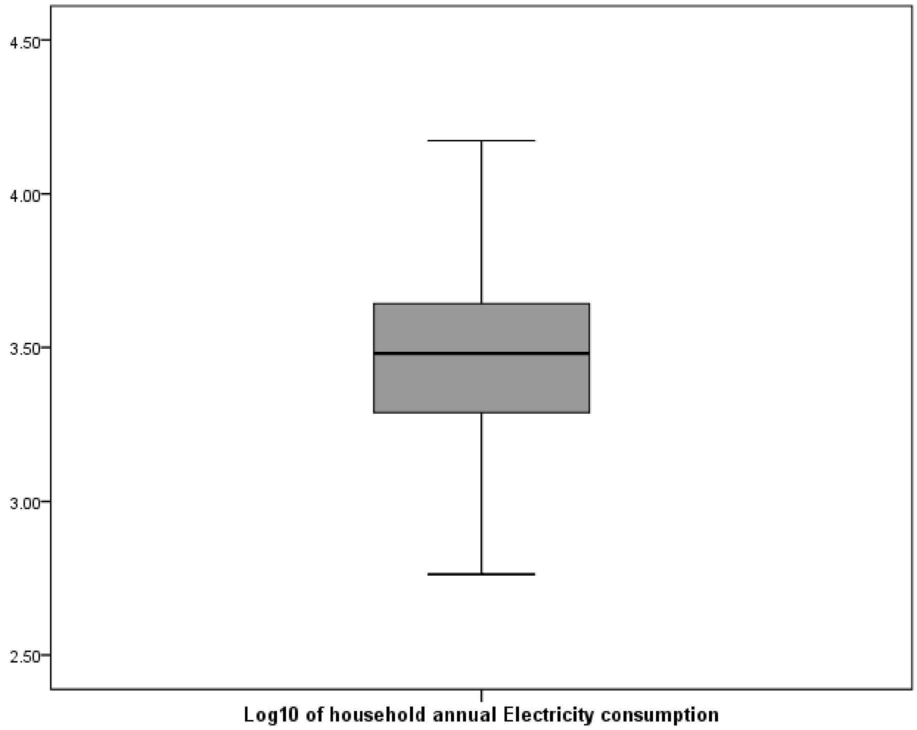



Figure A. 37.Whisker and box plot of the variable log 10 household annual electricity consumption before cleaning outliers

## F-The handling of missing data


Figure A. 38.The Overall summary of missing data in the BHPS database.

Figure A. 38 summarises the percentage of missing data in the BHPS database. Overall, it is evident that 61.54% of the variables and 17% of the cases have missing values. However, since the percentage of missing data in the total values is 2.77 %, which is less than 10%, this proportion of missing data can be tolerated according to Bennett (2001). Nevertheless, the missing values were predicted using multiple imputations method initiated by Rubin (2004). Furthermore, results with imputed and non-imputed will be compared.

### A.3.2 Characteristics of single non-elderly households in the BHPS dataset in 1991

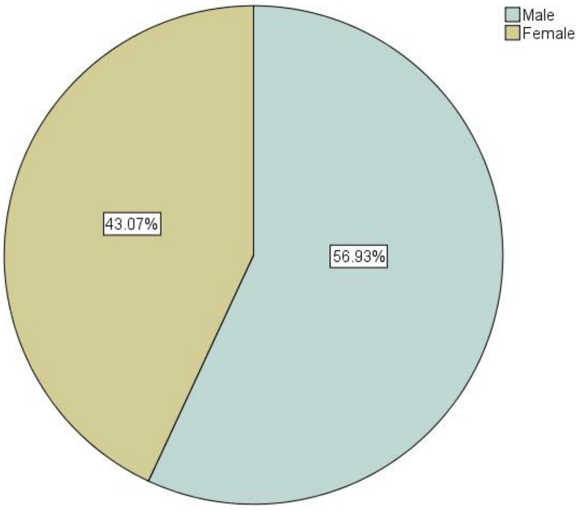



Figure A. 39.The proportion of males and females in the BHPS dataset

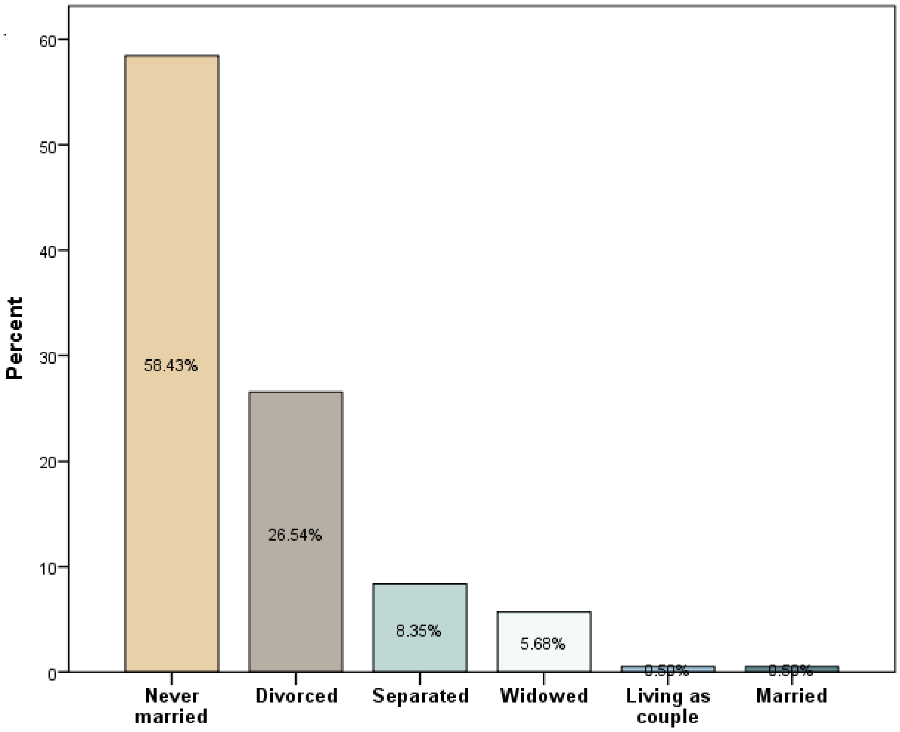



Figure A. 40.The marital status distribution of the participants

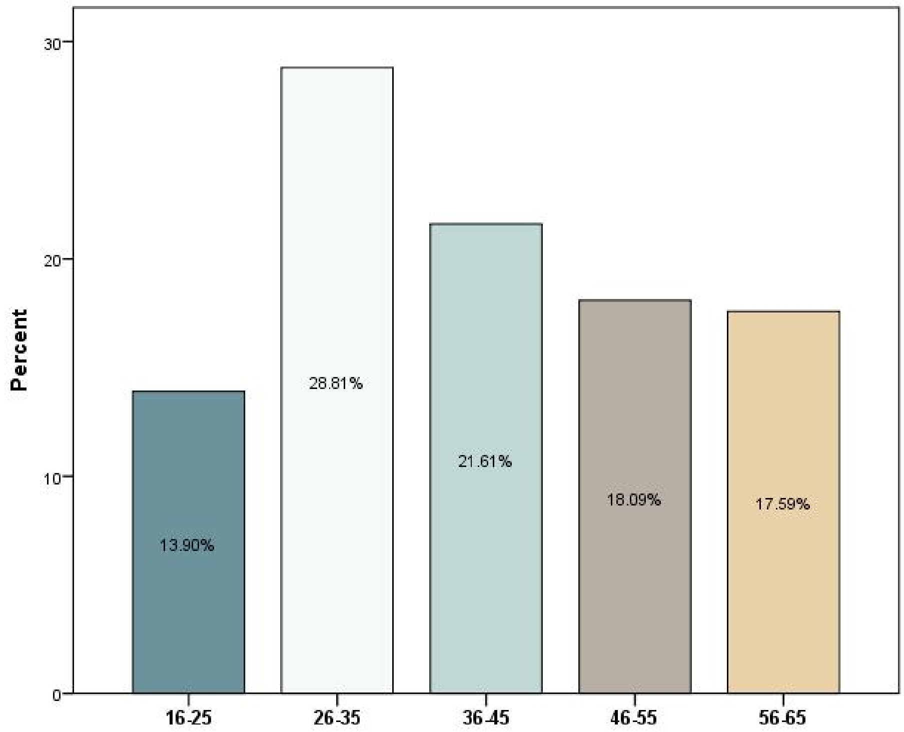



Figure A. 41.Age distribution of the BHPS respondents

Before addressing the different factors that contributed to the transitions of single non-elderly households from 1991 to 2009, it is necessary to gain insight into their initial characteristics in the year 1991. This will help to enhance the interpretation of study results. Moreover, enable the study of single non-elderly occurring transition patterns over this period in a meaningful way. Based on that, the below sub-section aims to analyse the socio-economic and demographic characteristics of this household type over this period.

### A-Gender, age, and marital status

Figure A. 39 and Figure A. 41 represent the gender and age distribution of single non-elderly households in wave 1 (year 1991), respectively. On the other hand, Figure A. 40 depicts their marital status. Overall, it is clear that the proportion of males (56.93%) was slightly higher than the one of females (43.07%). Furthermore, it is evident that approximately 58% of them were never married and mostly aged between 26 and 45. On the other hand, the minority were married or cohabiting couples (roughly 1%) and aged between 16-25 (13.90%).

As shown in Figure A. 40, the high proportion of never-married households (58.43%), of which 62% are males and 37.8% females, suggests possible future transitions to other family types including couples without children. Indeed, a recent government report indicated that over 67% of marriages in the UK and Wales were first marriages for both partners at an average age of 35.5 years (ONS, 2013e). Moreover, reported a 2.4% increase in the proportion of cohabiting couples from never-married single non-elderly over the last 10 years (ONS, 2015e).

Similarly, around 34% of households who were divorced (26.54%), separated (8.35%), and widowed (5.68%), are more likely to become lone parents according to a recent official report on lone parents with dependent children in the UK (ONS, 2015e). However, their transition to a couple with or without children is also possible. This is owing to the fact that remarriage accounts for around 15% of all marriages in the UK (ONS, 2013e).

In the light of the above analysis, it is clear that a considerable proportion of the studied single non-elderly household is expected to make transitions to other household types over this period.

**B-Income and electricity consumption**

Figure A. 42 represents the income distribution of single non-elderly households in year 1991, whereas Figure A. 43 shows their annual electricity consumption in function of different income bands. Generally speaking, it is clear that approximately 88% of the household earned between 0-£20k yearly, whereas the rest are paid over £20k. As for their annual electricity, it is evident that the majority of them consumed between 1000-4000 KWh, despite the nearly equal distribution of all energy consumption figures across all income bands.

For example, for households earning less than £10000 a year, 22.8% consumed 1000-2000KWh, 26.6% used between 2000-3000KWh, and 18.9% had 3000-4000KWh (Table A. 6). Moreover, a considerable proportion of around 28% consumed more than 4000KWh in comparison to only 3.5% for less than 1000 KWh.

Similarly, households with a yearly income ranging between £10000-20000 had similar figures to the ones earning less than £10000. For instance, 29.3% consumed 1000-2000KWh, whereas, 24.7% used between 2000-3000KWh. 19.5% consumed between 3000 and 4000KWh. On the other hand, the proportion of households using more than 4000KWh yearly was lower by roughly 8%.

In conclusion, the income distribution of single non-elderly households in addition to the roughly equal distribution of electricity consumption groups across their income bands, suggests that income has no or little effect on electricity usage patterns.

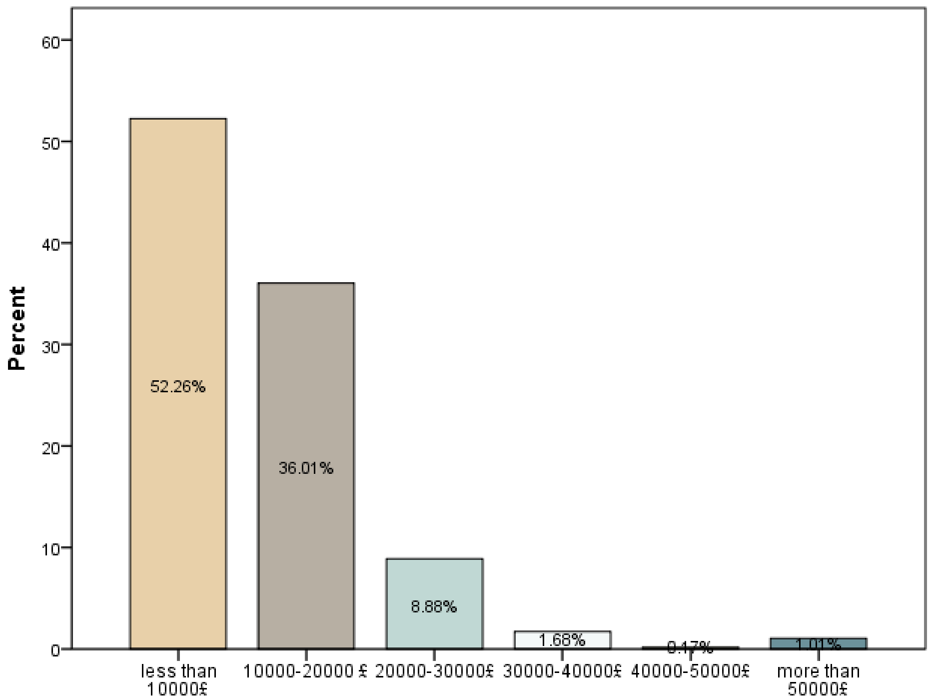



Figure A. 42.household income distribution in year 1991

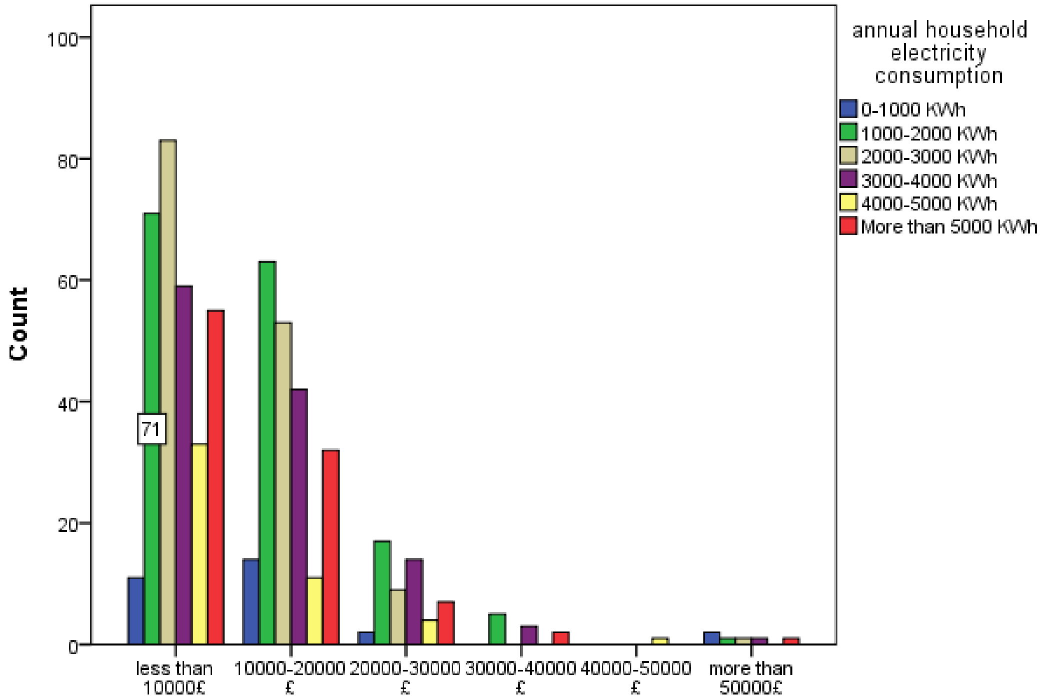



Figure A. 43.Households annual electricity consumption by income band

Table A. 6.households' yearly electricity consumption in function of their income bands (year 1991)

| | annual household electricity consumption | | | | | |
|---|---|---|---|---|---|---|
| | Less than 1000 KWh | 1000-2000 KWh | 2000-3000 KWh | 3000-4000 KWh | 4000-5000 KWh | More than 5000 KWh |
| less than 10000£ | 3.5% | 22.8% | 26.6% | 18.9% | 10.6% | 17.6% |
| 10000-20000 £ | 6.5% | 29.3% | 24.7% | 19.5% | 5.1% | 14.9% |
| 20000-30000£ | 3.8% | 32.1% | 17.0% | 26.4% | 7.5% | 13.2% |
| 30000-40000£ | 0.0% | 50.0% | 0.0% | 30.0% | 0.0% | 20.0% |
| 40000-50000£ | 0.0% | 0.0% | 0.0% | 0.0% | 100.0% | 0.0% |
| more than 50000£ | 33.3% | 16.7% | 16.7% | 16.7% | 0.0% | 16.7% |

## C-Income and gas consumption

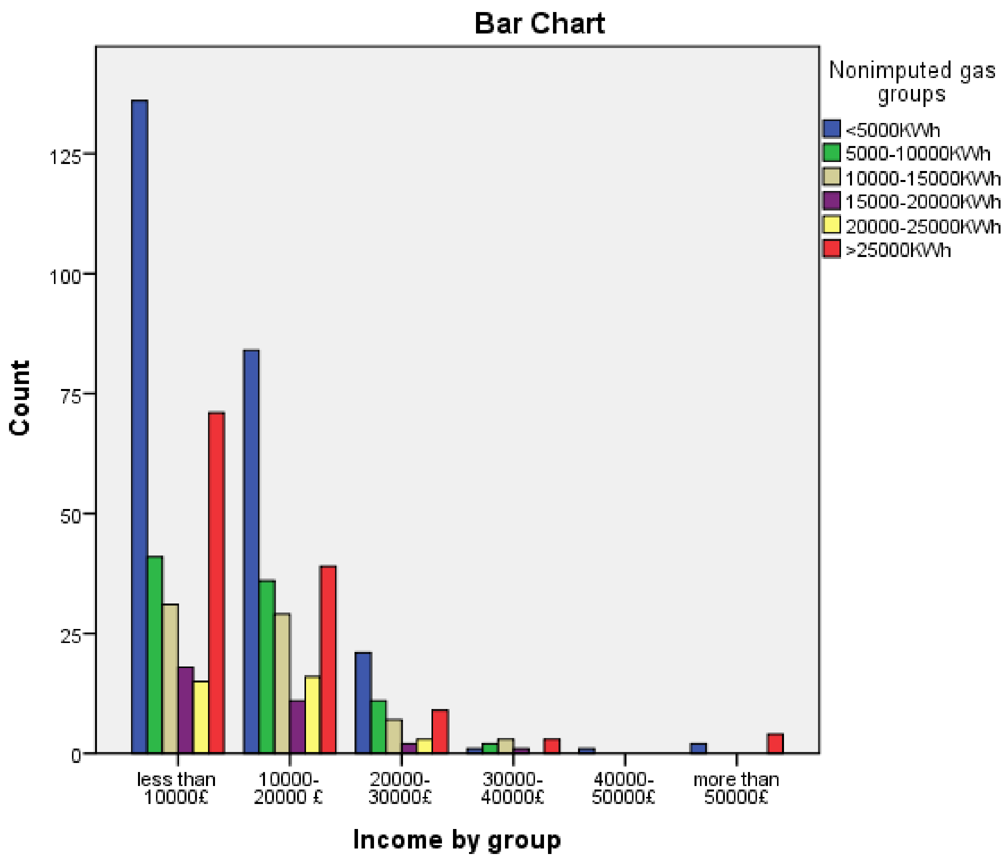



Figure A. 44..Households annual gas consumption by income band

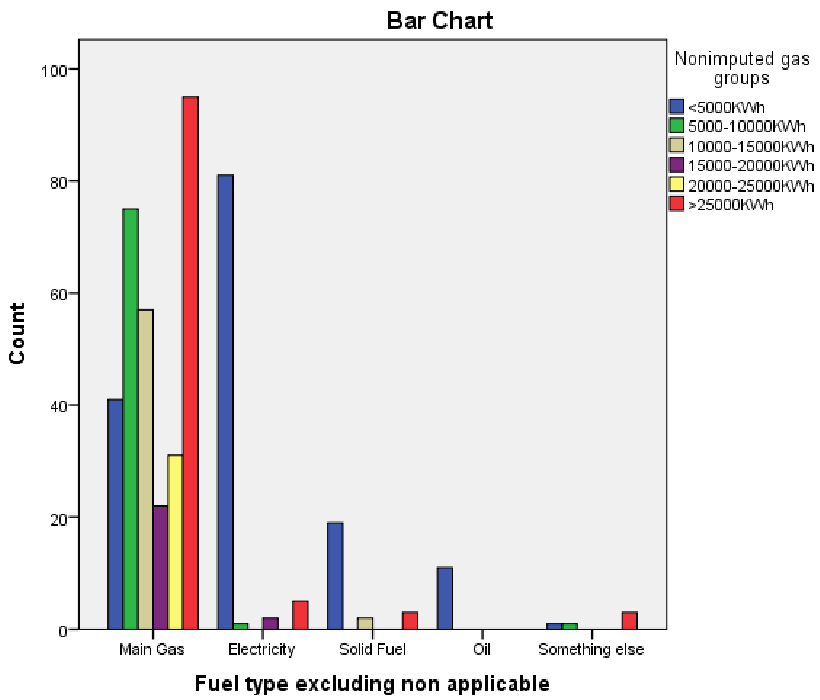



Figure A. 45.The household annual gas consumption by the type of main heating fuel

Figure A. 44 illustrates a bar graph representing single non-elderly annual gas consumption by income band. Overall, it is evident that the majority of the households, excluding the ones with all bills inclusive, consumed less than 5000KWh annually. This was followed by the subsequent annual consumption figure's; more than 25000KWh, 5000-10000KWh, 10000-15000KWh, 20000-25000KWh, and 15000-20000KWh, correspondingly.

The high percentage of households using less than 5000 KWh could be attributed to the employed central heating fuel type. Indeed, from analysing the Figure A. 45 below, which depicts the distribution of gas consumption bands by central heating fuel type, it is clear that the majority of those consumers employed electricity as their main heating fuel. Conversely, the majority of single non-elderly who consumed more than 25000 KWh per year had gas central heating.

At first glance, it could be argued that income should have an impact on these households' gas consumption patterns given their low-income figures. However, considering that the households who consumed more than 25000 KWh annually were mostly earning less than £10k, income is not expected to have a major impact.

**D-Socio-economic status and level of education**

Figure A. 46 depicts the level of education of the studied single non-elderly households in wave 1. On the other hand, Table A. 7 outlines their education levels in function of their Goldthorpe socio-economic profile. In general, the proportion of households with no formal or other qualifications formed the majority with 33.17%, whereas the ones with a higher education degree represented the minority (2.18%). As for their socio-economic status, the proportion of unemployed households was the highest by 27.68% followed by service class lower-grade professionals (21.64%), higher-grade professionals (15.6%), and semi/unskilled manual workers (7.72%). Conversely, agriculture workers, farmers/smallholders, small proprietors with employees, and personal service workers, were the least frequent socio-economic profiles with 1.01%, 0.5%, 1.17%, and 2.35%, respectively.

As illustrated in Table A. 7, 51.5% of those with other or no formal qualifications were unemployed, whereas 17.2% were semi or unskilled manual workers. The remaining 31.3% were almost equally divided across different socio-economic status such as foreman technicians.

As for households with a higher education degree, 92.3% of them belonged to service class higher and lower grade. 80% of households with a first degree and 73% of those possessing HND/HNC/teaching qualification.

For households with A-level and O-level qualifications, 22.4% and 15.9% were on unemployment. Moreover, 48.3% of households with A-level were service class lower and higher grade professionals in comparison to 36.7% for the ones with O-level qualifications.

On the other hand, households with a certificate of secondary education were more gravitated towards manual work since 20.8% and 16.7% of them belonged to semi/unskilled and skilled manual worker classes, correspondingly.

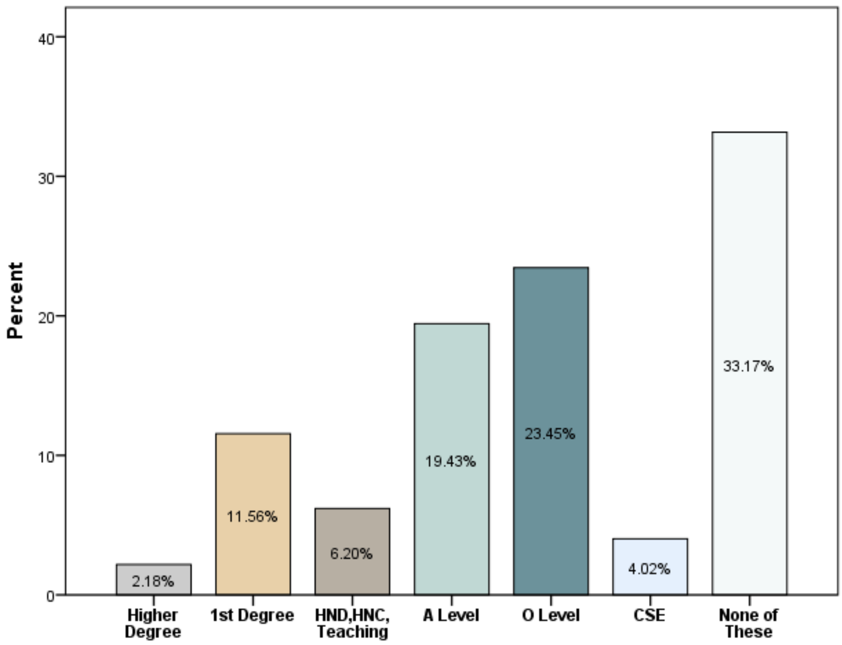



Figure A. 46. The distribution of education level among the BHPS single non-elderly households

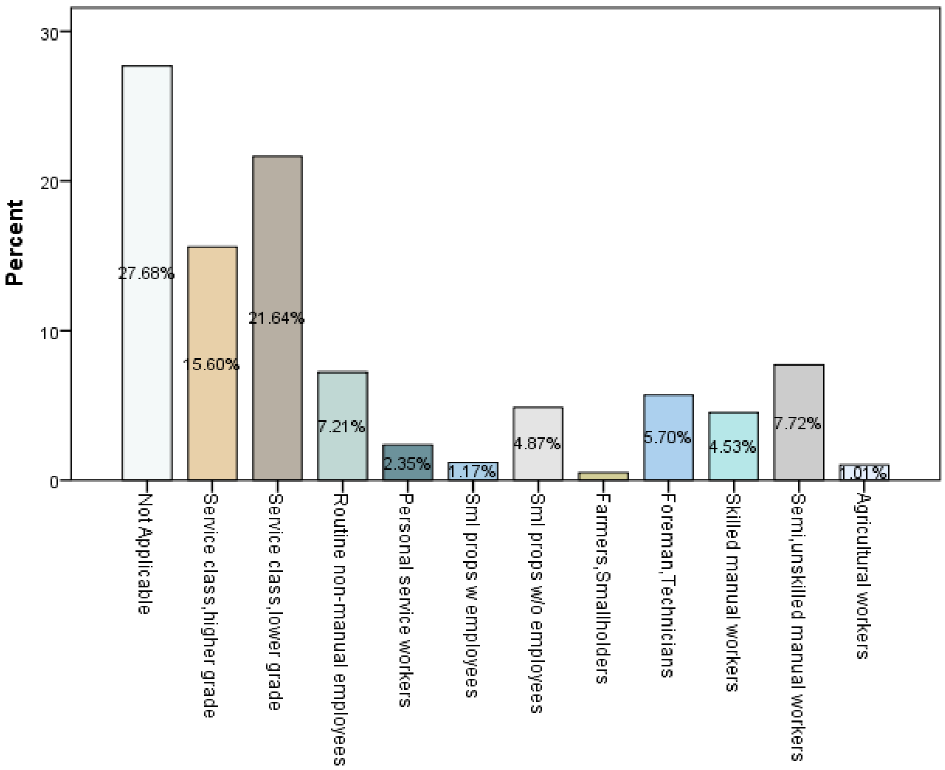



Figure A. 47. The distribution of single non-elderly BHPS households by socio-economic class

Table A. 7.outlines the households' education levels in function of their Goldthorpe socio-economic profile

| | Goldthorpe Social Class: present job | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Not Applicable | Service class, higher grade | Service class, lower grade | Routine non-manual employees | Personal service workers | Sml props w employees | Sml props w/o employees | Farmers, Smallholders | Foreman, Technicians | Skilled manual workers | Semi, unskilled manual workers | Agricultural workers |
| Higher Degree | 7.7% | 69.2 % | 23.1 % | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| 1st Degree | 8.7% | 52.2 % | 33.3 % | 2.9% | 0.0% | 1.4% | 1.4% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| HND, HNC, Teaching | 10.8% | 18.9 % | 54.1 % | 0.0% | 2.7% | 0.0% | 2.7% | 0.0% | 5.4% | 2.7% | 2.7% | 0.0% |
| A Level | 22.4% | 17.2 % | 31.0 % | 6.0% | 3.4% | 0.9% | 5.2% | 0.0% | 6.9% | 5.2% | 0.9% | 0.9% |
| O Level | 15.8% | 11.5 % | 25.2 % | 20.9% | 2.9% | 1.4% | 5.0% | 0.7% | 7.9% | 4.3% | 3.6% | 0.7% |
| CSE | 16.7% | 0.0% | 16.7 % | 4.2% | 4.2% | 0.0% | 16.7% | 0.0% | 0.0% | 16.7 % | 20.8 % | 4.2% |
| None of These | 51.5% | 2.5% | 4.0% | 2.0% | 2.0% | 1.5% | 5.1% | 1.0% | 6.6% | 5.1% | 17.2 % | 1.5% |

## E-Dwelling and tenure characteristics

*Dwelling type and size*

Figure A. 47 and Figure A. 48 provide an overview on the distribution of the studied households over different dwelling types and sizes, respectively. Generally speaking, it is clear that approximately 40% of single non-elderly households dwelled in flats. On the other hand, only 7.22% and 6.19% lived in end-terraced houses and other accommodation types, correspondingly. As for the dwelling size, the majority (27.81%) stayed in 3 bedroom dwellings. This is followed by households living in 2 bedrooms (24.79%), 4 bedrooms (21.94%), 5 bedrooms (11.89%), and finally 1 bedroom (7.54%).

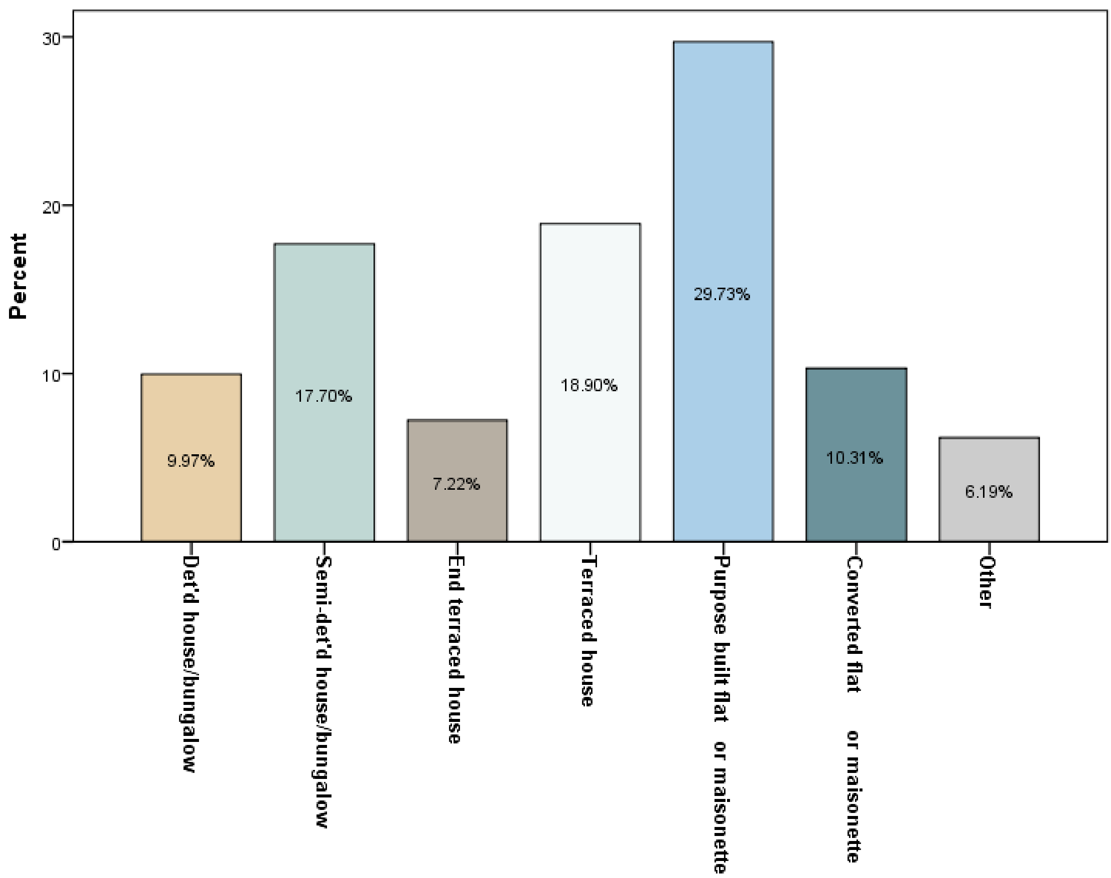



Figure A. 48. The distribution of BHPS households dwelling types

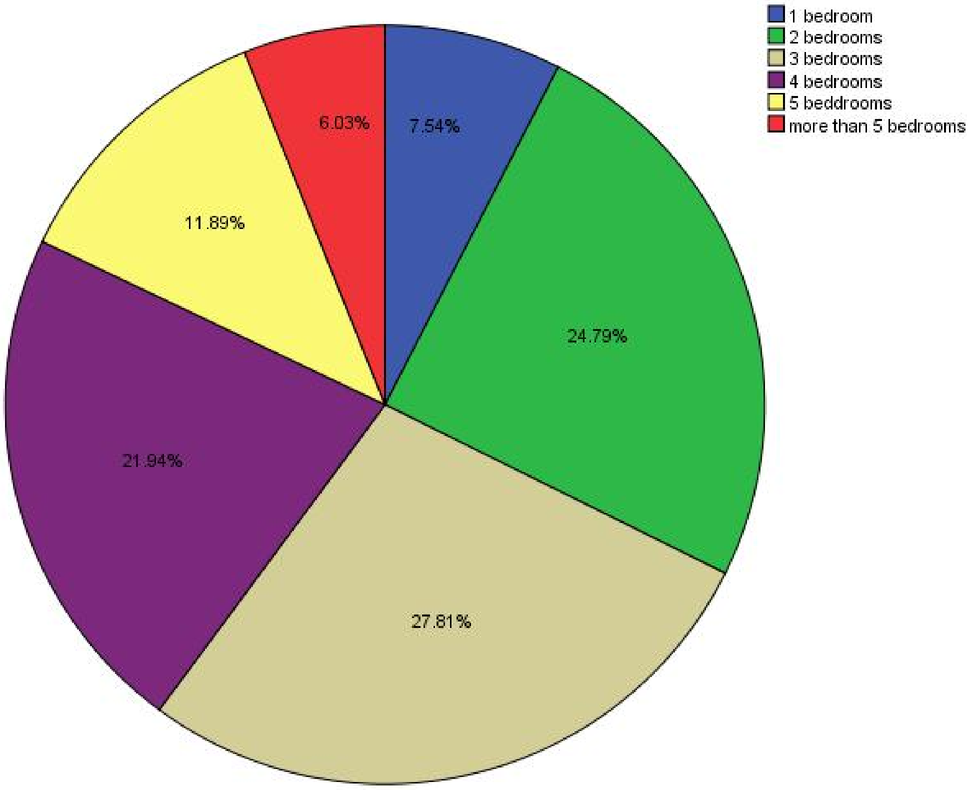



Figure A. 49.The dwelling size of the studied BHPS households

*Tenure mode*

Figure x is a pie chart of housing tenure pertaining to the observed single non-elderly households. Overall, it is evident that the percentage of households who own their accommodation with mortgage was the highest by 43.29%. Conversely, the proportion of those who rented from employers or on other tenure modes represented the minority by 2.45% and 0.34%, correspondingly.

As illustrated in table 1, 75% of households who owned their properties with mortgage preferred to dwell in terraced houses, purpose built flats, and semi-detached houses. However, converted flats and end-terraced houses were the least favoured as their percentages were 7.8% and 8.6%, respectively. Similarly, households who owned their properties outright had similar preferences except that the percentage of living in detached dwellings was higher by roughly 16%. Moreover, lower by approximately 10 % for purpose built flats.

On the other hand, 59.7 % of households living in local authority rented accommodations dwelled in purpose built-flats/maisonette while 11.8% lived semi-detached houses. Likewise, Households who rented their properties from housing associations had roughly similar figures except the percentage for other accommodations was higher by almost 35%. This could be due to the fact that around 31% of the studied households were from the regions with high urban density such as south-east and greater London (UKGOV, 2016).

Finally, purpose built /converted flats and terraced houses seems to be the favourite dwelling types of 67.6% of households who privately rented their properties. This could be attributed to the fact that these dwelling types are the most energy efficient due to their low floor area or/ and good insulation according to the English housing survey (ONS, 2015c).

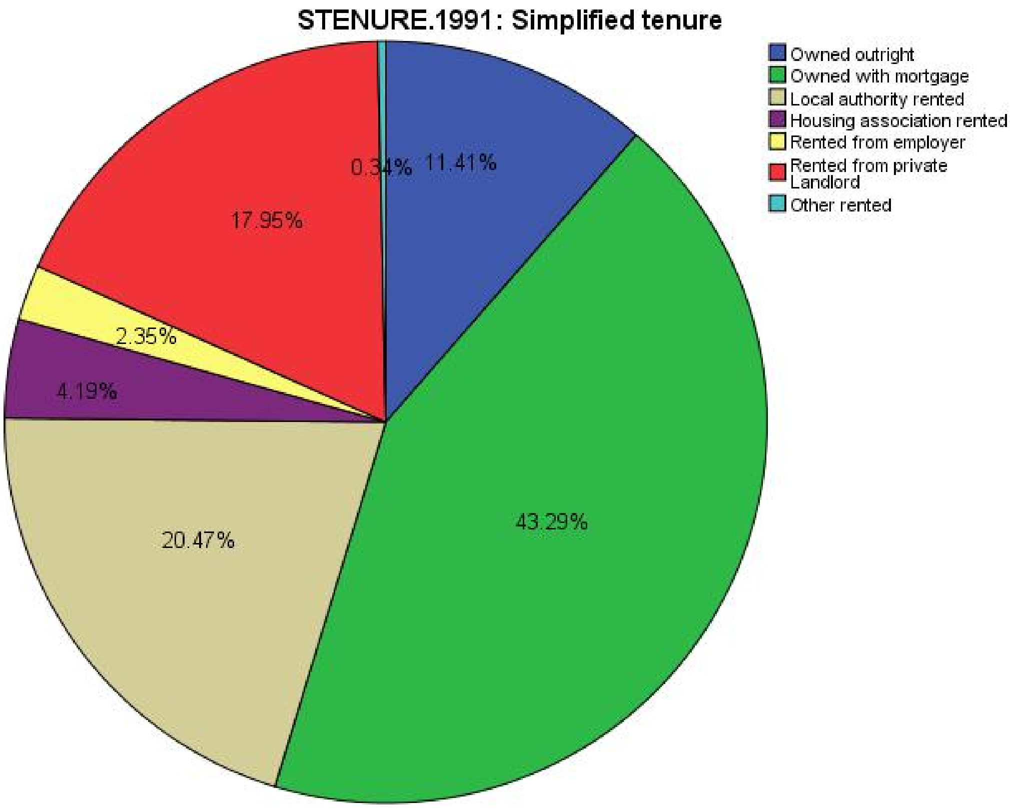



Figure A. 50. The distribution of the studied BHPS households by dwelling tenure

Table A. 8. Highlights the BHPS household tenure modes in function of their dwelling types

| | HSTYPE_NEW.1991: Simplified dwelling type | | | | | | |
|---|---|---|---|---|---|---|---|
| | Det'd house/bungalow | Semi-det'd house/bungalow | End terraced house | Terraced house | Purpose built flat or maisonette | Converted flat or maisonette | Other |
| Owned outright | 27.0% | 27.0% | 9.5% | 17.5% | 14.3% | 0.0% | 4.8% |
| Owned with mortgage | 11.7% | 21.9% | 8.6% | 25.4% | 24.2% | 7.8% | 0.4% |
| Local authority rented | 3.4% | 11.8% | 10.1% | 9.2% | 59.7% | 4.2% | 1.7% |
| Housing association rented | 0.0% | 4.5% | 0.0% | 4.5% | 45.5% | 9.1% | 36.4% |
| Rented from employer | 14.3% | 35.7% | 0.0% | 0.0% | 7.1% | 7.1% | 35.7% |
| Rented from private Landlord | 4.8% | 9.5% | 1.9% | 20.0% | 19.0% | 28.6% | 16.2% |

### A.3.3 Socio-economic and demographic factors affecting household transitions

#### A.3.3.1 Transition to single non-elderly households

This section aims to analyse in detail the logistic regression models representing the transitions to single non-elderly households in the next 3, 4, 6, 7, 8, and 9 years. These models are illustrated in Table A. 9 , Table A. 10, Table A. 11, and Table A. 12.

**Transition after 3 years**

The examination of the model representing the transition to single non-elderly households after 3 years, has revealed the following points. First, the model's independent variables explained 53.3% of the variability in the dependent variable. Secondly, there were no changes in the nature and the number of a significant positive variables on this transition. In other words, cohabiting couples and terraced dwellings were the only variables with significant positive impact. However, there were slight changes in their odds ratios in comparison to the previous model. For example, the odds ratio for cohabiting couples was 0.117 higher than in year 2 model. This could be owing to a possible increase in the proportion of break up rate amongst cohabiting couples from wave 3 to wave 4. Conversely, the odds ratio for terraced homes decreased from 2.445 in the previous model to 2.118 in the actual one. This could be attributed to a change in the income of some households especially the ones with low earning levels. Indeed, there has been approximately around a £1000 increase in the mean of total household annual income from wave 3 to wave 4, as shown in the below line graph Figure A. 51.

In contrast to the above, householder age, size, and log 10 total income had a significant negative impact on the third year transitions to single non-elderly households. This also applies to routine non-manual and households on benefit. In addition to that, for one additional year in the householder age, the odds ratio for being a single non-elderly in 3 years is lower by 37.1%, which is higher by around 11% than the previous model. This was expected since some householders who were aged 24 or less (3 years ago), moved into the 26-45 age band which is characterised by its high transition rates. Similarly, the odds ratio was lower by 47.3% for one additional person in the household. This was roughly in good agreement with the previous model. Surprisingly, for households with routine non-manual socio-economic status, the odds for being single non-elderly households was lower by 56.3%. This was in line with the study of Galobardes et al., (2006) who argued that a large proportion of routine non-manual tend to live in households containing two or more income units. Similarly, the odds of being a single non-elderly household was lower by 49.5% for households receiving benefits. This result is consistent with the UK government report on tax and benefit income which advises that the benefit claimed by single non-elderly households between 1977 and 2015 represents around 4% of the UK total welfare budget (ONS, 2015g). Unlike the previous model, for a one-unit increase in the log 10 of household gross income, the odds ratio was lower by a factor of 0.306. This suggests that households with higher income levels are more likely to move to other family structures such as couples with or without children.

Apart from the above significant factors, household type, tenure mode, and the remaining socio-economic status as well as dwelling types variables, were not found significant.

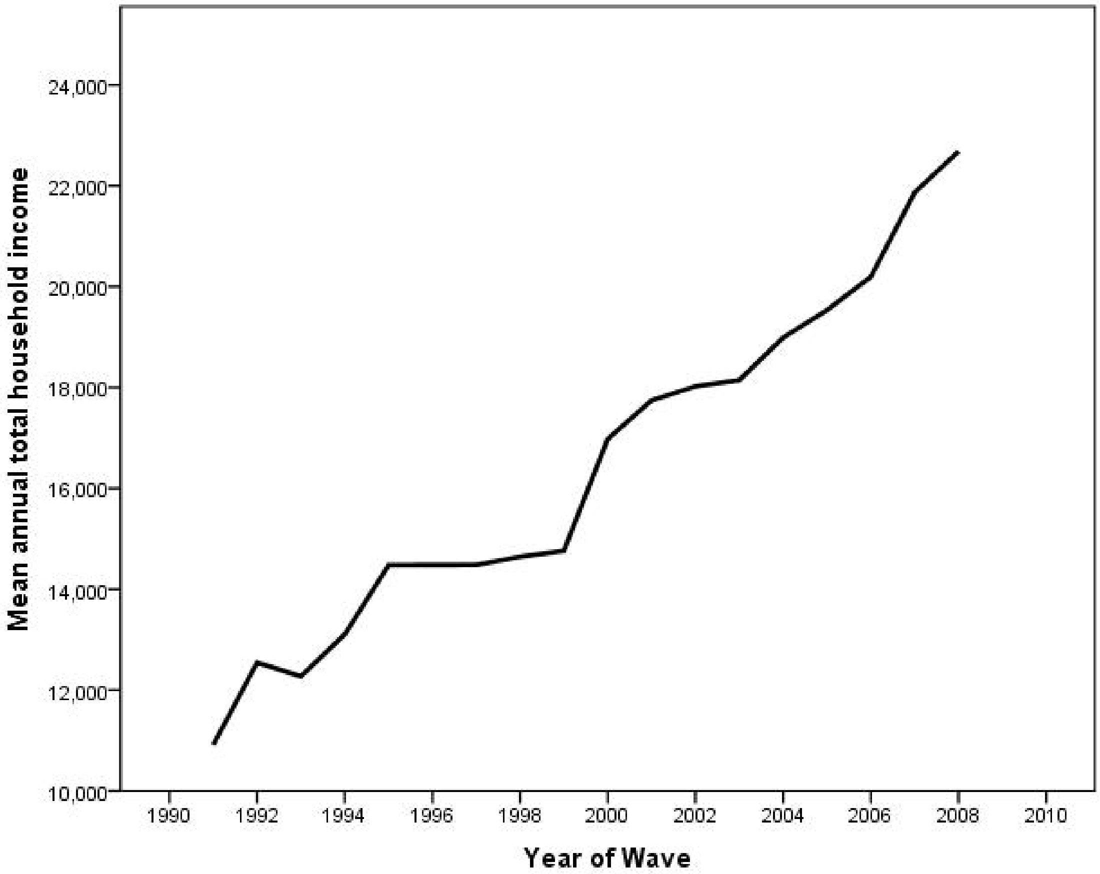



Figure A. 51. Illustrates the change in the mean annual income of single non-elderly household in the BHPS database between 1991 and 2008.

### Transition after 4 years

The study of the model outlining the transition to single non-elderly households in 4 years, has advised the following conclusions.

First, the included independent variables in the logistic regression model explain 59.1% of the variability in the dependent factor. Secondly, the direction and significance of factors comprising cohabiting couples and terraced homes, were consistent with the previous two models despite differences in their odds ratios. In particular, for cohabiting couples in year 4, the odds ratio was approximately twice lower than in the third year model. This was in good agreement with the work of Rosenfeld, (2014) who discovered that the rate of break-ups amongst cohabiting couples who lived together more than 2 years was 30-40% lower than the ones with 1-year relationship (Figure A. 52). Conversely, there was a minor difference of 0.588 in the odds ratio of households living in terraced dwellings in comparison to the model of year 3.

Apart from that, for households with higher and lower-grade socio-economic status, the odds of becoming single non-elderly after 4 years were 2.386 and 2.105, respectively. Although the majority of households belonging to these socio-economic classes are mostly couples without or with dependent children, around 40% of the UK single non-elderly pertain to both categories, according to the UK labour force survey (ONS, 2015d). Finally, the odds ratio for being single non-elderly was unexpectedly 6.25 higher for small proprietors with employees. This is simply owing to the fact that there has been a 24,000 annual growth in the number of this group over the last twenty years according to the UK labour force survey (ONS, 2015d).

On the other hand, single non-elderly, household age, and size had a significant negative effect on transitions to single non-elderly after 4 years. First, the odds ratio was 0.528 lower for one additional unit in the householder age. This was 0.217 and 0.101 lower than the odds ratios of year 2 and 3 models, correspondingly. Again, the increase in the number of population aged between 26-45 is the main reason behind these changes, as addressed previously. Similarly, as expected from the analysis of existing transition patterns between 1991 and 1996, single non-elderly are very likely to move to other family types after 4 years as their odds ratio was lower by 76.9%.

Finally, there were no major changes in the odds of household size in comparison to the previous two models. More precisely, for one additional member in the household, the odds ratio was lower by a factor of 44.1%. Apart from the above, tenure and income variables in addition to the majority of socio-economic, marital status, and dwelling types variables were not significant.

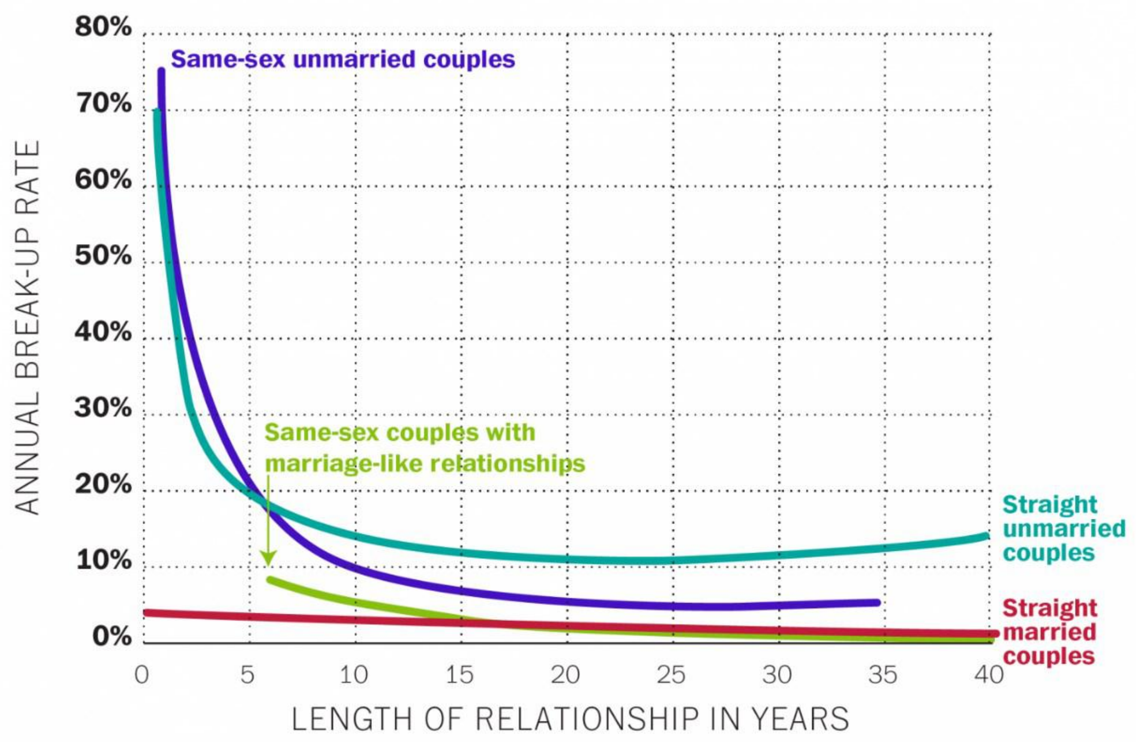



Figure A. 52.The percentage of annual break up rates between couples in function of the length of relationship in years (Rosenfeld, 2014).

Table A. 9.Logistic regression transition models to single non-elderly households in the next 3 and 4 years, (part1)

| | | (year 3) Single non-elderly | (year 4) Single non-elderly |
|---|---|---|---|
| Marital status | Cohabiting couples | 5.290*** (1.880) | 2.31*** (0.87) |
| | Widowed | - | 0.280 (0.211) |
| | Separated | - | 0.152 (0.152) |
| | Never Married | 2.192 (1.029) | - |
| Age and household type | Householder Age | 0.629*** (0.0163) | 0.528*** (0.0193) |
| | Single non-elderly | 1.385 (0.454) | 0.231*** (0.0955) |
| | Lone parent | - | 0.530 (0.369) |
| | Household size | 0.527*** (0.0988) | 0.559** (0.119) |
| Accommodation related variables | Terraced house | 2.188* (0.709) | 2.746** (1.065) |
| | Purpose built flat or maisonette | 1.703 (0.605) | 1.659 (0.776) |
| | Dwelling size | - | - |
| | Owned outright | 1.618 (0.530) | 1.900 (0.744) |
| | Rented from housing associations | 0.313 (0.238) | 0.264 (0.219) |

Note: standard errors in parentheses *Significance at the 95% level **Significance at the 99% level

Table A. 10.Logistic regression transition models to single non-elderly households in the next 3 and 4 years, (part2)

| | | (year 3) Single non-elderly | (year 4) Single non-elderly |
|---|---|---|---|
| Socio-economic class variables | Higher-grade professionals | - | 2.386* |
| | | | (0.960) |
| | Lower-grade professionals | - | 2.105* |
| | | | (0.695) |
| | Routine non-manual employees | 0.437* | - |
| | | (0.164) | |
| | Foreman or technicians | - | - |
| | Personal service workers | - | - |
| | Small proprietors with employees | - | 6.21* |
| | | | (5.66) |
| Income related variables | Log 10 total income | 0.306* | 0.593 |
| | | (0.145) | (0.287) |
| | Receiving benefit | 0.505* | 0.447 |
| | | (0.136) | (0.145) |
| | McFadden's R2 | 0.533 | 0.591 |
| | Type of employed models | Fixed-effects | Fixed-effects |
| | Sample size | 2631 | 2302 |

Note: standard errors in parentheses *Significance at the 95% level **Significance at the 99% level

### Transition after 6 years

The analysis of the logistic regression model representing the transitions to a single non-elderly household after 6 years has suggested that the model's independent variables explain 46.6% of the variation in the dependent factor. In addition to that, it has suggested that the majority of significant variables had a negative effect on the transition to single non-elderly households. This includes the following factors namely; Single non-elderly, householder age, purpose built flat, converted flat, and square root of household annual benefit income.

First, the odds ratios for becoming single non-elderly after 6 years were lower by 44.8% and 94.43% for one additional unit in the householder age and for single non-elderly households, respectively. In comparison to the previous models (year 2-5), it is clear that these odds ratios are the lowest. This implies that the effect of householder age and being single non-elderly on this transition type increases with time. As for households living in purpose built and converted flats, the odds of becoming single non-elderly within 6 years is lower by 87.3% and 97.54%, correspondingly. As addressed at an earlier stage, this is owing to the fact that flats are mostly occupied by couples. Finally, the square root of household annual benefit income was significant but its effect was so weak since for a one-unit increase in this variable, the odds ratio was lower by 1.1%.

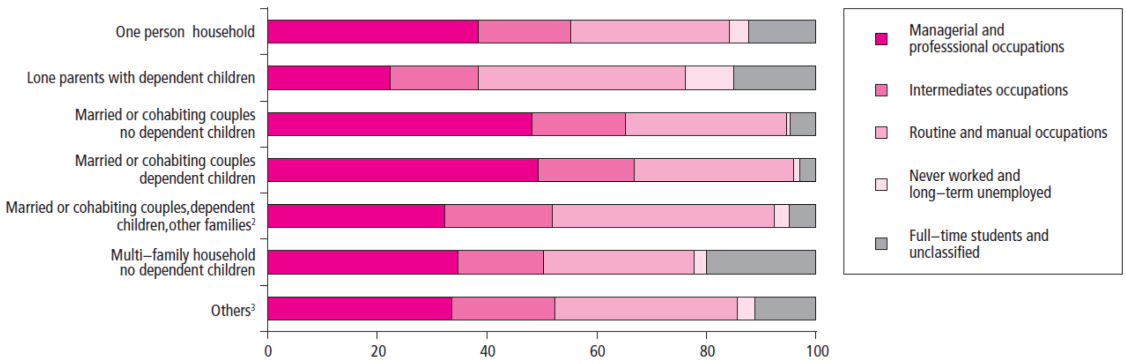
On the other hand, a lot of variables were not found significant at the 5% level. This comprises marital status, socio-economic status, dwelling characteristics, and tenure type variables. Moreover, the rest of household type and income factors.

### Transition after 7 years

In comparison to the year 6 model, the study of the model pertaining the 7$^{th}$ year of transitions to single non-elderly households has suggested similar findings except the significance of outright tenure, routine non-manual, and small proprietors with employees. In addition to that, it has indicated that the included independent factors in the model are responsible for 57.7% of the variation in the dependent variable.

In more details, the odds ratio was lower by 61.1% for households who own their properties outright. This was in line with the UK home ownership report claiming that around 50% of owned outright homes are occupied by at least 2 people (ONS, 2013d). Similarly, for routine non-manual employees and small proprietors with employees, the odds ratios were 0.417 and 0.0334, respectively. The odds ratio for the routine non-manual class was in good agreement with the ones of previous models and consequently with Galobardes et al. (2006) study, as discussed earlier. Conversely, the odds ratio related to small proprietors with employees was not in good agreement with year 4 model. Although it is controversial, it can be accepted since the majority of households belonging to this class are couples with and without dependent children (Figure A. 53). Nevertheless, the reason behind the high odds ratio in model 4, has to be further investigated in the future.

Finally, marital status factors as well as the remaining household type, dwelling characteristics, tenure mode, socio-economic, and income variables, were not significant.




1 Data are for working age only. Men aged 16–64, women aged 16-59.
2 Data in this category include cases where both families in the household have dependent children, or where only one family in a multi-family household has dependent children.
3 Others includes same-sex family units with or without other families present, lone parents with non-dependent children both with and without other families and lone parents with dependent children with other families present.

*Source: Labour Force Survey, Office for National Statistics*

Figure A. 53. illustrates the socio-economic class profile of different working UK household types (ONS, 2015d).

### Transition after 8 years

In contrast to the two previous models, the examination of the transition model to single non-elderly households in 8 years, suggested that the independent variables cause the dependent factor to vary by 59%. In addition to that, it has advised that four variables namely; higher-grade, lower-grade, foreman technicians, and receiving state benefit, had a positive significant effect.

In particular, the odds ratio for becoming single non-elderly over 8 years was 3.906 and 2.625 for higher and lower-grade professionals, correspondingly. This could be owing to a possible divorce occurrence amongst these socio-economic classes of which 60% of them tend to get married (Goodman and Greaves, 2010). Indeed, despite the family stability of households with professionals and managerial occupations, a recent UK official report claimed that the average marriage length in the UK ranges between 8.9 and 12.2 years. For those reasons, it is believed that these odds ratios are logical and consistent with existing governmental reports. Similarly, for foreman/ technicians, the odds of being single non-elderly was 3.046. This can be interpreted in two distinct ways. The first one could be attributed to possible divorces knowing that the divorce rate among this class is roughly 10% (ONS, 2013b). On the other hand, the second argument is built upon the fact that the proportion of single non-elderly is relatively considerable in this class (31.9%) (ONS, 2015h). However, since there is a lack of similar studies, these issues need to be further examined in the future. Finally, unlike the results pertaining to year 3 model, the odds of becoming single non-elderly in the next 8 years was higher by a factor of 2.425 for households receiving benefit. The reason behind this disagreement could be due to possible transitions of some lone parent households to a single non-elderly family while remaining on benefit.

In contrast to the above, variables namely; single non-elderly households, householder age, log 10 of total annual household income, and the square root of household benefit income, had a significant negative impact on the dependent variable. More specifically, the odds ratios of variables, including householder age, single non-elderly, and the square root of annual benefit income, were very close to the ones of year 6 and 7 models. For the log10 household annual income; however, the odds of becoming single non-elderly in 8 years was 80.9% lower for a one-unit increase in this variable. Indeed, this figure is supported since households with higher income are usually couples with or without dependent children (HM Treasury, 2014). Similarly, the odds ratio was lower by 30.5% for one additional bedroom in the accommodation. This aligns with the English household survey which suggests that larger dwellings are mostly occupied by families with children (ONS, 2015c). This implies a possible relationship between household size and dwelling size as discussed at an earlier stage.

Finally, unlike the previous two models, dwelling type and tenure mode variables were not found significant. This is possibly due to the fact that around 20% of one - person household in the UK own their dwellings outright according to Census 2011 (ONS, 2011d).

**Transition after 9 years**

Overall, the analysis of the model representing the transition possibilities of being single non-elderly households after 9 years, indicated that the independent factors are responsible for 57.1% of the variation in the dependent factor. Furthermore, it has suggested that all significant variables had a negative effect.

In addition to the shared negatively significant variables in the previous and actual models, the odds ratio of being single non-elderly in 9 years was lower by 70.5% for cohabiting couples. Similarly, there was a slight change in the odds ratio of single non-elderly households in comparison to the previous models. However, despite this change, the likelihood of remaining within the same household type was very small as the odds ratio was lower by 83.5%. For households living in other accommodations such as bedsits, business premises, and sheltered accommodations, the odds ratio for being single non-elderly was lower by 88.3%. This is because sheltered accommodations in the UK are mostly occupied by elderly, disables, or vulnerable households (GOVUK, 2015). Similarly, despite being a minority, households living in their business premises, including guesthouses and bed and breakfast, are usually couples without children (GOVUK, 2017d);(Hogan, 2016). Like in the model of year 7, households with "small propitiators with employees" socio-economic status were very unlikely to become single non-elderly as their odds ratio was lower by 96.23%.

Finally, the remaining marital status, dwelling type, and socio-economic variables in addition to dwelling size, receiving benefit, and tenure mode variables, were not found significant.

Table A. 11.Logistic regression transition models to single non-elderly households in the next 6, 7, 8, and 9 years. (part1)

| | | (year 6) Single non-elderly | (year 7) Single non-elderly | (year 8) Single non-elderly | (year 9) Single non-elderly |
|---|---|---|---|---|---|
| **Marital status** | Cohabiting couples | - | | - | $0.295^{*}$ (0.166) |
| | Separated | 0.303 (0.327) | 3.187 (2.777) | 8.788 (7.690) | - |
| | Never Married | - | 3.771 (2.767) | - | 1.933 (1.625) |
| **Household age and type** | Householder Age | $0.552^{***}$ (0.0217) | $0.45^{***}$ (0.0204) | $0.424^{***}$ (0.0233) | $0.423^{***}$ (0.0243) |
| | Single non-elderly | $0.0557^{***}$ (0.0232) | $0.0123^{***}$ (0.00605) | $0.0145^{***}$ (0.00983) | $0.165^{*}$ (0.122) |
| | Couple without children | - | - | 2.515 (1.438) | 3.362 (1.764) |
| | Lone parent | - | - | 2.456 (2.058) | - |
| | Other households | 0.449 (0.426) | | | |
| **Accommodation related variables** | Purpose built flat or maisonette | $0.127^{***}$ (0.0672) | $0.333^{*}$ (0.179) | - | 0.293 (0.251) |
| | Converted flat or maisonette | $0.0246^{***}$ (0.0233) | $0.153^{*}$ (0.116) | 0.285 (0.231) | |
| | Terraced dwelling | - | - | - | 0.472 (0.238) |
| | Other accommodations | - | 0.325 (0.251) | - | $0.117^{**}$ (0.0962) |
| | Dwelling size | 0.879 (0.105) | 0.833 (0.109) | $0.695^{*}$ (0.0982) | - |
| | Owned outright | - | 0.389* (0.197) | - | - |
| | Rented from authorities | 2.168 (1.490) | - | - | 3.478 (3.260) |
| | Rented from private landlord | - | - | - | 3.441 (2.454) |

Table A. 12.Logistic regression transition models to single non-elderly households in the next 6, 7, 8, and 9 years. (part2)

| | | (year 6) Single non-elderly | (year 7) Single non-elderly | (year 8) Single non-elderly | (year 9) Single non-elderly |
|---|---|---|---|---|---|
| Socio-economic class related variables | Higher-grade professionals | 1.362 | - | 3.906** | - |
| | | (0.485) | | (1.809) | |
| | Lower-grade professionals | - | 1.493 | 2.625* | - |
| | | | (0.501) | (1.065) | |
| | Routine non-manual employees | - | 0.417* | | |
| | | | (0.196) | | |
| | Foreman or technicians | - | - | 3.046* | 2.728 |
| | | | | (1.781) | (1.828) |
| | Personal service workers | - | - | | 0.0861 |
| | | | | | (0.123) |
| | Small proprietors with employees | 4.829 | 0.0334** | - | 0.0377* |
| | | (1.76) | (0.0430) | | (0.0684) |
| | Farmers/ small holders | 4.432 | | | |
| | | (5.983) | | | |
| Income related variables | Working full-time | - | - | 1.647 | - |
| | | | | (0.799) | |
| | Log 10 total annual household income | 0.850 | 0.650 | 0.191** | 0.148*** |
| | | (0.403) | (0.297) | (0.115) | (0.0834) |
| | Receiving benefit | 1.711 | 1.810 | 2.425* | - |
| | | (0.552) | (0.625) | (1.055) | |
| | Square root household annual benefit income | 0.989* | 0.990* | 0.993* | 0.998* |
| | | (0.00551) | (0.00463) | (0.00459) | (0.00532) |
| | Mc Fadden's R2 | 0.466 | 0.577 | 0.590 | 0.571 |
| | Type of used model | Fixed-effects | Fixed-effects | Fixed-effects | Fixed-effects |
| | Sample Size | 1534 | 1902 | 1612 | 1363 |

Note: standard errors in parentheses *Significance at the 95% level **Significance at the 99% level

#### A.3.3.2 Transition to couple without children households

This section intends to examine in depth the logistic regression models representing household transitions to couples without children in the next 2, 3, 4, 6, 7, 8, and 9 years. These models are illustrated in Table A. 13, Table A. 14, Table A. 15, Table A. 16, and Table A. 17.

**Transition after 2 years**

The examination of the logistic regression model related to transitions to couples without children after 2 years, advised that independent factors make the dependent variable to vary by 36.7%. This examination also suggested no major changes to report from the previous model. However, other variables namely; other households, labour-income, and higher-grade professionals, were found significant. In addition to that, there were minor changes in the odds ratios of certain variables which are common across the current and the previous model such as couple with children.

First, the odds ratio was higher by a factor of 3.745 for other household types, who are unrelated households composed of at least 2 occupants sharing the dwelling facilities. According to Hall and White (2005), young people are the ones that are most likely to live in unrelated groups, which justify their higher likelihood of moving to different household structures including couples without children. Secondly, for higher-grade professionals the odds ratio for becoming couple without children in 2 years was lower by 39%. This was in line with the results of the 4$^{th}$ model of single non-elderly transitions and certainly with findings of the UK labour force survey, which suggested that 40% of the UK single non-elderly hold this socio-economic status (ONS, 2015d). In addition to the above variables, the log 10 of household labour annual income was negatively significant but with minor effect as for a one-unit increase in the variable, the odds ratio was lower by 1.4%.

On the other hand, the odds ratios for never married, divorced, and couples with children almost doubled over this period. This suggests that the likelihood of being a couple without children for these particular households increases in function of time. A clear evidence on this claim is based on the fact that some mid-age couples with children are likely to be empty nesters as a result of children leave. In addition to the above factors, there was a 2.234 decrease in the odds ratio of households on pension in comparison to the previous model. This might be attributed to a possible decrease in the proportion of couple pensioners as a result of divorce. Indeed, a recent report from Pension Wise agency argued that around 16,000 couple pensioners in the UK got divorced, which results in most cases in splitting the pension equally between both parties (GOVUK, 2017b).

Finally, the remaining household type, socio-economic status, tenure more in addition to dwelling type variables, were not found statistically significant at both the 1% and 5% levels.

**Transition after 3 years**

The inspection of the results pertaining to the third year of transition to couples without children has suggested that the model's independent variables are responsible for 34.4% of the variation in the dependent factor. Furthermore, the analysis of this model advised similar findings to the previous models except for some changes in the odds ratios of certain variables (e.g. On pension). Finally, some variables which were significant in the previous model(s) such as separated, were found insignificant and reciprocally.

In more detail, for couples with children, the odds ratio for being couples without children in the coming three years was lower by a factor of 0.379, which is almost the triple of the previous model's odds ratio. This reinforces the argument made in the previous model with regards the growing likelihood of becoming empty nesters over time. Similarly, marital status variables namely; divorced, widowed, and never married, were also prone to these changes as their odds ratios increased by 0.129, 0.1, and 0.14, correspondingly. On the other hand, positively significant variables such as on pension, were also affected by changes in their odds ratios. For instance, there was 0.149 increase in the odds ratio of the variable other households from the previous model, whereas the odds ratio of households on pension, decreased by 1.11. As explained in the previous paragraph, this change might be associated with a possible decrease in the proportion of pensioner couples as a result of divorce. As for the variable "other households", Stone et al. (2011) and Daly (2005) argued that while there has been an increase in the proportion of individuals living in non-family arrangements in the UK, they are more likely to move to couple without children households mostly via cohabitation.

In contrast to the above, for households with foreman technicians' socio-economic status, the odds ratio for becoming couples without children in 3 years was higher by a factor of 2.589. Given the fact that the proportion of single non-elderly households belonging to this socio-economic class is relatively high (31.9%), as discussed at an earlier stage, it could be possible that some of them made the transition to a couple without children household over this period. Another probability might be due to the occurrence of possible transitions from a couple with children to a couple without children household as a result of children leaving the household. Conversely, separated and higher-grade professional variables, which were significant in the previous model, were not found significant in the actual one. This is might be because the proportion of separated households decreased dramatically in the BHPS dataset from 27% in wave 1 to 8% wave 3. Similarly, the main reason behind the non-significance of socio-economic class variables, including high-grade professional, is due to the non-significance of household annual labour income variable in this model.

Finally, apart from the non-significance of log10 of household annual labour income, the odds ratios of the remaining income variables namely; the square root of annual household benefit income and log 10 gross annual household income, remained fairly stable in comparison to the previous models.

### Transition after 4 years

Following the analysis of the logistic regression model representing transitions to couple without children households after 4 years, it is evident that the model's independent variables make the dependent factor to vary by 28.5%. This examination also advised that the significance of certain variables, including couple with children, was inconsistent with the previous model(s).

First, the marital status variables namely; widowed and separated, were not statistically significant. However, the former variable was significant at the 10% level while maintaining a similar magnitude to the previous model. Similarly, the variable couple with children was also not significant. Furthermore, given the 95% confidence interval of its odds ratio [0.55- 2.09], it is not possible to confirm the nature of its effect on this type of transition.

In contrast to the previous models (years 1, 2, and 3), the odds ratio for being couples without children was higher by a factor of 6.018 for personal service workers. This was very consistent with the results of the model related to transitions to single-non elderly households in 5 years. More specifically, given the age distribution of households belonging to this class, of which almost half are aged between 25 and 44, they are more likely to live in other family types than single non-elderly (Begum, 2004).

On the other hand, the odds ratios were 0.361, 0.221, and 1.179 for divorced, never married, and householder age variables, respectively. This shows their consistency with all the previous models, despite some changes in their odds ratios. Additionally, the odds ratio of the variable never married indicates that the probability of remaining single non-elderly households decreases over time.

Finally, tenure mode, dwelling types, and the remaining socio-economic variables were not significant.

Table A. 13. Logistics regression models of transition to couple without children households in the next 2, 3, and 4 years (Part1)

| | | (Year 2) | (Year 3) | (Year 4) |
|---|---|---|---|---|
| | | Couple No | Couple No | Couple No |
| Household age and type variables | Not main householder | - | - | - |
| | Age of the householder | 1.079*** | 1.126*** | 1.179*** |
| | | (0.0244) | (0.0259) | (0.0284) |
| | Couple with children | 0.132*** | 0.379** | - |
| | | (0.0388) | (0.117) | |
| | Lone parent | 0.155 | 0.307 | 0.289 |
| | | (0.199) | (0.305) | (0.381) |
| | Other households | 3.745* | 3.894* | 3.576* |
| | | (2.474) | (2.455) | (2.443) |
| Marital status | Divorced | 0.116*** | 0.245*** | 0.361* |
| | | (0.0445) | (0.0926) | (0.176) |
| | Widowed | 0.0257*** | 0.124** | 0.189 |
| | | (0.0226) | (0.0948) | (0.186) |
| | Separated | 0.292** | - | 1.883 |
| | | (0.129) | | (1.137) |
| | Never Married | 0.0864*** | 0.221*** | 0.472* |
| | | (0.0273) | (0.0740) | (0.180) |

Note: Couple No= couple without children / standard errors in parentheses

*Significance at the 95% level **Significance at the 99% level

Table A. 14.Logistics regression models of transition to couple without children households in the next 2, 3, and 4 years (Part2)

| | | (Year 2) Couple No | (Year 3) Couple No | (Year 4) Couple No |
|---|---|---|---|---|
| Dwelling related variables | Dwelling owned with mortgage | 0.551* | 0.741 | - |
| | | (0.141) | (0.211) | |
| | Renting from local authorities | 0.283 | - | - |
| | | (0.211) | | |
| | Renting from housing associations | 0.0966 | - | - |
| | | (0.132) | | |
| | Renting from employer | 0.0343 | - | - |
| | | (0.0648) | | |
| | Renting from private landlords | - | 0.562 | 0.617 |
| | | | (0.282) | (0.287) |
| | Dwelling size | - | - | - |
| Socio-economic class related variables | Personal service workers | 0.587 | - | 6.018* |
| | | (0.378) | | (5.038) |
| | Routine Non-manual | | | |
| | Foremen technicians | - | 2.589* | |
| | | | (1.185) | |
| | Small proprietors with employees | 0.370 | 0.262 | |
| | | (0.318) | (0.230) | |
| | Small proprietors without employees | 0.769 | 0.675 | 0.604 |
| | | (0.384) | (0.362) | (0.379) |
| | Farmers/ smallholders | | | 5.131 |
| | | | | (5.850) |
| | Skilled manual workers | | | 3.289* |
| | | | | (1.780) |

Table A. 15.Logistics regression models of transition to couple without children households in the next 2, 3, and 4 years (Part3)

| | | (Year 2) | (Year 3) | (Year 4) |
|---|---|---|---|---|
| | | Couple No | Couple No | Couple No |
| Income related variables | On Pension | 5.688*** | 4.576** | 7.07*** |
| | | (2.938) | (2.325) | (2.768) |
| | On Benefit | | | |
| | Square root of total household annual income | 1.027** | 1.021* | 1.007* |
| | | (0.00885) | (0.00867) | (0.00317) |
| | Square root of household annual benefit income | 0.977*** | 0.978*** | 0.970*** |
| | | (0.00539) | (0.00556) | (0.00590) |
| | Square root of household labour income | 0.986* | 0.989 | - |
| | | (0.00702) | (0.00709) | |
| | Working Full-time | 0.694 | 0.592 | 0.437** |
| | | (0.197) | (0.167) | (0.138) |
| | Mc Fadden's R2 | 0.367 | 0.345 | 0.285 |
| | Type of used model | Fixed-effects | Fixed-effects | Fixed-effects |
| | Sample Size | 1518 | 1359 | 1143 |

Note: Couple No= couple without children / standard errors in parentheses

*Significance at the 95% level **Significance at the 99% level

## Transition after 6 years

The analysis of the transition model of year 6 has advised that the model's independent factors account for 21.3% of the variation in the dependent variable. This examination also suggested a balance between positively and negatively significant variables. Furthermore, it has advocated the below changes from last model (year 5).

First, as expected given precedent changes in their odds ratios over time, the variables namely; never married and lower-grade professional, were statistically significant in comparison to the previous model. In particular, never married householders were very likely to move to couples without children after 6 years as their odds ratio was higher by a factor 2.117. Lower-grade professionals; however, were unlikely to be couples without children after 6 years as the odds ratio was lower by 65.8%. As explained in the previous models, higher and lower- grade professionals, who belong to "managerial and professional occupations" category according to the labour force survey report (ONS, 2015d), are characterised by their financial stability. Therefore, their transition to couples with children is quite fast. This is evident in the high proportion (63%) of young mothers aged 30 years and over in this socio-economic class, according to the UK government report on births by parent characteristics (ONS, 2015a). In contrast to the above and the previous model, variables such as renting from local authorities, renting from private landlords, living in purpose-built flats, routine non-manual, on pension, and log10 gross household annual income, were insignificant.

On the other hand, the subsequent variables with high odds ratios namely; couple with children, separated, age of householder, and living in converted flats, were consistent with the ones of the precedent model. More specifically, for couple with children households, the odds ratio for becoming couple no children households in 6 years was higher by 3.795. Similarly, for a one-unit increase in the householder age, the odds for being couples without children over 6 years was higher by 9.2%. However, there was a significant change of 2.42 and 3.67 in the odds ratios of separated and living in converted flats, respectively. Premised on the argument made when analysing the model of year 5, it is possible that a proportion of couples without children living in converted flats became couples with children after the first childbirth; consequently, they moved to another dwelling type (e.g. terraced). Hence, the lower odds ratio.

Finally, the odds ratios of the following negatively significant variables namely; higher-grade professionals, foreman technicians, semi-unskilled manual workers, the square root of total annual benefit income, and working full-time, were in line with the ones of the previous model despite minor changes of 0.08 on average.

**Transition after 7 years**

The analysis of the model representing transition probabilities to couple without children households in the next 7 years, has advocated that the independent factors are responsible for 19.5% of the variation in the dependent variable. This investigation also advised that the number of variables with a negative impact was slightly superior to the ones with positive effect. Furthermore, it has demonstrated that some factors (e.g. lower-grade), which were significant in the previous model, were insignificant in the actual one and vice versa.

First, the odds ratios of variables with positive effect namely; couple with children, separated, never married, and age of the householder, have increased next to the ones of previous models. For instance, the odds ratio for couples with children was 7.022 which is approximately 1.85 times the previous model's odds ratio. Similarly, the odds ratio of the variable never married in the actual model was roughly 1.6 times the one of the previous model. In addition to that, the odds ratio of the variable higher-grade, which is negatively associated with the dependent variable, rose by 69% from the precedent year's model. As addressed at an earlier stage, the likelihood of becoming couples without children for households with those socio-economic characteristics (e.g. couple with children) increases in function of time. Indeed, after performing another regression analysis, where the year of the wave is the independent variable, it was found that for one additional year, the odds ratio for being a couple without children family increases by 23%.

In contrast to the above, the odds ratios of other variables with a negative effect on this transition model namely; foreman technicians, semi-unskilled manual workers, the square root of total annual benefit income, and working full-time, remained relatively stable over the last three models. For example, there has been a 0.7% rise in the odds ratio of the variable working full-time from the precedent year's model.

Finally, owned with a mortgage, on Benefit, and on pension, were significant variables in this model. In particular, for households who have been owning their accommodation for 6 years, the odds of being couples without children in the following year was lower by 48.6%. This was in line with the result of year 1 model. Additionally, the odds ratio for being couples without children in 7 years was lower by 54.2% for households on benefit. Thus, this outcome was in good agreement with the one of the year1 model. Conversely, the following variables namely; living in converted flats and lower-grade-professionals, were statistically insignificant.

**Transition after 8 years**

In contrast to the previous model, the examination of the logistic regression model representing the transition probabilities to a couple without children family after 8 years, has suggested that the independent variables cause the dependent factor to vary by 18.8%. Moreover, this analysis also advised the subsequent changes.

First, the odds ratios of the variables; couples with children and separated decreased by 1.78 and 7.1, correspondingly. The latter result indicates that there is a possible relationship between the length of separation and the likelihood of becoming couple without children. However, after conducting some research via analysing various resources, including journal articles, government reports and others, there was not enough evidence to draw conclusions. Thus, more research is needed in this respect. As for the variable couple with children's odds ratio, it could be attributed to an increase in the proportion of young couples with children in the BHPS dataset.

Conversely, as expected from analysing precedent models, never married and householder's age variables witnessed a rise of 2.68 and 0.08 in their odds ratios, respectively.

On the other hand, variables with a significant negative impact namely; higher-grade professionals and semi-unskilled workers, had a decrease of 0.191 and 0.118 in their odds ratio, correspondingly. Apart from that, the variable routine non-manual was found significant in this model with an odds ratio of 0.345. Those results were excepted given the fact that the majority of households from those socio-economic classes live in couple with children families as indicated previously in year 6 model.

Finally, the variables; on pension, on benefit, working full-time, and foreman technicians, were not found statistically significant in this model.

**Transition after 9 years**

After analysing the model outlining the transition probabilities to a couple without children household in 9 years, it is evident that the independent factors are responsible for 18% of the variation in the dependent variable. In addition to that, the examination of the model table suggested that variables which were statistically significant across the previous 8 years such as couple with children, were not found significant in this model. Finally, the number of variables with a negative effect was nearly equal to the ones with a positive impact.

First, "couple with children", "separated", and "routine non-manual" were not found significant in the model. However, this does not signify the absence of an effect but instead there is not enough evidence in explaining the variation of the dependent variables in function of those predictors. On the other hand, for households who were never married 8 years ago, the odds for being couple without children in the following year was higher by a factor of 7.561. Similarly, for one additional year in the householder age, the odds for becoming couple no children after 9 years was higher by 41.9%.

In contrast to the above, for higher and lower-grade professionals, the odds for being couples without children in 9 years was lower by 69% and 68.4%, respectively. Also, the odds ratio was lower by 93.39% for semi-unskilled manual workers; whereas only 0.6% lower for one-unit increase in the square root of total household annual benefit income.

Finally, like in the previous models, variables pertaining to dwelling type, tenure mode in addition to the remaining socio-economic class, marital status, and income variables were not found significant both at the 5 and 1% levels.

Table A. 16.Logistics regression models of transition to couple without children households in the next 6, 7, 8, and 9 years (Part1)

| | | (Year 6) | (Year 7) | (Year 8) | (Year 9) |
|---|---|---|---|---|---|
| | | Couple No | Couple No | Couple No | Couple No |
| Marital status / household age and type variables | Couple with children | 3.795*** | 7.022*** | 5.242** | 1.848 |
| | | (1.525) | (3.435) | (2.958) | (1.068) |
| | Lone parent | 0.319 | 0.220 | 0.157 | 0.169 |
| | | (0.332) | (0.232) | (0.176) | (0.189) |
| | Separated | 8.771** | 10.74*** | 3.729* | 1.02896 |
| | | (6.280) | (7.570) | (2.408) | (0.8783) |
| | Never Married | 2.117* | 3.378* | 6.063* | 7.561* |
| | | (1.054) | (2.085) | (5.335) | (7.433) |
| | Age of the householder | 1.092*** | 1.218*** | 1.302*** | 1.419*** |
| | | (0.0282) | (0.0340) | (0.0419) | (0.0564) |
| Dwelling type variables | Terraced Dwelling | 1.830 | | | |
| | | (0.874) | | | |
| | Purpose built flat | 2.477 | 0.647 | 0.512 | 0.244 |
| | | (1.385) | (0.377) | (0.3933) | (0.227) |
| | Converted flat | 4.413* | 1.449 | 0.564 | 0.395 |
| | | (2.966) | (0.958) | (0.517) | (0.410) |
| | | (0.203) | | (2.029) | (3.080) |

Note: Couple No= couple without children / standard errors in parentheses

*Significance at the 95% level **Significance at the 99% level

Table A. 17.Logistics regression models of transition to couple without children households in the next 6, 7, 8, and 9 years (Part2)

| | | (Year 6) Couple No | (Year 7) Couple No | (Year 8) Couple No | (Year 9) Couple No |
|---|---|---|---|---|---|
| Socio-economic class related variables | Higher-grade professionals | 0.267*** | 0.452* | 0.261** | 0.310** |
| | | (0.0903) | (0.147) | (0.110) | (0.137) |
| | Lower-grade professionals | 0.342*** | 0.616 | 0.460 | 0.316** |
| | | (0.110) | (0.186) | (0.186) | (0.137) |
| | Routine Non-manual | 0.545 | | 0.345* | 0.369 |
| | | (0.224) | | (0.174) | (0.197) |
| | Foremen technicians | 0.280* | 0.286* | 0.293 | 0.299 |
| | | (0.145) | (0.164) | (0.198) | (0.219) |
| | Small proprietors with employees | 0.258 | 0.416 | 0.271 | 0.166 |
| | | (0.224) | (0.367) | (0.248) | (0.191) |
| | Small proprietors without employees | | 3.646 | 2.192 | |
| | | | (2.412) | (1.453) | |
| | Semi-skilled manual workers | 0.207** | 0.286* | 0.168** | 0.0661** |
| | | (0.116) | (0.164) | (0.114) | (0.0564) |
| Income related variables | On Pension | 2.326 | 4.125* | 0.770 | 0.371 |
| | | (1.139) | (2.369) | (0.517) | (0.277) |
| | On Benefit | | 0.458* | 0.699 | 1.22 |
| | | | (0.144) | (0.238) | (0.4384) |
| | Square root of household annual benefit income | 0.990** | 0.996* | 0.997* | 0.994* |
| | | (0.00346) | (0.00331) | (0.00324) | (0.00284) |
| | Working Full-time | 0.441* | 0.474* | 0.896 | 1.22 |
| | | (0.143) | (0.157) | (0.334) | (0.509) |
| | Mc Fadden's R2 | 0.213 | 0.195 | 0.188 | 0.180 |
| | Type of used model | Fixed-effects | Fixed-effects | Fixed-effects | Fixed-effects |
| | Sample Size | 1107 | 1096 | 924 | 776 |

Note: Couple No= couple without children / standard errors in parentheses

*Significance at the 95% level **Significance at the 99% level

### A.3.3.3 Transition to Lone parent households

This section aims to analyse the logistic regression models representing household transitions to lone parent households in the next 2, 3, 4, 6, 7, 8, and 9 years. These models are illustrated below in Table A. 19, Table A. 20, Table A. 21, Table A. 22, Table A. 23, and Table A. 24.

**Transition after year 2**

The examination of the model outlining the transition probabilities to lone parent households in the next 2 years, has advised that the independent factors are responsible for 26.8% of the variation in the dependent variable. Moreover, this analysis also advocated the insignificance of certain variables which were significant in the previous model such as skilled manual workers. Finally, it has suggested an increase in the odds ratios of the majority of the variable.

First, like in the previous model, the effect of household type variables was negligible. In particular, the odds ratios were 0.206, 0.0283, and 0.0308 for single non-elderly, couple without children, and couple with children, respectively. However, these odds ratios were 1.5, 4.2, 48 times the odds of single non-elderly, couple without children, and couple with children in the precedent model, respectively. This indicates a possible relationship between time and likelihood of becoming lone parent households for these types.

On the other hand, variables with significant impact on the dependent variable in the previous model namely; foreman technicians, skilled-manual workers, and semi-unskilled manual workers, were not found significant in the actual model. Even though a considerable proportion of lone parents belong to these socio-economic status, as indicated previously, lone parent mothers are prone to job instability. Indeed, based on the British household panel data survey, Iacovou and Berthoud

(2000) ; Blundell et al. (2005) found that 52% of lone parent mothers on part-time jobs were unlikely to remain in their jobs for more than a year. Conversely, the variable "living in end-terraced" was statistically significant as the odds ratio for being lone parent households in two years was higher by 5.469. This aligns with the English housing survey (ONS, 2011b) suggesting that around 38% of lone parent households lived in terraced houses in 2007, as depicted in the below Table A. 18 . Apart from that, variables namely; renting from private landlords, routine non-manual employees, and on pension, witnessed a 15% increase in their odds ratio on average. However, the odds ratios of the subsequent variables namely; female, aged 46-55 and the square root of household annual benefit income, remained fairly consistent with the ones of the previous model.

Finally, the odds ratios of divorced and household size decreased by 1.84 and 5.185, correspondingly. This could be attributed to an increase in the proportion of households with non-dependent children, where a divorce would result other family types other than lone parent such as single non-elderly.

**Transition after 3 years**

The analysis of the regression model outlining the transition probabilities to a lone parent household in the next 3 years has suggested that the model's independent factors account for 23.1% of the variation in the dependent variable. In addition to that, the examination of the model's coefficient permitted to retrieve the subsequent points.

First, unlike the previous models, household types variables, except couple with children, were not significant. The same applies to other variables, including rented from private landlords, and routine non-manual. As indicated above, the inconsistency of significant socio-economic status variables across models could be linked to the job instability of this household type. The insignificance of "privately rented" variable; however, could be related to the fact that lone parents do not stay longer in privately rented accommodation waiting for their social housing application decisions since they are classed as a high priority category. However, due to the lack of evidence to support this claim, more research is needed in this respect. In contrast to the above, the variables namely; living in terraced houses, farmers /smallholders, and renting from local authorities were found significant in this model. More specifically, the odds ratio for becoming lone parent households in three years were 2.240, 0.00462, and 0.275 for households living in terraced dwellings, renting their dwellings from local authorities, and the ones belonging to farmers/smallholders socio-economic class, respectively. The former result reinforces the argument made in the previous model with regards the high proportion of lone parents living in terraced houses. Conversely, the latter outcome was expected since almost 70% of farmers and smallholders in the UK were married according to the UK report on farm household (DEFRA, 2012); (GOVUK, 2002). Finally, although around 23% of lone parents' households in the UK rented their accommodations from local authorities, the proportion of single non-elderly with similar tenure mode was almost double (40%) (GOVUK, 2012). Hence, the lower odds ratio considering that the initial sample in this research contained 100% of single non-elderly households.

Secondly, despite some slight changes, the odds ratios of the following variables namely; divorced, household size, on pension, and square root annual income benefit were fairly consistent with the odds of the previous model. Conversely, there was a decrease of 2.45 and 2.551 in the odds ratio of the subsequent variables; aged 46-55 and living in end terraced houses, correspondingly. An explanation for the former odds ratio changes could be based on the fact that a considerable proportion of households ageing between 46-55 already moved to lone parent households, hence the lower odds ratio. Similarly, the high likelihood of household living in terraced houses could have impacted the odds ratio for terraced houses. Another possibility is that some end-terraced dwellers moved to other dwelling types, including semi-detached houses.

Table A. 18. The accommodation profile of different UK household types (ONS, 2011b).

| Great Britain | | | | | | Percentages |
|---|---|---|---|---|---|---|
| | House or bungalow | | | Flat or maisonette | | |
| | Detached | Semi-detached | Terraced | Purpose-built | Other[1] | All dwellings[2] |
| **One person households** | | | | | | |
| Under state pension age[3] | 11 | 22 | 31 | 28 | 9 | 100 |
| Over state pension age[3] | 18 | 28 | 25 | 25 | 3 | 100 |
| **One family households** | | | | | | |
| Couple[4] | | | | | | |
| No children | 31 | 31 | 24 | 10 | 3 | 100 |
| Dependent children[5] | 28 | 36 | 28 | 7 | 1 | 100 |
| Non-dependent children only | 31 | 36 | 28 | 5 | 0 | 100 |
| Lone parent[4] | | | | | | |
| Dependent children[5] | 8 | 30 | 38 | 20 | 4 | 100 |
| Non-dependent children only | 13 | 40 | 37 | 8 | 1 | 100 |
| **Other households**[6] | 19 | 27 | 34 | 14 | 6 | 100 |
| **All households**[7] | 23 | 31 | 28 | 15 | 3 | 100 |

1 Includes converted flats, part of a house and rooms.
2 Includes other types of accommodation, such as mobile homes.
3 State pension age is currently 65 for men and 60 for women.
4 Other individuals who were not family members may also be included.
5 See Appendix, Part 2: Families. May also include non-dependent children.
6 Comprising two or more unrelated adults or two or more families.
7 Includes a small number of same-sex couples.

*Source: General Household Survey (Longitudinal), Office for National Statistics*

## Transition after 4 years

The examination of the model showing the transition probabilities to lone parent households after 4 years has advised that the model's independent factors cause the dependent factor to vary by 19.6%. In addition to that, the analysis of the model suggested the subsequent conclusions.

First, it is clear that the majority of significant variables had a positive effect on the dependent variable. For instance, the odds for being lone parents in the next 4 years was higher by a factor of 4.334, 3.959, and 4.541, for divorced householders, the ones living in terraced and end-terraced dwellings, correspondingly. Similarly, for one additional member in the household, the odds ratio was higher by 3.004. However, this model has witnessed a change in the direction of significance of the variable single non-elderly, which had an odds ratio of 4.061. This was logical and consistent with an official source suggesting that around 54% of lone parents in the UK had a never married marital status (ONS, 2015e). Apart from that, the variable living in purpose-built flat was found significant with an odds ratio of 12.02. This could be because around 40% of lone parent households live in accommodations rented from the social sector (local authorities and housing associations) in which terraced houses and flats constitute a considerable proportion (ONS, 2015h). Finally, the square root of annual benefit income had a negligible effect since a one-unit increase in this variable leads to a 1.4% lower odds ratio.

In contrast to the above, the odds ratio for being a lone parent in the 4 years was lower by 87.5% for household living on a pension. Conversely, variables namely; couple with children, rented from private landlords, rented from local authorities, farmers/smallholders, working full-time, and age variables, were not significant. This suggests that there is not enough evidence to draw conclusions on the nature of their effect on the dependent variable.

Table A. 19.Logistics regression models of transition to Lone parent households in the next 2, 3, and 4 years (Part1)

| | | (Year 2)<br>Lone parent | (Year 3)<br>Lone parent | (Year 4)<br>Lone parent |
|---|---|---|---|---|
| **Accommodation related variables** | Renting from private landlords | $8.558^{**}$ | 1.034 | 4.234 |
| | | (5.826) | (0.7122) | (3.873) |
| | Dwelling owned with mortgage | 2.536 | - | 0.478 |
| | | (1.219) | | (0.327) |
| | Renting from local authorities | - | $0.275^{*}$ | 1.258 |
| | | | (0.169) | (1.1967) |
| | Living in end-terraced dwellings | $5.469^{*}$ | $2.918^{*}$ | $3.959^{*}$ |
| | | (3.799) | (1.865) | (3.334) |
| | Living in terraced dwellings | 2.089 | $2.240^{*}$ | $4.541^{*}$ |
| | | (1.046) | (1.106) | (3.208) |
| | Living in purpose built flats | - | - | $12.02^{**}$ |
| | | | | (10.16) |

Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table A. 20.Logistics regression models of transition to Lone parent households in the next 2, 3, and 4 years (Part2)

| | | (Year 2) Lone parent | (Year 3) Lone parent | (Year 4) Lone parent |
|---|---|---|---|---|
| Household type, gender and marital status | Single non-elderly | 0.206** | 0.882 | 4.061* |
| | | (0.116) | (0.450) | (2.936) |
| | Couple without children | 0.0283*** | 0.354 | |
| | | (0.0284) | (0.278) | - |
| | Couple with children | 0.0308*** | 0.0117*** | 0.120 |
| | | (0.0258) | (0.0122) | (0.142) |
| | Female | 3.47* | 3.07* | - |
| | | (2.420) | (2.23) | |
| | Divorced | 2.567* | 3.158* | 4.334* |
| | | (1.529) | (1.806) | (4.064) |
| Householder age and household size variables | Aged 26-35 | - | - | 0.342 |
| | | | | (0.251) |
| | Aged 36-45 | 1.222 | - | 1.212 |
| | | (0.610) | | (1.39) |
| | Aged 46-55 | 5.345** | 2.896** | 1.015 |
| | | (2.819) | (1.165) | (0.7341) |
| | Household size | 2.548** | 3.483*** | 3.004* |
| | | (0.790) | (1.057) | (1.457) |

Note: standard errors in parentheses

*Significance at the 95% level **Significance at the 99% level

Table A. 21.Logistics regression models of transition to Lone parent households in the next 2, 3, and 4 years (Part3)

| | | (Year 2) Lone parent | (Year 3) Lone parent | (Year 4) Lone parent |
|---|---|---|---|---|
| Socio-economic class variables | Higher-grade professionals | - | 0.370 (0.238) | - |
| | Routine non-manual employees | 5.124* (3.410) | 1.9120 (1.24) | 0.218 (0.268) |
| | Personal service workers | 6.713 (7.406) | - | - |
| | Farmers/small holders | - | 0.00462** (0.00820) | 0.0756 (0.126) |
| | Foreman technicians | 6.181 (6.337) | 3.777 (3.482) | 19.22* (25.03) |
| | Skilled manual workers | 0.174 (0.207) | - | - |
| Income related variables | On Pension | 0.202* (0.152) | 0.111** (0.0763) | 0.125* (0.158) |
| | Receiving benefit | 0.443 (0.257) | 2.692 (1.378) | - |
| | Square root of total annual benefit income | 1.030*** (0.00821) | 1.012* (0.00556) | 1.014* (0.00786) |
| | Square root of annual investment income | 0.988 (0.00875) | - | - |
| | Working full-time | - | - | 0.125 (0.158) |
| | McKelvey & Zavoina's R2/ Mc Fadden's R2 | 0.268 | 0.231 | 0.197 |
| | Type of the used model | Random effect | Random effect | Fixed- effect |
| | Sample size | 4597 | 4128 | 300 |

Note: standard errors in parentheses

*Significance at the 95% level **Significance at the 99% level

## Transition after 6 years

The examination of the logistic regression model pertaining to the transition probabilities to a lone parent family after 6 years has suggested that the model's independent variables explain only 21.2% of the variation in the dependent factors. In addition to that, this analysis also advised the following points.

First, it is evident that the majority of statistically significant variables in this model had a positive effect on the dependent variable. Furthermore, most of these factors were prone to changes in their odds ratios. For example, for single non-elderly, the odds of becoming a lone parent household over 6 years was higher by a factor of 5.945. This was 2.209 higher than the odds ratio of the precedent model. This implies that there might be a positive relationship between time and the likelihood of single non-elderly to become a lone parent household. However, the reason behind this relationship needs to be further investigated through qualitative research methods. Similarly, the odds ratio for divorced householders almost halved in comparison to the last model. As argued previously, this could because this variable's odds ratio is sensitive to changes in the divorce rates over time. On the other hand, the odds of household size decreased by almost 4.5 times from the model of year 1. This means that the effect of one additional person in the household decreases with time. Indeed, the recent UK government report on divorces suggested that the chance of divorce decreases dramatically after the $5^{th}$ year of marriage anniversary, as depicted below in Figure A. 54 (ONS, 2013b). The odds ratio of household living in purpose-built flats (4.46) and females (3.175) remained fairly consistent with the previous model. Conversely, the variables namely; dwelling size and aged 26-35, were found significant since the year 1 model. In particular, for householders aged between 26-35, the odds of being lone parent households in 6 years was higher by a factor of 5.413. This aligns with findings of the previous models suggesting that half of divorces were amongst middle-aged householders. Similarly, for one additional room in the dwelling, the odds ratio is higher by 53%.

In contrast to the above, the odds ratios were lower by 87.6% and 72.8% for households renting from local authorities and high-grade professionals, respectively. However, variables namely; terraced, end terraced, routine non-manual, farmers-small holders, on pension in addition to income variables, were not found statistically significant.

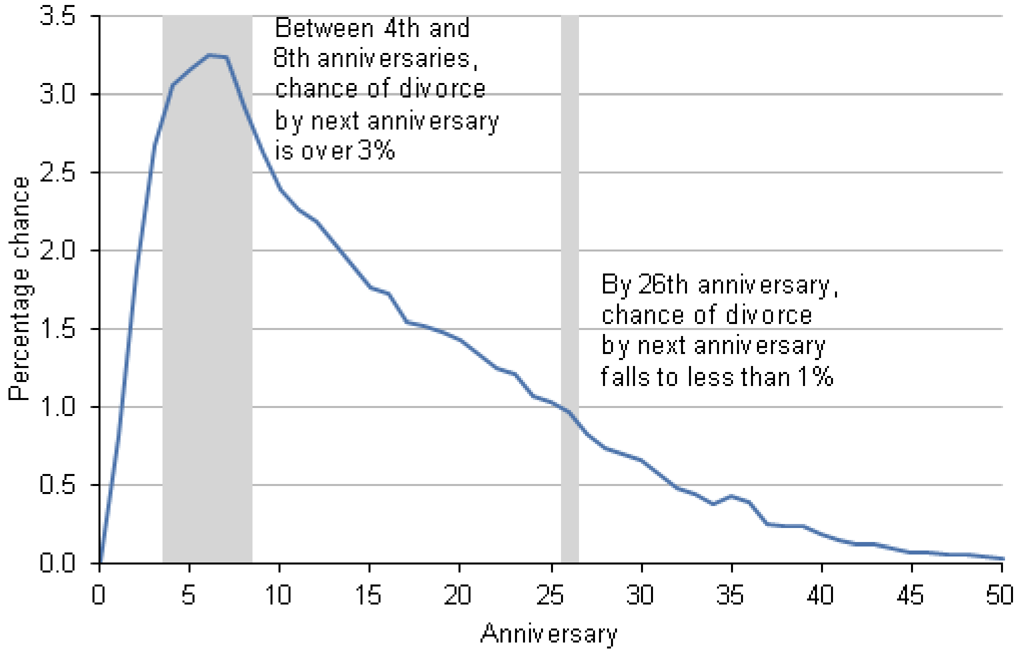



Figure A. 54. Illustrates the relationship between the length of marriage and divorce occurrence (ONS, 2013b)

### Transition after 7 years

After analysing the logistic regression model depicting the transition possibilities to a lone parent family after 7 years, it is evident that the model's independent factors are responsible for 18.7% of the variation in the dependent variable. In addition to that, the examination of the model's coefficients led to the below conclusions.

First, it is clear that there were not many changes to report from the last model except the non-significance the following variables namely; divorced and living in purpose-built flats. Moreover, the significance of the variable "Living in semi-detached dwellings" with an odds ratio of 4.042. This reinforces the argument made at an earlier stage with regards lone parents households remaining in privately rented houses before moving to social houses including; semi-detached dwellings. Indeed, based on the above table, it is clear that semi-detached was the second most common dwelling type for lone parents with dependent children after terraced houses (ONS, 2011b).

Apart from the above, the variables; single non-elderly, female, aged 26-35, household and dwelling size, remained positively significant in comparison to the previous model. However, some were prone to significant changes in their odds ratios. For instance, the odds ratio for being lone parents in seven years was higher by a factor of 6.498 for females, which is almost the double of the previous model's odds ratio. Considering the variation in the odds ratio of this variable across all the previous models, it could be argued that females are more likely to become lone parents with time. This possibility becomes even higher after taking account of other key factors, including the average marriage length in the UK (8.9 -12.2 years). Similarly, the odds ratio of the variable aged 26-35 was 1.5 times higher than in the precedent model. This was excepted since most divorces occur at an age 40 and 55 as indicated previously. In addition to the above, the odds ratio was 7.619 higher for households living end-terraced houses, which was consistent with the results of the models of year 1 to 5. Finally, the variable; rented from local authorities, had a negative impact on the dependent variable with an odds ratio lower by 81.8%.

### Transition after 8 years

The examination of the logistic regression model representing the transition probabilities to a lone parent family after 8 years has suggested that the model's independent variables are responsible for 15.4% of the variation in the dependent factor. The analysis of this transition model has also permitted to obtain the subsequent findings.

First, it is clear that the majority of the statistically significant variables in this model had a positive effect on the dependent variable. Secondly, unlike in the previous model, factors namely; household size and living in semi-detached dwellings, were not found significant in the actual model. However, skilled-manual workers, higher-grade professionals, and working full-time, were found significant. More specifically, the odds of being a lone parent in 8 years was 8.078 higher for householders belonging to skilled manual workers socio-economic class. This was in line with the findings of the model of year 1, the 2011 UK census report (ONS, 2011c), and other studies, such as the work of workers Hall, (2006). On the other hand, for households on a full-time employment and the ones with higher-grade professional socio-economic status, the odds of being lone parents in 8 years were lower by 67.16% and 88.5%, correspondingly. In this respect, it could be argued that the reason behind the lower odds ratio for householders working full-time is due to the fact that a considerable proportion of lone parents in the UK (around 56.2% in 1996 and 38% in 2015), are unemployed (ONS, 2013c). Furthermore, around 60% of those who are employed are working part-time (ONS, 2015h).

Apart from the above, the variables; single non-elderly, female, aged-26-35, renting from local authorities, and dwelling size, remained significant in comparison to the previous models. Furthermore, the direction of their association with the dependent variable was also consistent since their odds ratios were 2.803, 3.360, 5.975, 0.152, 1.904, respectively.

### Transition after 9 years

The analysis of the logistic regression model representing the transition possibilities to a lone parent family after 9 years has advised that the independent variables cause the dependent factor to vary by 13.2%. In addition to that, the examination of the model's coefficients advised the following changes.

First, the variables; single non-elderly and dwelling size, were not found significant in the actual model, whereas the factors; aged 35-45, aged 46-55, small proprietors with employees, and on pension were significant. In particular, the odds ratios were lower by 76.3% and 83.8% for householders aged 36-45 and those who are 46-55, respectively. Similarly, for households belonging to the small proprietors with employees' socio-economic class, the odds ratio was higher by a factor 23.31. This could be attributed to the fact that more than 20% of lone parents are small employers and own account workers (ONS, 2013c). Conversely, the likelihood of being lone parents was lower for pensioners since the odds ratio was 0.156, which was consistent with the models of year 1-5.

Apart from that, the following variables namely; rented from local authorities, higher-grade professionals, female, skilled manual-workers, and working full-time, were consistent with the previous models and had the subsequent odds ratios; 0.118, 0.219, 3.658, 36.48, and 0.415, correspondingly.

Table A. 22.Logistics regression models of transition to Lone parent households in the next 6, 7, 8, and 9 years (Part1)

| | | (Year 6) Lone parent | (Year 7) Lone parent | (Year 8) Lone parent | (Year 9) Lone parent |
|---|---|---|---|---|---|
| Household type and marital status variables | Single non-elderly | 5.945** | 4.289* | 2.803* | 3.058 |
| | | (3.621) | (2.785) | (1.574) | (1.999) |
| | Couple with children | 0.221 | 2.915 | - | - |
| | | (0.234) | (3.195) | | |
| | Female | 3.175* | 6.498* | 3.360* | 3.658* |
| | | (2.180) | (5.102) | (2.686) | (3.757) |
| | Divorced | 11.84*** | 0.948 | - | - |
| | | (7.942) | (0.705) | | |
| | Aged 26-35 | 5.413** | 7.929** | 5.975* | - |
| | | (3.458) | (5.405) | (4.221) | |
| Householder age and household size variables | Aged 36-45 | - | - | - | 0.237* |
| | | | | | (0.168) |
| | Aged 46-55 | - | - | - | 0.162* |
| | | | | | (0.128) |
| | Household size | 1.712* | 1.301* | - | - |
| | | (0.709) | (0.151) | | |

Note: standard errors in parentheses

*Significance at the 95% level **Significance at the 99% level

Table A. 23.Logistics regression models of transition to Lone parent households in the next 6, 7, 8, and 9 years (Part2)

| | | (Year 6) | (Year 7) | (Year 8) | (Year 9) |
|---|---|---|---|---|---|
| | | Lone parent | Lone parent | Lone parent | Lone parent |
| Accommodation related variables | Renting from private landlords | - | - | 0.139 | - |
| | | | | (0.193) | |
| | Renting from local authorities | 0.124** | 0.182* | 0.152* | 0.118* |
| | | (0.0992) | (0.158) | (0.142) | (0.135) |
| | Rented from housing associations | 0.192 | - | - | - |
| | | (0.194) | | | |
| | Living in semi-detached houses | - | 4.042* | 1.677 | - |
| | | | (2.394) | (0.915) | |
| | Living in end-terraced dwellings | 3.359 | 7.619* | - | - |
| | | (2.741) | (7.006) | | |
| | Living in terraced dwellings | 1.382 | - | - | - |
| | | (0.911) | | | |
| | Living in purpose built flats | 4.464* | - | 2.408 | 0.111 |
| | | (3.428) | | (2.316) | (0.144) |
| | Dwelling size | 1.530* | 1.904*** | 1.767** | - |
| | | (0.274) | (0.370) | (0.356) | |
| | Square root monthly rent | | 0.965 | | |
| | | | (0.0325) | | |

Note: standard errors in parentheses

*Significance at the 95% level **Significance at the 99% level

Table A. 24.Logistics regression models of transition to Lone parent households in the next 6, 7, 8, and 9 years (Part3)

| | | (Year 6) Lone parent | (Year 7) Lone parent | (Year 8) Lone parent | (Year 9) Lone parent |
|---|---|---|---|---|---|
| Socio-economic class variables | Higher-grade professionals | 0.272* | 0.545 | 0.115* | 0.219* |
| | | (0.205) | (0.426) | (0.129) | (0.224) |
| | Routine non-manual employees | 0.522 | - | - | - |
| | | (0.447) | | | |
| | Small proprietors with employees | - | 3.119 | 5.036 | 23.31* |
| | | | (3.259) | (5.403) | (32.44) |
| | Farmers/small holders | 0.104 | - | - | - |
| | | (0.167) | | | |
| | Foreman technicians | - | 0.336 | - | - |
| | | | (0.375) | | |
| | Skilled manual workers | - | 5.926 | 8.078* | 36.48** |
| | | | (5.694) | (8.034) | (44.14) |
| Income related variables | On Pension | 0.697 | 0.642 | 0.742 | 0.156* |
| | | (0.504) | (0.520) | (0.459) | (0.147) |
| | Square root of annual investment income | 0.982 | 0.986 | 0.991 | - |
| | | (0.0108) | (0.0114) | (0.00999) | |
| | Working full-time | - | - | 0.3284* | 0.415* |
| | | | | (0.18815) | (0.269) |
| | McKelvey & Zavoina's R2 | 0.212 | 0.187 | 0.154 | 0.132 |
| | Type of the used model | Random effect | Random effect | Random effect | Random effect |
| | Sample size | 3235 | 3237 | 2867 | 2605 |

Note: standard errors in parentheses

*Significance at the 95% level **Significance at the 99% level

#### A.3.3.4 Couple with children transitions

This section intends to examine the logistic regression models representing household transitions to couple with children households in the next 2, 3, 4, 6, 7, 8, and 9 years. These models are depicted below in Table A. 25, Table A. 26, Table A. 27, Table A. 28, Table A. 29, and Table A. 30.

**Transition after 2 years**

The examination of the logistic regression model outlining the transition possibilities to a couple with children family after two years has advised that the model's independent factors are responsible for 49.5% of the variation in the dependent variable. In addition to that, this analysis also suggested the subsequent changes.

First, unlike the precedent model, the variables; lone parent, higher-grade professionals, routine non-manual workers, and on benefit, were statistically significant. On the other hand, lower-grade professionals, semi-unskilled, foreman/technicians, farmers/ smallholders, and rented from private landlord, were insignificant variables.

In the light of the above, the odds ratios for becoming a couple with children in two years were higher by a factor of 7.513 and 2.393 for lone parent households and families receiving state benefit, correspondingly. The former odds ratio could be attributed to the fact that a percentage of separated householders could be reunited after a short period. Indeed, a recent study, published by the independent, suggested that more than 50% of separated couples in the UK reunited within a period of 5 years (Independent, 2016). As for the high odds ratio of households on benefit, this is because couples with children in the UK, except for those with an annual income beyond £50,000, are fully entitled to child benefit income; not to mention other benefits (GOVUK, 2017a). Similarly, for higher-grade professionals and routine non-manual workers, the odds were 1.743 and 3.302, respectively. These findings were excepted given the financial stability of these socio-economic classes. Indeed, a recent report from the UK government suggests that around 22% of the UK couples with children belong to managerial and professional classes (ONS, 2013c).

In contrast to the above, the odds ratios of remaining variables were consistent with the ones of the previous model. For example, the odds ratios were 2.253, 2.807, 3.708, and 7.799 for households aged 46-55, household members other than the main householder, householders living in detached dwellings, the ones who owned their dwelling outright, respectively. On the other hand, for the ones aged between 26-35, cohabiting couples, divorced householders, and never married, the odds were lower by 85.3%, 87.1%, 99.55%, and 99.7%, correspondingly. Finally, for households living in terraced houses, purpose-built, and converted flats, the odds were lower by a factor of 0.241, 0.143, and 0.0324, respectively.

### Transition after 3 years

The analysis of the model showing the transition probabilities to a couple with children household in 3 years has advised that the model's independent factors cause the dependent variable to vary by 45%. Moreover, the examination of the model's coefficients suggested the following points.

First, it is evident that the variables; not main householder and living in terraced houses, were not significant in the actual model. Conversely, the following factors namely; semi-detached, farmers-smallholders, and agriculture workers, were significant. Finally, there were some changes in the magnitude of certain variables' odds ratios such as never married.

In line with the above, the odds ratios were higher by 3.792 and 10.39 for households living in semi-detached dwellings and the ones belonging to farmers/smallholders socio-economic class, respectively. The former odds ratio was consistent with a UK government report (Table A. 18) suggesting that the majority of UK couples with children (36%) live in semi-detached houses (ONS, 2013d). Similarly, the odds ratio of the variable "farmers/smallholders" was in line with the findings of year 3 model of transition to lone parents. Moreover, the UK report on farm households advising that around 70% of them were married (DEFRA, 2012; GOVUK, 2002). As for the lower odds ratio of agriculture workers (0.0562), it could be due to the fact that a considerable proportion of this class is composed of foreign labour, who mostly reside in other European countries (BBC, 2016b). Indeed, the proportion of British working in the agriculture sector was 1% in 2011 (ONS, 2013a).

In contrast to the above, the odds ratio of never married was 6.7 times higher than in the previous model, which implies that single non-elderly households are more likely to become couples with children with time. Similarly, the odds ratio of living in detached dwellings almost double from the last model. Finally, the odds for higher-grade increase by 1.453 which signifies that the likelihood of being a couple with children for this group increases with time.

Finally, the odds ratios of the remaining age, marital status, socio-economic class, and income variables, stayed fairly consisted with the previous model.

### Transition after 4 years

In comparison to the previous model, the examination of the logistic regression model representing the transition possibilities to couples with children after 4 years has indicated that the model's independent factors are responsible for 43% of the variation in the dependent variable. The analysis of this transition model also advised the following points.

First, the subsequent variable namely; lone parent, living in converted flats, agriculture workers, and on benefit, were not found significant. However, the variables; "Rented from local authorities" and "other households" were found statistically significant. More precisely, for individuals living in non-family arrangements, the odds ratio for becoming a couple with children family was higher by a factor of 11.82. This was consistent with the work of Stone et al. (2011) and Daly (2005), suggesting that these households are more likely to move to couple without children households. As a result, they should make transitions to couples with children after some time. Like in the previous models, for householders aged between 46 and 55, the odds ratio was higher by a factor of 4.575. Similarly, the odds ratios for the following dwelling types namely; detached and semi-detached, were 4.812 and 2.205, respectively. Additionally, householders belonging to the following socio-economic classes; higher-grade professionals, lower-grade professionals, routine-non manual, small proprietors without employees, and farmers/smallholders, were more likely to be couple with children families in the next 4 years. The high proportion of self-employed householders living in couple with children families (around 1 million), is the reason behind the high odds ratio of small proprietors with employees (CitizenAdvice, 2015). Finally, the square root of benefit income was significant but its impact was small as for a one-unit increase in this variable, the odds ratio is higher by 2.5%.

In contrast to the above, the odds ratio for those who rented their accommodation from local authorities was lower by 0.0874. This finding was in a good agreement with the UK official report on home ownership and renting (ONS, 2013d). Similarly, the odds ratios of the subsequent dwelling type factors namely; terraced and purpose built flat, were 0.217 and 0.140, correspondingly. The same applies to the variables; cohabiting couples, divorced, and never married, in which the odds ratios were lower by 83.1%, 97.41%, and 87.6%, respectively. Finally, for householders aged between 26 and 35, the odds of becoming a couple with children in 4 years was lower by a factor of 85%.

Finally, the odds ratios of the square root of household annual benefit income and of annual investment income remained fairly consistent with the previous models.

Table A. 25.Logistics regression models of transition to Couple with children households in the next 2, 3, and 4 years (Part1)

| | | (Year 2) Couple CH2 | (Year 3) Couple CH3 | (Year 4) Couple CH4 |
|---|---|---|---|---|
| Household type and householder age | Lone parent | 7.513* | 6.995* | 6.412 |
| | | (6.650) | (6.349) | (6.206) |
| | Other households | - | 4.956 | 11.82** |
| | | | (4.050) | (10.46) |
| | Aged 26-35 | 0.147** | 0.146** | 0.150** |
| | | (0.0435) | (0.0457) | (0.0476) |
| | Aged 46-55 | 2.253* | 2.804** | 4.575** |
| | | (0.865) | (1.035) | (1.709) |
| Marital status and existence of a household reference person | Cohabiting couples | 0.129** | 0.145** | 0.169** |
| | | (0.0576) | (0.0651) | (0.0862) |
| | Divorced | 0.00447** | 0.00910** | 0.0251** |
| | | (0.00484) | (0.00892) | (0.0255) |
| | Never Married | 0.00264** | 0.0163** | 0.124** |
| | | (0.00203) | (0.0127) | (0.0908) |
| | Not main householder | 2.807** | 1.822 | 1.582 |
| | | (1.102) | (0.711) | (0.640) |

Couple CH: couples with children / Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table A. 26.Logistics regression models of transition to Couple with children households in the next 2, 3, and 4 years (Part2)

| | | (Year 2) Couple CH2 | (Year 3) Couple CH3 | (Year 4) Couple CH4 |
|---|---|---|---|---|
| Accommodation related variables | Living in detached dwellings | 3.708** | 6.149** | 4.812** |
| | | (1.814) | (2.983) | (2.445) |
| | Living is semi-detached dwellings | 1.694 | 3.792** | 2.205* |
| | | (0.714) | (1.711) | (1.048) |
| | Living in Terraced dwellings | 0.241** | 0.576 | 0.217** |
| | | (0.125) | (0.292) | (0.116) |
| | Living in Purpose built flats | 0.143** | 0.233* | 0.140* |
| | | (0.0869) | (0.166) | (0.121) |
| | Living in converted flats | 0.0324** | 0.0322* | 0.391 |
| | | (0.0410) | (0.0473) | (0.491) |
| | Dwelling owned outright | 7.799** | 2.690* | 4.535** |
| | | (4.171) | (1.313) | (2.467) |
| | Dwelling Rented from employer | 0.0757* | - | - |
| | | (0.0978) | | |
| | Dwelling rented from local authorities | - | - | 0.0874* |
| | | | | (0.0906) |
| | Dwelling rented from private landlords | 0.342 | - | - |
| | | (0.204) | | |

Couple CH: couples with children / Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table A. 27.Logistics regression models of transition to Couple with children households in the next 2, 3, and 4 years (Part3)

| | | (Year 2) Couple CH2 | (Year 3) Couple CH3 | (Year 4) Couple CH4 |
|---|---|---|---|---|
| Socio-economic class variables | Higher-grade professionals | 1.743* | 3.196** | 2.724* |
| | | (0.574) | (1.263) | (1.136) |
| | Lower-grade professionals | - | 3.031** | 5.117** |
| | | | (1.208) | (2.088) |
| | Routine non-manual employees | 3.302* | 2.33** | 4.62** |
| | | (1.765) | (1.26) | (2.036) |
| | Small proprietors with employees | 2.412 | 3.035 | - |
| | | (1.877) | (2.296) | |
| | Small proprietors without employees | - | 9.822** | 16.11** |
| | | | (7.004) | (12.56) |
| | Skilled manual workers | 2.806 | 2.612 | - |
| | | (1.932) | (1.677) | |
| | Foreman technicians | - | - | 2.982 |
| | | | | (1.731) |
| | Farmers/smallholders | - | 10.39* | 7.188* |
| | | | (10.02) | (6.543) |
| | Agriculture workers | 0.0981 | 0.0562* | |
| | | (0.188) | (0.0804) | |
| Income related variables | On benefit | 2.393** | 1.917* | 1.628 |
| | | (0.784) | (0.635) | (0.524) |
| | Square root of annual benefit income | 1.012* | 1.032** | 1.035** |
| | | (0.00714) | (0.00785) | (0.00751) |
| | Square root of annual investment income | 0.980* | 0.966** | 0.968** |
| | | (0.00832) | (0.00838) | (0.00823) |
| | McFadden's R2 | 0.495 | 0.450 | 0.430 |
| | Type of model | Fixed effects | Fixed effects | Fixed effects |
| | Sample size | 1108 | 1024 | 918 |

### Transition after 6 years

In comparison to the previous models of year 1-5, the analysis of the transition model belonging to the $6^{th}$ year of transitions to couple with children households shows that the independent factors are responsible for 28.5% of the variation in the dependent variable. This examination also indicated that the vast majority of the factors had a positive impact on the predicted variable. In addition to that, many variables were insignificant from the precedent model. These include; divorced, never married, semi-detached, owned outright, routine non-manual, small proprietors without employees, semi-unskilled manual workers, and investment income.

As for the significant variables, the odds of most of them remained fairly consistent with the previous model. For examples, for unrelated households, the odds for being a couple with children in 6 years was higher by a factor of 5.507. Similarly, the odds ratios were 2.243, 3.611, and 4.497 higher, for those living in detached dwellings and belonging to higher and lower-grade professional classes, correspondingly. The same applies to the ones working full-time as the odds ratio was higher by a factor of 2.312. However, variables namely; on a benefit, and aged 36-45 were also found significant in this model. The odds ratio of the former variable (2.678) was in line with the ones of year 1 and year 2 models. As for the odd ratio of householders aged between 36 and 45, this was consistent with the UK report on births by parents (ONS, 2015a).

In contrast to the above, the factor cohabiting couples was the only variable with a negative impact. More specifically, the odds for being couples with children for cohabiting couples in 6 years was lower by a factor of 0.396, which is almost 2.67 times higher than the odds ratio of the previous model. This, in turn, advises that the likelihood of being couples with children increases in function of the length of the relationship between cohabiting couples.

### Transition after 7 years

The investigation of the transition probabilities to couples with children after 7 years, has demonstrated that the model's independent factors cause the dependent variable to vary by 25.7%. The examination of the model's coefficients also suggested that the majority of the significant variables had a positive impact on the dependent factor. Furthermore, it has showed that variables namely; other households, lower-grade professionals, and routine non-manual workers, were not found significant. Conversely, this model has witnessed the significance of new variables, in comparison to the previous one. These include; divorced, terraced, and skilled manual workers.

In addition to the above, the remaining variables were consistent with the previous model. For example, for householders aged 36-45 and 46-55, the odds ratio for being in a couple with children household were higher by 3.744 and 5.54, respectively. Similarly, for householders who were divorced 6 years ago, the likelihood of being a couple with children in the year after was high by a factor of 4.189. This could be attributed to a possible occurrence of remarriage after divorce, which accounts for around 15% of all UK marriages (ONS, 2013e). Finally, the odds ratios were 2.761, 1.891, 3.644, and 2.881, for the variables; dwelling owned outright, higher-grade professionals, on benefit, and working full-time, correspondingly.

On the other hand, surprisingly, the odds ratio for being couples with children in 7 years, was lower for householders lodging in semi-detached dwellings by 53.5%. This could be because this dwelling type is also popular for other household structures such as couples without children, as indicated in Table A. 18 (above). However, there was not enough evidence to draw further conclusions. Thus, more research is needed in this regard. Conversely, the odds ratio for households living in terraced houses was lower by a factor of 0.215, which was in good agreement with previous models. In other words, terraced houses were mostly occupied by lone parents and single non-elderly households. Finally, for skilled manual workers, the odds for being a couple with children was lower by 77.3%. Again, this was in line with lone parent transition models and the UK 2011 census report (ONS, 2011c), suggesting that this socio-economic class is more likely to be a lone parent household.

### Transition after 8 years

The analysis of the model reflecting the transition probabilities to couple with children households in 8 years has suggested that the independent factors are responsible for 22% of the variation in the dependent variable. The examination of the model's coefficients also advised minor changes in comparison to the previous models. These comprise; the insignificance of the variables namely; dwelling owned outright, working full-time, and skilled-manual workers. Furthermore, the significance of the factor "living in end-terraced houses".

In addition to the above, the variables namely; aged 36-45, aged 46-55, divorced, higher-grade professionals, on benefit, and the square root of annual benefit income, had a positive impact on the dependent variable. In particular, the odds ratios were higher by 3.867 and 5.876, for householders aged 36-45 and 46-55, respectively. The odds ratio of the variable "divorced"; however, almost doubled from the last model to reach 10.99. This reinforces the statement made previously with regards the likelihood of remarrying after divorced. Furthermore, it was in line with the report on divorce and remarriage in England and Wales, which advises that divorces usually occur at a younger age. Thus, the chances of having remarried for those is high (Haskey, 1998). Like in the previous models, the odds ratios were 3.669 and 2.702, for higher-grade professionals and households on benefit. Finally, for a one-unit increase in the square root of annual benefit income, the odds of being couples with children after 8 years was higher by 2.3%.

On the other hand, the subsequent factors namely; cohabiting couples, living in semi-detached dwellings, living in terraced houses, and end-terraced dwellings, had a negative impact on the dependent variable. More precisely, their odds ratios were

0.435, 0.291, 0.390, and 0.163, respectively. These findings were in good agreement with the previous models and with the UK official report on home ownership and renting (ONS, 2013d).

### Transition after 9 years

The analysis of the model showing the transition possibilities to couples with children in 9 years has shown that the model's independent factors can cause the dependent variable to vary by 18.4%. The examination of the model also indicated that variables with a negative effect in the precedent model(s) were insignificant in the actual one. As a result, all statistically significant variables had a positive impact on the dependent factor.

In more details, for householders aged 36-45 and 46-55, the odds ratio for being a couple with children family in 9 years were higher by a factor 2.265 and 4.987, correspondingly. Similarly, the odds ratios were 3.448 and 2.398, for higher-grade professionals and for those receiving benefit, respectively. Additionally, for a one-unit increase in the square root of annual benefit income, the possibility of living in a couple with children arrangement was higher by 2.8%. The variable; divorced; on the other hand, was not significant at the 5% level.

In contrast to the above, cohabiting couples, living in semi-detached dwellings, living in terraced houses, and end-terraced were not significant at both the 1% and 5% levels. This means that there was not enough evidence to determine the direction of the association between these factors and the dependent variables.

Table A. 28.Logistics regression models of transition to Couple with children households in the next 6, 7, 8, and 9 years (Part1)

| | | (Year 6) Couple CH6 | (Year 7) Couple CH7 | (Year 8) Couple CH8 | (Year 9) Couple CH9 |
|---|---|---|---|---|---|
| **Household type and householder age variables** | Lone parent | - | 3.794 | - | - |
| | | | (3.415) | | |
| | Other households | 5.507* | - | - | - |
| | | (4.675) | | | |
| | Aged 36-45 | 4.479** | 3.744* | 3.867* | 2.465* |
| | | (1.621) | (1.157) | (1.352) | (1.106) |
| | Aged 46-55 | 6.01** | 5.54** | 5.876** | 4.987** |
| | | (2.01) | (1.78) | (1.568) | (1.987) |
| **Marital status variables** | Cohabiting couples | 0.396* | 0.365* | 0.435** | 0.362 |
| | | (0.163) | (0.155) | (0.212) | (0.227) |
| | Divorced | - | 4.189* | 10.99** | 5.841 |
| | | | (2.764) | (8.946) | (5.371) |
| | Separated | 4.629 | - | - | - |
| | | (4.412) | | | |
| | Never Married | - | - | 0.235 | 0.285 |
| | | | | (0.219) | (0.246) |

Couple CH: couples with children / Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table A. 29.Logistics regression models of transition to Couple with children households in the next 6, 7, 8, and 9 years (Part2)

| | | (Year 6) Couple CH6 | (Year 7) Couple CH7 | (Year 8) Couple CH8 | (Year 9) Couple CH9 |
|---|---|---|---|---|---|
| Accommodation related variables | Living in detached dwellings | 2.243* (0.990) | - | - | - |
| | Living is semi-detached dwellings | 1.765 (0.711) | 0.465* (0.163) | 0.291** (0.139) | 0.463 (0.202) |
| | Living in Terraced dwellings | - | 0.215** (0.102) | 0.390* (0.239) | 2.455 (1.512) |
| | Living in Purpose built flats | - | 0.277 (0.185) | 0.329 (0.282) | - |
| | Living in converted flats | 2.308 (2.555) | - | - | - |
| | Living in End-terraced houses | - | - | 0.163* (0.138) | - |
| | Dwelling owned outright | 2.131 (1.150) | 2.761* (1.418) | - | - |
| | Dwelling rented from local authorities | 0.316 (0.335) | - | - | - |

Couple CH: couples with children / Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

Table A. 30.Logistics regression models of transition to Couple with children households in the next 6, 7, 8, and 9 years (Part3)

| | | (Year 6) Couple CH6 | (Year 7) Couple CH7 | (Year 8) Couple CH8 | (Year 9) Couple CH9 |
|---|---|---|---|---|---|
| Socio-economic class variables | Higher-grade professionals | 3.611** | 1.891* | 3.669** | 3.448* |
| | | (1.545) | (0.587) | (1.707) | (1.827) |
| | Lower-grade professionals | 4.497*** | - | 2.050 | 1.863 |
| | | (1.994) | | (0.834) | (0.869) |
| | Routine non-manual employees | 2.892 | 1.644 | - | - |
| | | (1.631) | (0.799) | | |
| | Small proprietors with employees | 0.515 | - | - | - |
| | | (0.494) | | | |
| | Small proprietors without employees | 2.661 | 0.506 | 0.369 | 0.406 |
| | | (1.729) | (0.288) | (0.243) | (0.318) |
| | Skilled manual workers | - | 0.227* | 0.418 | 0.178 |
| | | | (0.147) | (0.309) | (0.169) |
| | Semi-unskilled manual workers | 3.660 | 1.869 | - | - |
| | | (2.470) | (1.109) | | |
| Income related variables | On benefit | 2.678** | 3.644*** | 2.702** | 2.398* |
| | | (0.890) | (1.159) | (0.986) | (0.999) |
| | Square root of annual benefit income | 1.013** | 1.013** | 1.023*** | 1.028*** |
| | | (0.00433) | (0.00397) | (0.00517) | (0.00634) |
| | Working full-time | 2.312* | 2.881** | 1.725 | 1.256 |
| | | (0.847) | (0.963) | (0.606) | (0.522) |
| | McFadden's R2 | 0.285 | 0.257 | 0.22 | 0.184 |
| | Type of model | Fixed effects | Fixed effects | Fixed effects | Fixed effects |
| | Sample size | 780 | 777 | 659 | 580 |

Couple CH: couples with children / Note: standard errors in parentheses
*Significance at the 95% level **Significance at the 99% level

### A.3.4 Tools and techniques used in developing EvoEnergy

Unlike the old Unity Legacy GUI (graphical user interface) system, the new unity 3D GUI (graphical user interface) system is non-programing centric which means that only a fairly small amount of scripting is needed to build and govern the behaviour of a user-interface. The first step towards creating a user interface consists of adding a canvas. This latter acts as the paint board of the user-interface controls. It should be noted that the new Canvas system can be created as an overlay, in a 2D camera with a perspective, or even totally emerged in a scene as a traditional 3D object.

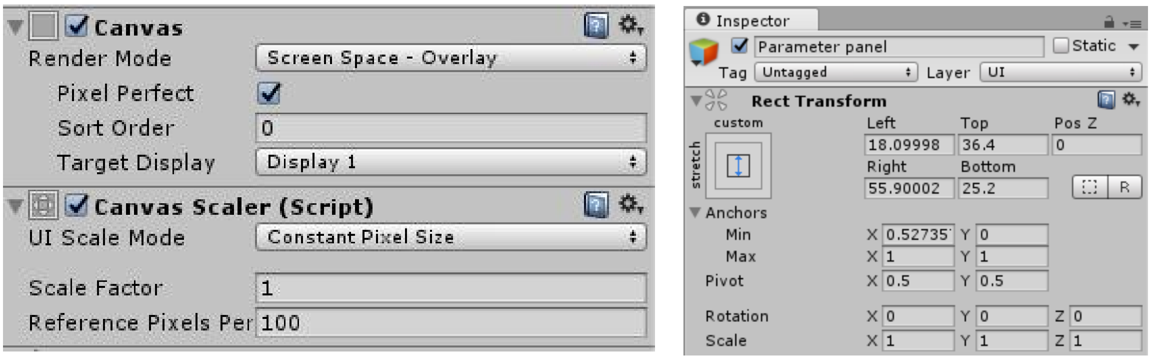
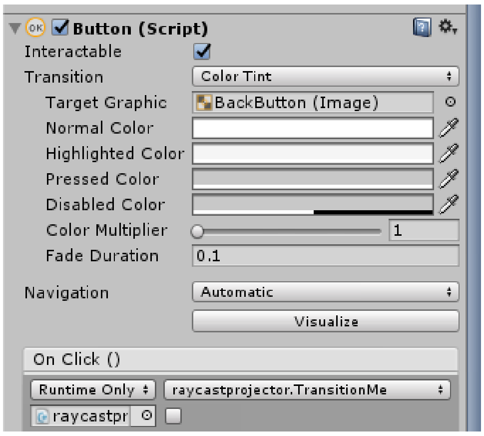



Figure A. 55.The unity canvas tab (left) and rect transform tab (right)

Once the canvas is in place, the following step comprises adding a panel, which is a highlighted background with a pre-set image that helps grouping different UI elements such as buttons, sliders, texts, and others. Depending on the design of the user-interface, the “Rect Transform” component (Figure A. 55) allows to resize, rotate, and scale different UI elements. Moreover, control their pivot and anchor positions. Finally, the last step involves defining the behaviour of each UI element using a script. For example, as depicted in Figure A. 56, a simple script was written to enable the user to return to the main menu once the back button is clicked. The script consists of a public function which returns a void and takes a Boolean value as a parameter. If the button is clicked and the Boolean value is not open, the household transition menu will be disabled, and the main menu will appear.




```
public void TransitionMe(bool Open) {
    if (Open)
    {
        TransitionMenu.gameObject.SetActive(true);
        MainMenu.gameObject.SetActive(false);
    }
    if (!Open) {

        TransitionMenu.gameObject.SetActive(false);
        MainMenu.gameObject.SetActive(true);
    }
}
```

Figure A. 56.Unity UI Button tab (left) and its script on the right

# Bibliography

1001bitStandard 2017. 1001bit.com - Architectural Tools for Sketchup. [Accessed 30 May 2017]. Available from: http://www.1001bit.com/1001bit_standard.html.

3DIS 2016. CityGML2CAD - 3DIS GmbH. [Accessed 29 May 2017]. Available from: https://www.3dis.de/downloads/downloadscitygml2cad/.

Abrahamse, W. and Steg, L. 2009. How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? *Journal of economic psychology.* **30**(5),pp.711–720.

Agami Reddy, T. 2006. Literature review on calibration of building energy simulation programs: Uses, problems, procedures, uncertainty, and tools. *ASHRAE Transactions.*,pp.226–240.

Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H. and Saidur, R. 2014. A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews.* **33**,pp.102–109.

Akay, D. and Atak, M. 2007. Grey prediction with rolling mechanism for electricity demand forecasting of Turkey. *Energy.* **32**(9),pp.1670–1675.

Alam, N., Coors, V., Zlatanova, S. and Van Oosterom, P.J.M. 2012. Shadow effect on photovoltaic potentiality analysis using 3D city models *In*: *XXII ISPRS Congress, Commission VIII, Melbourne, Australia, 25 August-1 September 2012; IAPRS XXXIX-B8, 2012.* International Society for Photogrammetry and Remote Sensing (ISPRS).

Ali, P.A.U. and Yano, K. 2004. *Eco-finance: the legal design and regulation of market-based environmental instruments.* Kluwer Law International.

Allison, P.D. 2006. Fixed effects regression methods in SAS. *Thirty-first Annual SAS.* [Online]. Available from: https://www.researchgate.net/profile/Paul_Allison/publication/228767539_Fixed_effects_regression_methods_in_SAS/links/0fcfd50e5de875b37b000000.pdf.

Amirebrahimi, S., Rajabifard, A., Mendis, P. and Ngo, T. 2015. A Data Model for Integrating GIS and BIM for Assessment and 3D Visualisation of Flood Damage to Building. *Locate.* **15**,pp.10–12.

Anderson, L., Newell, K. and Kilcoyne, J. 1999. '"SELLING BLOOD"': CHARACTERISTICS AND MOTIVATIONS OF STUDENT PLASMA DONORS. *Sociological Spectrum.* **19**(2),pp.137–162.

APA 2017. Socioeconomic Status. [Accessed 20 May 2017]. Available from: http://www.apa.org/topics/socioeconomic-status/.

Ardolino, A., Arnaud, R., Berinstein, P., Franco, S., Herubel, A., McCutchan, J., Nedelcu, N., Nitschke, B., Robinet, F., Ronchi, C., Samour, G., Turkowski, R. and Walter, R. 2014. Geometry and Models: 3D Format Conversion (FBX, COLLADA) *In*: *Game Development Tool Essentials* [Online]. Berkeley, CA: Apress, pp. 19–37. Available from: http://dx.doi.org/10.1007/978-1-4302-6701-0_3.

Arellano, M. 2003. *Panel data econometrics.* Oxford University Press.

Asadi, S., Amiri, S.S. and Mottahedi, M. 2014. On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. *Energy and Buildings.* [Online]. **85**,pp.246–255. Available from: http://www.sciencedirect.com/science/article/pii/S0378778814007154.

ASHRAE 1997. ANSI/ASHRAE standard 129-1997, measuring air-change effectiveness. *ASHRAE Standard.*

ASHRAE 2001. Fundamentals. *American Society of Heating, Refrigerating and Air Conditioning Engineers, Atlanta.* **111**.

astronavigation 2015. The Demystified Astro Navigation Course – Unit 3 Part 1 | Astro Navigation Demystified. [Accessed 30 July 2018]. Available from: https://astronavigationdemystified.com/the-demystified-astro-navigation-course-unit-3/.

Aviva 2013. *Family Finances report July 2013* [Online]. [Accessed 12 April 2017]. Available from: https://www.financialplanning.org.uk/news/2013/07/aviva-family-finances-report-july-2013-there's-never-been-better-time-talk-about-family.

AVIVA 2014. Cost of divorce reaches £44,000 for UK couples - Aviva. [Accessed 6 March 2017]. Available from: http://www.aviva.co.uk/media-centre/story/17337/cost-of-divorce-reaches-44000-for-uk-couples/.

Aydin, N.Y., Kentel, E. and Sebnem Duzgun, H. 2013. GIS-based site selection methodology for hybrid renewable energy systems: A case study from western Turkey. *Energy Conversion and Management.* [Online]. **70**,pp.90–106. Available from: http://www.sciencedirect.com/science/article/pii/S0196890413000848.

Aydinalp, M., Ismet Ugursal, V. and Fung, A.S. 2002. Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. *Applied Energy.* [Online]. **71**(2),pp.87–110. Available from: http://www.sciencedirect.com/science/article/pii/S0306261901000496.

Aydinalp, M. and Ugursal, V.I. 2008. Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. *Applied Energy.* [Online]. **85**(4),pp.271–296. Available from: http://www.sciencedirect.com/science/article/pii/S030626190600136X.

Babbie, E. 1992. The practice of social research . California: Wadsworth.

Bahrami 2016. Forced Convection Heat Transfer. [Accessed 23 July 2018]. Available from: https://www.sfu.ca/~mbahrami/ENSC 388/Notes/Forced Convection.pdf.

Bahu, J.M., Koch, A., Kremers, E. and Murshed, S.M. 2013. Towards a 3D spatial urban energy modelling approach *In*: *ISPRS Annals of the Photogrammetry, Remote Sensing* *and Spatial Information Sciences. Proceedings of the ISPRS 8th 3D GeoInfo Conference & WG II/2 Workshop.*, pp. 33–41.

Baker, K.J. and Rylatt, R.M. 2008. Improving the prediction of UK domestic energy-demand using annual consumption-data. *Applied Energy.* [Online]. **85**(6),pp.475–482. [Accessed 13 May 2017]. Available from: http://www.sciencedirect.com/science/article/pii/S030626190700147X.

Baker, P., Blundell, R. and Micklewright, J. 1989. Modelling household energy expenditures using micro-data. *The Economic Journal.* **99**(397),pp.720–738.

Ball, M. 2010. The UK private rented sector as a source of affordable accommodation. *York: Joseph Rowntree Foundation.*

Balnaves, M. and Caputi, P. 2001. *Introduction to quantitative research methods: An investigative approach.* Sage.

Bartiaux, F. and Gram-Hanssen, K. 2005. Socio-political factors influencing household electricity consumption : A comparison between Denmark and Belgium. *ECEE 2005 Summer Study.*,pp.1313–1325.

Batty, M. 2007. Planning support systems: progress, predictions, and speculations on the shape of things to come.

BBC 2016a. Fifth of UK couples close to break-up - Relate study - BBC News. [Accessed 14 February 2017]. Available from: http://www.bbc.co.uk/news/uk-36373299.

BBC 2016b. What will happen to the UK's European farm workers? - BBC News. [Accessed 9 April 2017]. Available from: http://www.bbc.co.uk/news/magazine-36656969.

Beckel, C., Sadamori, L. and Santini, S. 2012. Towards automatic classification of private households using electricity consumption data *In*: *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings.* ACM, pp. 169–176.

Bedir, M., Hasselaar, E. and Itard, L. 2013. Determinants of electricity consumption in Dutch dwellings. *Energy and buildings.* **58**,pp.194–207.

Begum, N. 2004. Employment by occupation and industry. *Labour Market Trends.* **112**(6),pp.227–234.

BEIS 2013. *Energy consumption in the United Kindgdom* [Online]. Available from: http://webarchive.nationalarchives.gov.uk/+/http:/www.berr.gov.uk/files/file11250.pdf.

BEIS 2016. *Energy consumption in the United Kingdom* [Online]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/541163/ECUK_2016.pdf.

Bellotti, E. 2014. *Qualitative networks: Mixed methods in sociological research.* Routledge.

Bennett, D.A. 2001. How can I deal with missing data in my study? *Australian and New Zealand journal of public health.* **25**(5),pp.464–469.

Benson, H. 2013. *Unmarried parents account for one fifth of couples but half of all family* *breakdown.*

Bentzen, J. and Engsted, T. 2001. A revival of the autoregressive distributed lag model in estimating energy demand relationships. *Energy.* **26**(1),pp.45–55.

Berger, P. and Luckmann, T. 1967. The social construction of reality: A treatise on the sociology of education.

Bergman, L. 2005. CGE modeling of environmental policy and resource management. *Handbook of environmental economics.* **3**,pp.1273–1306.

Berkhout, P.H.G., Ferrer-i-Carbonell, A. and Muskens, J.C. 2004. The ex post impact of an energy tax on household energy demand. *Energy Economics.* **26**(3),pp.297–317.

Van Berlo, L. 2009. CityGML extension for Building Information Modelling (BIM) and IFC. *Free and Open Source Software for Geospatial (FOSS4G), Sydney, Australia.*

Bhatia, A. 2017. Cooling Load Calculations and Principles Credit: 6 PDH. [Accessed 23 July 2018]. Available from: https://www.cedengineering.com/userfiles/Cooling Load Calculations and Principles.pdf.

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S. and Çöltekin, A. 2015. Applications of 3D city models: state of the art review. *ISPRS Int J Geo-Inf.* [Online]. **4**. Available from: http://dx.doi.org/10.3390/ijgi4042842.

Blaikie, N. 2009. *Designing social research.* Polity.

Blight, T.S. and Coley, D.A. 2013. Sensitivity analysis of the effect of occupant behaviour on the energy consumption of passive house dwellings. *Energy and Buildings.* **66**,pp.183–192.

Blundell, R., Dearden, L. and Sianesi, B. 2005. Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* **168**(3),pp.473–512.

Böhringer, C. and Rutherford, T.F. 2009. Integrated assessment of energy policies: decomposing top-down and bottom-up. *Journal of Economic Dynamics and Control.* **33**(9),pp.1648–1661.

Boyce, C. and Neale, P. 2006. *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input.* Pathfinder International Watertown, MA.

Boyes, G., Thomson, C. and Ellul, C. 2015. Integrating BIM and GIS: Exploring the use of IFC space objects and boundaries.

Bramlett, M.D. and Mosher, W.D. 2012. Cohabitation, marriage, divorce, and remarriage in the United States. *Vital health statistics.* **23**(22),pp.1–32.

BRE 2013. *Report 3: Metered fuel consumption.*

Brohmann, B., Heinzle, S., Rennings, K., Schleich, J. and Wüstenhagen, R. 2009. What's driving sustainable energy consumption? A survey of the empirical literature.

van den Brom, P., Meijer, A. and Visscher, H. 2017. Performance gaps in energy consumption:

household groups and building characteristics. *Building Research & Information.* [Online],pp.1–17. Available from: http://dx.doi.org/10.1080/09613218.2017.1312897.

Brounen, D., Kok, N. and Quigley, J.M. 2012. Residential energy use and conservation: Economics and demographics. *European Economic Review.* **56**(5),pp.931–945.

Bryman, A. 2007. Barriers to integrating quantitative and qualitative research. *Journal of mixed methods research.* **1**(1),pp.8–22.

Bryman, A. 2008. Why do researchers integrate/combine/mesh/blend/mix/merge/fuse quantitative and qualitative research. *Advances in mixed methods research.*,pp.87–100.

Bunn, P. and Rostom, M. 2015. Staff Working Paper No. 554 Household debt and spending in the United Kingdom Household debt and spending in the United Kingdom. [Accessed 15 July 2017]. Available from: www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx.

Burholt, V. and Windle, G. 2006. Keeping warm? Self-reported housing and home energy efficiency factors impacting on older people heating homes in North Wales. *Energy Policy.* **34**(10),pp.1198–1208.

Campanale, C., Fugazza, C. and Gomes, F. 2015. Life-cycle portfolio choice with liquid and illiquid financial assets. *Journal of Monetary Economics.* **71**,pp.67–83.

Campbell, J.C. and Bunting, S. 1991. Voices and paradigms: Perspectives on critical and feminist theory in nursing. *Advances in Nursing Science.* **13**(3),pp.1–15.

Canyurt, O.E., Ozturk, H.K., Hepbasli, A. and Utlu, Z. 2005. Estimating the Turkish residential–commercial energy output based on genetic algorithm (GA) approaches. *Energy Policy.* **33**(8),pp.1011–1019.

Carrión, D., Lorenz, A. and Kolbe, T.H. 2010. Estimation of the energetic rehabilitation state of buildings for the city of Berlin using a 3D city model represented in CityGML. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences.* **38**,pp.31–35.

Carter, A., Craigwell, R. and Moore, W. 2012. Price reform and household demand for electricity. *Journal of Policy Modeling.* **34**(2),pp.242–252.

Carter, B.E. and McGoldrick, M.E. 1988. *The changing family life cycle: A framework for family therapy.* Gardner Press.

Chalal, M.L. and Balbo, R. 2014. Framing Digital Tools and Techniques in Built Heritage 3D Modelling: The Problem of Level of Detail in a Simplified Environment. *International Journal of the Constructed Environment.* **4**(2).

Chambers, R. 2002. *Participatory workshops: a sourcebook of 21 sets of ideas and activities.* Earthscan.

Chatterton, T.J., Anable, J., Barnes, J. and Yeboah, G. 2016. Mapping household direct energy consumption in the United Kingdom to provide a new perspective on energy justice. *Energy Research & Social Science.* [Online]. **18**,pp.71–87. Available from: http://www.sciencedirect.com/science/article/pii/S2214629616300755.

Chaturvedi, K., Smyth, C.S., Gesquière, G., Kutzner, T. and Kolbe, T.H. 2017. Managing versions and history within semantic 3D city models for the next generation of CityGML. *Lecture Notes in Geoinformation and Cartography.*

Chen, Y., Hong, T. and Piette, M.A. 2017. City-Scale Building Retrofit Analysis: A Case Study using CityBES.

Chun-sheng, Z., Shu-Wen, N.I.U. and Xin, Z. 2012. Effects of household energy consumption on environment and its influence factors in rural and urban areas. *Energy Procedia.* **14**,pp.805–811.

CitizenAdvice 2015. *Who are the self-employed?* [Online]. [Accessed 2 April 2017]. Available from: https://www.citizensadvice.org.uk/Global/CitizensAdvice/Work Publications/Whoaretheselfemployed.pdf.

Cohen, L., Manion, L. and Morrison, K. 2013. *Research methods in education.* Routledge.

Corrado, V., Mechri, H.E. and Fabrizio, E. 2007. Building energy performance assessment through simplified models: application of the ISO 13790 quasi-steady state method *In: Proceedings of building simulation.*, pp. 79–86.

Cortes, C. and Vapnik, V. 1995. Support-vector networks. *Machine Learning.* **20**(3),pp.273–297.

Cox, S., Cuthbert, A., Daisey, P., Davidson, J., Johnson, S., Keighan, E., Lake, R., Mabrouk, M., Margoulies, S. and Martell, R. 2002. OpenGIS® Geography Markup Language (GML) Implementation Specification, version.

Cramer, J.C., Miller, N., Craig, P., Hackett, B.M., Dietz, T.M., Vine, E.L., Levine, M.D. and Kowalczyk, D.J. 1985. Social and engineering determinants and their equity implications in residential electricity use. *Energy.* **10**(12),pp.1283–1291.

Creswell, J.W. 2013. *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage publications.

Creswell, J.W. and Clark, V.L.P. 2007. Designing and conducting mixed methods research.

Creswell, J.W., Plano Clark, V.L., Gutmann, M.L. and Hanson, W.E. 2003. Advanced mixed methods research designs. *Handbook of mixed methods in social and behavioral research.*,pp.209–240.

Crotty, M. 1998. *The foundations of social research: Meaning and perspective in the research process.* Sage.

Daly, M. 2005. Gender mainstreaming in theory and practice. *Social Politics: International Studies in Gender, State & Society.* **12**(3),pp.433–450.

DECC 2015a. *Annual Fuel Poverty Statistics Report, 2015.*

DECC 2015b. *Energy Consumption in the UK (2015)* [Online]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/449102/ECUK_Chapter_1_-_Overall_factsheet.pdf.

DECC 2014a. *Energy efficient products -helping us cut energy use* [Online]. Available from:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/328083/Energy_efficient_products_-helping_us_to_cut_energy_use_-_publication_version_final.pdf.

DECC 2013. *further Analysis of Data from the Household Electricity Usage Study:Consumer Archetypes.*

DECC 2014b. *How heating controls affect domestic energy demand: A Rapid Evidence Assessment* [Online]. [Accessed 12 June 2017]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/277552/FINALHow_heating_controls_affect_domestic_energy_demand_-_A_Rapid_Evidence_Assessment.pdf.

DECC 2015c. National Energy Efficiency Data-Framework (NEED) report: summary of analysis 2015 - Publications - GOV.UK. Available from: https://www.gov.uk/government/statistics/national-energy-efficiency-data-framework-need-report-summary-of-analysis-2015.

DECC 2015d. RPI: fuel & light: electricity (Jan 1987=100) - Office for National Statistics. [Accessed 18 April 2016]. Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/dobx.

DECC 2012a. *UK Renewable Energy Roadmap* [Online]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/48128/2167-uk-renewable-energy-roadmap.pdf.

DECC 2012b. *United Kingdom Housing Energy Fact File* [Online]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachmentdata/file/201167/ukhousingactfile2012.pdf/.

DEFRA 2012. *Farm Household Income and Household Composition : Results from the Farm Business Survey , England 2012 /1 3* [Online]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/454626/fbs-householdincome201213-statsnotice-28aug14.pdf.

Delgado, F., Martínez-González, M.M. and Finat, J. 2013. An evaluation of ontology matching techniques on geospatial ontologies. *International Journal of Geographical Information Science.* **27**(12),pp.2279–2301.

Depuru, S.S.S.R., Wang, L. and Devabhaktuni, V. 2011. Smart meters for power grid: Challenges, issues, advantages and status. *Renewable and sustainable energy reviews.* **15**(6),pp.2736–2742.

Dombaycı, Ö.A. 2010. The prediction of heating energy consumption in a model house by using artificial neural networks in Denizli–Turkey. *Advances in Engineering Software.* [Online]. **41**(2),pp.141–147. Available from: http://www.sciencedirect.com/science/article/pii/S0965997809002014.

Dong, B., Cao, C. and Lee, S.E. 2005. Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings.* **37**(5),pp.545–553.

Donkers, S., Ledoux, H., Zhao, J. and Stoter, J. 2015. Automatic conversion of IFC datasets to geometrically and semantically correct CityGML LOD3 buildings. *Transactions in GIS*. [Online],p.n/a-n/a. Available from: http://dx.doi.org/10.1111/tgis.12162.

Dresner, S. and Ekins, P. 2006. Economic instruments to improve UK home energy efficiency without negative social impacts. *Fiscal Studies*. **27**(1),pp.47–74.

Druckman, A. and Jackson, T. 2008. Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model. *Energy Policy*. **36**(8),pp.3177–3192.

DTI 2005. *Annex 3B Regional and local use of energy in the domestic sector* [Online]. [Accessed 12 May 2017]. Available from: http://webarchive.nationalarchives.gov.uk/20060829152920/dti.gov.uk/files/file20328.pdf.

Du, R.Y. and Kamakura, W.A. 2006. Household Life Cycles and Lifestyles in the United States. *Journal of Marketing Research*. [Online]. **43**(1),pp.121–132. Available from: http://journals.ama.org/doi/abs/10.1509/jmkr.43.1.121.

Eakins, J. 2013. An analysis of the determinants of household energy expenditures: Empirical evidence from the Irish household budget survey. *Degree of Doctor of Philosophy in Economics from the University of Surrey, Surrey Energy Economics Centre (SEEC), School of Economics, University of Surrey*.

EEA 2013. Energy intensity — European Environment Agency. [Accessed 21 December 2016]. Available from: http://www.eea.europa.eu/data-and-maps/indicators/total-primary-energy-intensity-1.

EIA 2016. International Energy Outlook 2016-Buildings sector energy consumption - Energy Information Administration. Available from: https://www.eia.gov/forecasts/ieo/buildings.cfm.

EIA 2012. Residential Energy Consumption Survey data show decreased energy consumption per household - Today in Energy - U.S. Energy Information Administration (EIA). Available from: http://www.eia.gov/todayinenergy/detail.php?id=6570.

Ekici, B.B. and Aksoy, U.T. 2009. Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*. [Online]. **40**(5),pp.356–362. Available from: http://www.sciencedirect.com/science/article/pii/S0965997808001105.

Ekonomou, L. 2010. Greek long-term energy consumption prediction using artificial neural networks. *Energy*. **35**(2),pp.512–517.

El-Mekawy, M., Östman, A. and Shahzad, K. 2011. Towards Interoperating CityGML and IFC Building Models: A Unified Model Based Approach *In*: *Advances in 3D Geo-Information Sciences* [Online]. Springer Berlin Heidelberg, pp. 73–93. Available from: http://link.springer.com/10.1007/978-3-642-12670-3_5.

Elliott, J., Heesterbeek, S., Lukensmeyer, C.J. and Slocum, N. 2005. Participatory Methods Toolkit: A practitioner's manual. *King Baudouin Foundation and the Flemish Institute for Science and Technology Assessment (viWTA)*.

Engineeringtoolbox 2015. Radiation Heat Transfer. [Accessed 26 July 2018]. Available from: https://www.engineeringtoolbox.com/radiation-heat-transfer-d_431.html.

EngineeringToolBox 2015. Dry Bulb, Wet Bulb and Dew Point Temperatures. [Accessed 30 July 2018]. Available from: https://www.engineeringtoolbox.com/dry-wet-bulb-dew-point-air-d_682.html.

Erdogdu, E. 2007. Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey. *Energy policy.* **35**(2),pp.1129–1146.

Erhart, D.K., Bialk, S., Botero, L., Eicker, U. and Kurth, D. 2013. Development of an energy efficient refurbishment concept and new ways of energy supply for a post-war quarter in Ludwigsburg/Germany. *Development.* **26**,p.29th.

Ermoliev, Y., Makowski, M. and Marti, K. 2012. *Managing Safety of Heterogeneous Systems: Decisions Under Uncertainties and Risks.* Springer Science & Business Media.

Fawcett, T. 2016. Policy and extreme energy consumption.

Fell, D. and King, G. 2012. Domestic energy use study: to understand why comparable households use different amounts of energy. *Department for Energy and Climate Change Report.*

Finch, N. 2002. Demographic Trends in the UK. *Social Policy Research Unit, University of York: Heslington, York.*

FirstGreen 2012. Differentiate between the DNI, DHI and GHI? |. [Accessed 30 July 2018]. Available from: https://firstgreenconsulting.wordpress.com/2012/04/26/differentiate-between-the-dni-dhi-and-ghi/.

Ford-Gilboe, M., Campbell, J. and Berman, H. 1995. Stories and numbers: Coexistence without compromise. *Advances in Nursing Science.* **18**(1),pp.14–26.

Fotheringham, S. and Rogerson, P. 2013. *Spatial analysis and GIS.* CRC Press.

Foucquier, A., Robert, S., Suard, F., Stéphan, L. and Jay, A. 2013. State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews.* [Online]. **23**(0),pp.272–288. Available from: http://www.sciencedirect.com/science/article/pii/S1364032113001536.

Frederiks, E.R., Stenner, K. and Hobman, E. V 2015. The socio-demographic and psychological predictors of residential energy consumption: A comprehensive review. *Energies.* **8**(1),pp.573–609.

FreeMapTools 2017. Radius From UK Postcode. [Accessed 14 January 2017]. Available from: https://www.freemaptools.com/radius-from-uk-postcode.htm.

Freitas, H., Oliveira, M., Jenkins, M. and Popjoy, O. 1998. The Focus Group, a qualitative research method. *ISRC, Merrick School of Business, University of Baltimore (MD, EUA).* **1**,pp.1–22.

Fumo, N. 2014. A review on the basics of building energy estimation. *Renewable and Sustainable Energy Reviews.* [Online]. **31**(0),pp.53–60. Available from: http://www.sciencedirect.com/science/article/pii/S1364032113007892.

Fumo, N. and Rafe Biswas, M.A. 2015. Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews.* [Online]. **47**(0),pp.332–343. Available from: http://www.sciencedirect.com/science/article/pii/S1364032115001884.

Furlong, P. and Marsh, D. 2010. A skin not a sweater: ontology and epistemology in political science. *Theory and methods in political science.*,pp.184–211.

Gale, C.G. and Longley, P.A. 2013. Temporal uncertainty in a small area open geodemographic classification. *Transactions in GIS.* **17**(4),pp.563–588.

Galobardes, B., Shaw, M., Lawlor, D.A., Lynch, J.W. and Smith, G.D. 2006. Glossary: Indicators of socioeconomic position (Part 2). *Journal of Epidemiology and Community Health (1979-).* **60**(2),pp.95–101.

Geem, Z.W. and Roper, W.E. 2009. Energy demand estimation of South Korea using artificial neural network. *Energy policy.* **37**(10),pp.4049–4054.

Genjo, K., Tanabe, S.I., Matsumoto, S.I., Hasegawa, K.I. and Yoshino, H. 2005. Relationship between possession of electric appliances and electricity for lighting and others in Japanese households. *Energy and Buildings.* **37**(3),pp.259–272.

GeoInformationGroup 2017. The GeoInformation Group | UKSurvey. [Accessed 6 July 2017]. Available from: http://www.geoinformationgroup.co.uk/uksurvey.

Gerring, J. 2008. Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques *In*: *The Oxford handbook of political methodology.*

Ghauche, A. 2010. Integrated transportation and energy activity-based model.

Ghauri, P.N. and Grønhaug, K. 2005. *Research methods in business studies: A practical guide.* Pearson Education.

Ghiaus, C. 2006. Experimental estimation of building energy performance by robust regression. *Energy and Buildings.* [Online]. **38**(6),pp.582–587. Available from: http://www.sciencedirect.com/science/article/pii/S0378778805001799.

Gill, Z.M., Tierney, M.J., Pegg, I.M. and Allan, N. 2010. Low-energy dwellings: the contribution of behaviours to actual performance. *Building Research & Information.* **38**(5),pp.491–508.

Gillingham, K., Newell, R.G. and Pizer, W.A. 2008. Modeling endogenous technological change for climate policy analysis. *Energy Economics.* **30**(6),pp.2734–2753.

Goodman, A. and Greaves, E. 2010. *Cohabitation , marriage and child outcomes* [Online]. Available from: http://www.ifs.org.uk/comms/comm114.pdf.

GOVUK 2015. *2010 to 2015 government policy: housing for older and vulnerable people - GOV.UK* [Online]. [Accessed 2 March 2017]. Available from: https://www.gov.uk/government/publications/2010-to-2015-government-policy-housing-for-older-and-vulnerable-people/2010-to-2015-government-policy-housing-for-older-and-vulnerable-people.

GOVUK 2017a. Child Tax Credit - GOV.UK. [Accessed 8 April 2017]. Available from: https://www.gov.uk/child-tax-credit/overview.

GOVUK 2017b. *Divorce and pensions - Pension Wise* [Online]. [Accessed 9 March 2017]. Available from: https://www.pensionwise.gov.uk/divorce.

GOVUK 2017c. Housing Benefit - GOV.UK. [Accessed 19 March 2017]. Available from: https://www.gov.uk/housing-benefit/eligibility.

GOVUK 2017d. *Simplified expenses if you're self-employed - GOV.UK* [Online]. [Accessed 2 March 2017]. Available from: https://www.gov.uk/simpler-income-tax-simplified-expenses/living-at-your-business-premises.

GOVUK 2002. *Status, Percentage distribution of farmers by marital* [Online]. Available from: https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwj73qjxluvSAhXDLMAKHTZJCj8QFgghMAA&url=http%3A%2F%2Fwww.equality.nisra.gov.uk%2FPercentage%2520distribution%2520of%2520farmers%2520by%2520marital%2520status%2520farm%2520type%2520.

GOVUK 2012. *Table 750 General needs social lettings1 made by housing associatons (HA) and local authority (LA) landlords, 2007/08 to 2011/12, England: household type* [Online]. Available from: https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwijv5aFiPTSAhWUHsAKHbURCOkQFgglMAA&url=https%3A%2F%2Fwww.gov.uk%2Fgovernment%2Fuploads%2Fsystem%2Fuploads%2Fattachment_data%2Ffile%2F39278%2FLive_Table_750.xls&usg=AFQjCNGkstGoibR.

Gram-Hanssen, K., Kofod, C. and Petersen, K.N. 2004. Different everyday lives. *Proceedings of the 2004 American Council for an Energy Efficient Economy.*

Greene, J.C., Caracelli, V.J. and Graham, W.F. 1989. Toward a conceptual framework for mixed-method evaluation designs. *Educational evaluation and policy analysis.* **11**(3),pp.255–274.

Greene, W.H. 2008. The Econometric Approach to Efficiency Analysis *In*: *The Measurement of Productive Efficiency and Productivity Change.*

Greene, W.W.H.. 2012. *Econometric analysis* [Online]. Available from: http://www.mendeley.com/research/econometric-analysis-195/%5Cnhttps://scholar.google.com/scholar?hl=en&q=econometric+analysis&btnG=&as_sdt=1,10&as_sdtp=#1.

GreenSpec 2018. Heat transfer. [Accessed 22 July 2018]. Available from: http://www.greenspec.co.uk/building-design/heat-transfer-conduction-convection-radiation/.

Greenwood, P., Sago, J., Richmond, S. and Chau, V. 2009. Using game engine technology to create real-time interactive environments to assist in planning and visual assessment for infrastructure *In*: *18th World IMACS/MODSIM Congress.*, pp. 13–17.

Gröger, G., Kolbe, T., Nagel, C. and Häfele, K.-H. 2012. OGC City Geography Markup Language (CityGML) En-coding Standard. *Ogc.* [Online],pp.1–344. Available from: https://portal.opengeospatial.org/files/?artifact_id=47842.

Gröger, G. and Plümer, L. 2012. CityGML – Interoperable semantic 3D city models. *ISPRS*

*Journal of Photogrammetry and Remote Sensing.* [Online]. **71**,pp.12–33. Available from: http://www.sciencedirect.com/science/article/pii/S0924271612000779.

de Groot, E., Spiekman, M. and Opstelten, I. 2006. 361: Dutch research into user behaviour in relation to energy use of residences. *Natural gas.*,p.2.

Grubb, M. 2014. *Planetary economics: energy, climate change and the three domains of sustainable development.* Routledge.

Guba, E.G. and Lincoln, Y.S. 1994. Competing paradigms in qualitative research. *Handbook of qualitative research.* **2**(163–194),p.105.

Guerra Santin, O. 2010. Actual energy consumption in dwellings: the effect of energy performance regulations and occupant behaviour. *Sustainable Urban Areas 33.*

Guerra Santin, O. 2011. Behavioural Patterns and User Profiles related to energy consumption for heating. *Energy and Buildings.* [Online]. **43**(10),pp.2662–2672. Available from: http://www.sciencedirect.com/science/article/pii/S0378778811002738.

Guerra Santin, O., Itard, L. and Visscher, H. 2009. The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. *Energy and Buildings.* [Online]. **41**(11),pp.1223–1232. Available from: http://www.sciencedirect.com/science/article/pii/S0378778809001388.

Guler, B., Ugursal, V.I., Fung, A.S. and Aydinalp-Koksal, M. 2008. Impact of energy efficiency upgrade retrofits on the residential energy consumption and Greenhouse Gas emissions in Canada. *International Journal of Environmental Technology and Management.* **9**(4),pp.434–444.

Günay, M.E. 2016. Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey. *Energy Policy.* **90**,pp.92–101.

Guo, J.J., Wu, J.Y. and Wang, R.Z. 2011. A new approach to energy consumption prediction of domestic heat pump water heater based on grey system theory. *Energy and Buildings.* **43**(6),pp.1273–1279.

Haas, R. and Schipper, L. 1998. Residential energy demand in OECD-countries and the role of irreversible efficiency improvements. *Energy economics.* **20**(4),pp.421–442.

Haben, S., Singleton, C. and Grindrod, P. 2016. Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Transactions on Smart Grid.* **7**(1),pp.136–144.

Hall, C. 2006. A picture of the United Kingdom using the National Statistics Socio-economic Classification. *POPULATION TRENDS-LONDON-.* **125**,p.7.

Hall, R. 2013. Mixed methods: In search of a paradigm. *L. Lê (Éd.), Conducting research in a changing and challenging World.*,pp.71–78.

Hall, R. and White, P. 2005. *Europes Population.* Routledge.

Harold, J., Lyons, S. and Cullinan, J. 2015. The determinants of residential gas demand in

Ireland. *Energy Economics*. [Online]. **51**,pp.475–483. [Accessed 20 May 2017]. Available from: http://www.sciencedirect.com/science/article/pii/S0140988315002388.

Haskey, J. 1998. Divorce and remarriage in England and Wales. *Population Trends*. (95),pp.18–22.

Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica: Journal of the Econometric Society.*,pp.1251–1271.

Hawkins, D.I., Mothersbaugh, D.L. and Best, R.J. 2013. *Consumer behavior: Building marketing strategy*. McGraw-Hill Irwin.

Heiple, S. and Sailor, D.J. 2008. Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles. *Energy and Buildings*. [Online]. **40**(8),pp.1426–1436. Available from: http://www.sciencedirect.com/science/article/pii/S0378778808000200.

Henry, C. and Kemmis, S. 1985. A point-by-point guide to action research for teachers. *The Australian Administrator*. **6**(4),pp.1–4.

Herbst, A., Toro, F., Reitze, F. and Jochem, E. 2012. Introduction to energy systems modelling. *Swiss journal of economics and statistics*. **148**(2),pp.111–135.

Hijazi, I., Ehlers, M., Zlatanova, S. and Isikdag, U. 2009. IFC to CityGML transformation framework for geo-analysis: a water utility network case *In*: *4th International Workshop on 3D Geo-Information, 4-5 November 2009, Ghent, Belgium*.

Hirst, E., Lin, W. and Cope, J. 1977. Simulation model of household fuel uses. *Applied Mathematical Modelling*. **1**(4),pp.177–180.

HM Treasury 2014. *Impact on households: distributional analysis to accompany budget 2014* [Online]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/293738/budget_2014_distributional_analysis.pdf.

Hoaglin, D.C., Iglewicz, B. and Tukey, J.W. 1986. Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*. **81**(396),pp.991–999.

Hogan, C. 2016. Does the Owner Have to Live in Their B&B? *startabedandbreakfast*. [Online]. Available from: http://www.startabedandbreakfast.co.uk/does-owner-have-live-bb.html.

Honkatukia, J. 2013. The VATTAGE regional model VERM-a dynamic, regional, applied general equilibrium model of the Finnish economy. *VATT Research Reports*. (171).

Houthakker, H.S. 1951. Electricity tariffs in theory and practice. *The Economic Journal*. **61**(241),pp.1–25.

Howell, S., Rezgui, Y. and Beach, T. 2017. Integrating building and urban semantics to empower smart water solutions. *Automation in Construction*.

Hsiao, C. 1985. Benefits and limitations of panel data. *Econometric Reviews*. **4**(1),pp.121–174.

Huebner, G.M., Hamilton, I., Chalabi, Z., Shipworth, D. and Oreszczyn, T. 2015. Explaining domestic energy consumption - The comparative contribution of building factors, sociodemographics, behaviours and attitudes. *Applied Energy.* **159**,pp.589–600.

Hughes, M. and Moreno, J.G. 2013. *Further Analysis of Data From the Household Electricity Usage Study: Consumer Archetypes.* Tech. Rep.). Cambridge: Department of Energy and Climate Change, Department for the Environment Food and Rural Affairs.

Hygh, J.S., DeCarolis, J.F., Hill, D.B. and Ranjithan, S.R. 2012. Multivariate regression as an energy assessment tool in early building design. *Building and Environment.* **57**,pp.165–175.

Iacovou, M. and Berthoud, R. 2000. Parents and employment. *RESEARCH REPORT-DEPARTMENT OF SOCIAL SECURITY.*,p.all-all.

IBM 2007. The benefits of software architecting. [Accessed 17 January 2018]. Available from: https://www.ibm.com/developerworks/rational/library/may06/eeles/.

IFS 2015. *The effect of the coalition's tax and benefit changes on household incomes and work incentives - Election 2015 - IFS* [Online]. [Accessed 8 March 2017]. Available from: http://election2015.ifs.org.uk/article/the-effect-of-the-coalition-s-tax-and-benefit-changes-on-household-incomes-and-work-incentives.

Independent 2016. Half of Brits get back together with their ex after breaking up, research reveals | The Independent. [Accessed 8 April 2017]. Available from: http://www.independent.co.uk/life-style/love-sex/half-of-brits-get-back-together-with-their-ex-after-breaking-up-research-reveals-a7438606.html.

InSMART 2013. InSMART:Integrative Smart City Planning. [Accessed 2 June 2017]. Available from: https://www.nottingham.ac.uk/research/groups/lucas/research/insmart.aspx.

ISER 2016. British Household Panel Survey (BHPS) - Institute for Social and Economic Research (ISER). [Accessed 16 April 2016]. Available from: https://www.iser.essex.ac.uk/bhps.

Isikdag, U. and Zlatanova, S. 2009. Towards Defining a Framework for Automatic Generation of Buildings in CityGML Using Building Information Models *In*: *3D Geo-Information Sciences* [Online]. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 79–96. Available from: http://link.springer.com/10.1007/978-3-540-87395-2_6.

IVAP3 2015. Self-Organizing Maps – the Kohonen's algorithm explained | Iván Vallés-Pérez. [Accessed 20 June 2017]. Available from: http://ivape3.blogs.uv.es/2015/03/15/self-organizing-maps-the-kohonens-algorithm-explained/.

Jain, A.K. 2010. Data clustering: 50 years beyond K-means. *Pattern recognition letters.* **31**(8),pp.651–666.

Jain, R.K., Smith, K.M., Culligan, P.J. and Taylor, J.E. 2014. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy.* [Online]. **123**,pp.168–178. Available from: http://www.sciencedirect.com/science/article/pii/S0306261914002013.

Jamasb, T. and Meier, H. 2011. Energy spending and vulnerable households.

Janke, J.R. 2010. Multicriteria GIS modeling of wind and solar farms in Colorado. *Renewable Energy*. **35**(10),pp.2228–2234.

Jiménez, M.J. and Heras, M.R. 2005. Application of multi-output ARX models for estimation of the U and g values of building components in outdoor testing. *Solar Energy*. **79**(3),pp.302–310.

Jimenez, R. and Yépez-García, A. 2016. *Composition and Sensitivity of Residential Energy Consumption*. Inter-American Development Bank.

Johnson, R.B., Onwuegbuzie, A.J. and Turner, L.A. 2007. Toward a definition of mixed methods research. *Journal of mixed methods research*. **1**(2),pp.112–133.

Jokela, J. 2016. CityGML building model production from airborne laser scanning.

Jones, P.J., Lannon, S. and Williams, J. 2001. Modelling building energy use at urban scale *In*: *seventh international IBPSA conference, Rio de Janeiro, Brazil.*, pp. 175–180.

Jones, R. V. and Lomas, K.J. 2016. Determinants of high electrical energy demand in UK homes: Appliance ownership and use. *Energy and Buildings*. [Online]. **117**,pp.71–82. [Accessed 20 May 2017]. Available from: http://www.sciencedirect.com/science/article/pii/S0378778816300676.

Jones, R. V, Fuertes, A. and Lomas, K.J. 2015. The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. *Renewable and Sustainable Energy Reviews*. **43**,pp.901–917.

Jones, R. V and Lomas, K.J. 2015. Determinants of high electrical energy demand in UK homes: Socio-economic and dwelling characteristics. *Energy and Buildings*. **101**,pp.24–34.

JRF 2005. The effect of parents' employment on outcomes for children | JRF. [Accessed 12 March 2017]. Available from: https://www.jrf.org.uk/report/effect-parents-employment-outcomes-children.

Kaden, R. and Kolbe, T.H. 2013. City-wide total energy demand estimation of buildings using semantic 3D city models and statistical data. *Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. **2**,p.W1.

Kalogirou, S.A. and Bojic, M. 2000. Artificial neural networks for the prediction of the energy consumption of a passive solar building. *Energy*. **25**(5),pp.479–491.

Kardinal Jusuf, S., Mousseau, B., Godfroid, G. and Soh Jin Hui, V. 2017. Integrated modeling of CityGML and IFC for city/neighborhood development for urban microclimates analysis *In*: *Energy Procedia*.

Karjalainen, J., Käkönen, M., Luukkanen, J. and Vehmas, J. 2014. Energy Models and Scenarios in the Era of Climate Change. *FFRC-Finland Futures Research Centre, Turku*.

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z. and Djurovic-Petrovic, M. 2010. A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment*. **45**(7),pp.1683–1697.

Kavousian, A., Rajagopal, R. and Fischer, M. 2013. Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy*. **55**,pp.184–194.

Kilian, L. 2008. The economic effects of energy price shocks. *Journal of Economic Literature*. **46**(4),pp.871–909.

Koehl, M. and Grussenmeyer, P. 2000. SEMANTIC ASPECTS FOR 3D MODELING. [Accessed 30 July 2018]. Available from: http://www.isprs.org/proceedings/XXXIII/congress/part4/507_XXXIII-part4.pdf.

Kohonen, T. and Somervuo, P. 1998. Self-organizing maps of symbol strings. *Neurocomputing*. **21**(1),pp.19–30.

Kornbrot, D. 2005. Point biserial correlation. *Wiley StatsRef: Statistics Reference Online*.

Krarti, M. and Ihm, P. 2009. Implementation of a building foundation heat tranfer model in EnergyPlus. [Accessed 23 July 2018]. Available from: https://www.tandfonline.com/doi/pdf/10.1080/19401490802613610?needAccess=true.

Kreith, F. 2017. *CRC Handbook of Thermal Engineering Second Edition* [Online]. CRC Press. [Accessed 24 July 2018]. Available from: https://www.taylorfrancis.com/books/9781315119717.

Krüger, A. and Kolbe, T.H. 2012. Building analysis for urban energy planning using key indicators on virtual 3D city models—the energy atlas of Berlin *In*: *Proceedings of the ISPRS Congress*.

Krysiak, F.C. and Weigt, H. 2015. The demand side in economic models of energy markets: the challenge of representing consumer behavior. *Frontiers in Energy Research*. **3**,p.24.

Kubota, T., Surahman, U. and Higashi, O. 2015. A Comparative Analysis of Household Energy Consumption in Jakarta and Bandung.

Kucukali, S. and Baris, K. 2010. Turkey's short-term gross annual electricity demand forecast by fuzzy logic approach. *Energy Policy*. **38**(5),pp.2438–2445.

Kucuksari, S., Khaleghi, A.M., Hamidi, M., Zhang, Y., Szidarovszky, F., Bayraksan, G. and Son, Y.-J. 2014. An Integrated GIS, optimization and simulation framework for optimal PV size and location in campus area environments. *Applied Energy*. [Online]. **113**,pp.1601–1613. Available from: http://www.sciencedirect.com/science/article/pii/S0306261913007381.

Kutz, M. 2015. *Mechanical Engineers' Handbook, Volume 1: Materials and Engineering Mechanics*. John Wiley & Sons.

Lai, F., Magoules, F. and Lherminier, F. 2008. Vapnik's learning theory applied to energy consumption forecasts in residential buildings. *International Journal of Computer Mathematics*. **85**(10),pp.1563–1588.

Lambert, P. 2006. The british household panel survey: introduction to a longitudinal data resource. *Longitudinal Data Analysis for Social Science Researchers Working Paper*. **2**.

Laurie, H. 2010. *Continuity and innovation in the design of Understanding Society: the UK*

*Household Longitudinal Study*. Understanding Society at the Institute for Social and Economic Research.

Leahy, E. and Lyons, S. 2010. Energy use and appliance ownership in Ireland. *Energy Policy*. **38**(8),pp.4265–4279.

Lee, Y.M., Liu, F., An, L., Jiang, H., Reddy, C., Horesh, R., Nevill, P., Meliksetian, E., Chowdhary, P. and Mills, N. 2011. Modeling and simulation of building energy performance for portfolios of public buildings *In*: *Simulation Conference (WSC), Proceedings of the 2011 Winter*. IEEE, pp. 915–927.

Leticia, B., Boogen, N. and Filippini, M. 2012. *Residential electricity demand for Spain: new empirical evidence using aggregated data*. CEPE Center for Energy Policy and Economics, ETH Zurich.

Lewis, C.D. 1982. Industrial and business forecasting methods: A Radical guide to exponential smoothing and curve fitting.

Li, D.H.W., Yang, L. and Lam, J.C. 2012. Impact of climate change on energy use in the built environment in different climate zones–a review. *Energy*. **42**(1),pp.103–112.

Li, J., Johnson, S.E., Han, W.-J., Andrews, S., Kendall, G., Strazdins, L. and Dockery, A. 2014. Parents' nonstandard work schedules and child well-being: A critical review of the literature. *The journal of primary prevention*. **35**(1),pp.53–73.

Li, Q., Meng, Q., Cai, J., Yoshino, H. and Mochida, A. 2009. Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*. [Online]. **86**(10),pp.2249–2256. Available from: http://www.sciencedirect.com/science/article/pii/S0306261908003176.

Li, Q., Ren, P. and Meng, Q. 2010. Prediction model of annual energy consumption of residential buildings *In*: *Advances in Energy Engineering (ICAEE), 2010 International Conference on*. IEEE, pp. 223–226.

Lieberman, E.S. 2005. Nested analysis as a mixed-method strategy for comparative research. *American Political Science Review*. **99**(03),pp.435–452.

Lincoln, Y.S., Lynham, S.A. and Guba, E.G. 2011. Paradigmatic controversies, contradictions, and emerging confluences, revisited. *The Sage handbook of qualitative research*. **4**,pp.97–128.

Longhi, S. 2015. Residential energy expenditures and the relevance of changes in household circumstances. *Energy Economics*. [Online]. **49**,pp.440–450. [Accessed 22 July 2017]. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0140988315001103.

Longhi, S. 2014. *Residential energy use and the relevance of changes in household circumstances*. ISER Working Paper Series.

Longhi, S. and Nandi, A. 2014. *A practical guide to using panel data*. Sage.

Loomis, C.P. 1936. The Study of the Life Cycle of Families. *Rural Sociology*. **1**(2),p.180.

Löschel, A. 2002. Technological change in economic models of environmental policy: a survey.

*Ecological economics*. **43**(2),pp.105–126.

Lowry, R. 2014. Concepts and applications of inferential statistics.

Lu, X. and M, V. 2011. Heat Convection - an overview | ScienceDirect Topics. [Accessed 23 July 2018]. Available from: https://www.sciencedirect.com/topics/chemistry/heat-convection.

Maile, T., Fischer, M. and Bazjanac, V. 2007. Building energy performance simulation tools-a life-cycle and interoperable perspective. *Center for Integrated Facility Engineering (CIFE) Working Paper*. **107**,pp.1–49.

Martín-Merino, M. and Román, J. 2006. A new SOM algorithm for electricity load forecasting *In*: *International Conference on Neural Information Processing*. Springer, pp. 995–1003.

MathWorks 2016. Genetic Algorithm - MATLAB. Available from: https://www.mathworks.com/discovery/genetic-algorithm.html?s_tid=gn_loc_drop.

Mavrogianni, A., Davies, M., Kolokotroni, M. and Hamilton, I. 2009. A GIS-based bottom-up space heating demand model of the London domestic stock *In*: *Proceddings 11th International IBPSA Conference, Building Simulation*.

McDonald, J.H. 2009. *Handbook of biological statistics*. Sparky House Publishing Baltimore, MD.

McFall, S.L. and Garrington, C. 2011. Understanding society: early findings from the first wave of the UK's Household Longitudinal Study.

McFarland, J.R., Reilly, J.M. and Herzog, H.J. 2004. Representing energy technologies in top-down economic models using bottom-up information. *Energy Economics*. **26**(4),pp.685–707.

McLoughlin, F., Duffy, A. and Conlon, M. 2012. Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. *Energy and Buildings*. **48**,pp.240–248.

Medina, M.A. 1999. A quasi-steady-state heat balance model of residential walls. *Mathematical and Computer Modelling*.

Meier, H. and Rehdanz, K. 2010. Determinants of residential space heating expenditures in Great Britain. *Energy Economics*. **32**(5),pp.949–959.

Mejri, O., Palomo Del Barrio, E. and Ghrab-Morcos, N. 2011. Energy performance assessment of occupied buildings using model identification techniques. *Energy and Buildings*. [Online]. **43**(2–3),pp.285–299. Available from: http://www.sciencedirect.com/science/article/pii/S0378778810003208.

MENR 2016. Republic of Turkey Ministry of Energy and Natural Resources - MENR Strategic Plan 2015-2019 Presentation. Available from: http://www.enerji.gov.tr/en-US/Activity/MENR-Strategic-Plan-2015-2019-Presentation.

Merriam, S.B. 2014. *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

Merriam, S.B. and Tisdell, E.J. 2015. *Qualitative research: A guide to design and implementation.* John Wiley & Sons.

Mertens, D.M. 2010. Transformative mixed methods research. *Qualitative inquiry.*

Metashapes 2016. 3D File Formats - metashapes.com. [Accessed 14 June 2017]. Available from: https://metashapes.com/support/3d-file-formats/.

Mignard, C. and Nicolle, C. 2014. Merging BIM and GIS using ontologies application to urban facility management in ACTIVe3D. *Computers in Industry.* **65**(9),pp.1276–1290.

Mihalakakou, G., Santamouris, M. and Tsangrassoulis, A. 2002. On the energy consumption in residential buildings. *Energy and Buildings.* **34**(7),pp.727–736.

Mills, B.F. and Schleich, J. 2010. Why don't households see the light?: Explaining the diffusion of compact fluorescent lamps. *Resource and Energy Economics.* **32**(3),pp.363–378.

Min, J., Hausfather, Z. and Lin, Q.F. 2010. A High-Resolution Statistical Model of Residential Energy End Use Characteristics for the United States. *Journal of Industrial Ecology.* **14**(5),pp.791–807.

MIT 2017. Heat Transfer Modes. [Accessed 22 July 2018]. Available from: http://web.mit.edu/16.unified/www/FALL/thermodynamics/notes/node115.html.

Molotch, N. 2010. Lecture 5: Surface Energy Balance Energy Budget by Latitude. [Accessed 24 July 2018]. Available from: https://www.colorado.edu/geography/class_homepages/geog_1001_f10/lectures/Lecture_5_surface_energy_balance_Molotch_guest lecture.pdf.

Monien, D., Strzalka, A., Koukofikis, A., Coors, V. and Eicker, U. 2017. Comparison of building modelling assumptions and methods for urban scale heat demand forecasting. *Future Cities and Environment.* [Online]. **3**(1),p.2. Available from: http://dx.doi.org/10.1186/s40984-017-0025-7.

Moore, R. 2012. Definitions of fuel poverty: Implications for policy. *Energy Policy.* **49**,pp.19–26.

Morgan, D.L. 1996. *Focus groups as qualitative research.* Sage publications.

Morgan, D.L. 1998. Practical strategies for combining qualitative and quantitative methods: Applications to health research. *Qualitative health research.* **8**(3),pp.362–376.

Morse, J.M. 2003. Principles of mixed methods and multimethod research design. *Handbook of mixed methods in social and behavioral research.*,pp.189–208.

Nagel, C. and Kolbe, T.H. 2007. Conversion of IFC to CityGML *In*: *Meeting of the OGC 3DIM Working Group at OGC TC/PC Meeting, Paris (Frankreich).*

Nagel, C., Stadler, A. and Kolbe, T.H. 2009. Conceptual requirements for the automatic reconstruction of building information models from uninterpreted 3D models. *Proceedings of the International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences.*,pp.46–53.

NationalGeographic 2015. latitude - National Geographic. [Accessed 27 July 2018]. Available from: https://www.nationalgeographic.org/encyclopedia/latitude/.

NCT 2014. Barriers remain for parents returning to work. Available from: https://www.nct.org.uk/press-release/barriers-remain-parents-returning-work.

Ndiaye, D. and Gabriel, K. 2011. Principal component analysis of the electricity consumption in residential dwellings. *Energy and buildings.* **43**(2),pp.446–453.

Nielsen, L. 1993. How to get the birds in the bush into your hand: results from a Danish research project on electricity savings. *Energy policy.* **21**(11),pp.1133–1144.

Nielsen, S. and Möller, B. 2013. GIS based analysis of future district heating potential in Denmark. *Energy.* **57**,pp.458–468.

Nottingham Council 2016. Insight web family - Demography (2016). [Accessed 23 May 2017]. Available from: http://jsna.nottinghamcity.gov.uk/insight/Strategic-Framework/Nottingham-JSNA/Related-documents/Demography-2016.aspx.

Nouvel, R., Mastrucci, A., Leopold, U., Baume, O., Coors, V. and Eicker, U. 2015. Combining GIS--based statistical and engineering urban heat consumption models: towards a new framework for multi-scale policy support. *Energy Build.* [Online]. **107**. Available from: http://dx.doi.org/10.1016/j.enbuild.2015.08.021.

Nouvel, R., Schulte, C., Eicker, U., Pietruschka, D. and Coors, V. 2013. CityGML-based 3D City Model for energy diagnostics and urban energy policy support. *IBPSA World.* **1**.

Nouvel, R., Zirak, M., Dastageeri, H., Coors, V. and Eicker, U. 2014. URBAN ENERGY ANALYSIS BASED ON 3D CITY MODEL FOR NATIONAL SCALE APPLICATIONS *In*: *Presented at the IBPSA Germany Conference.*

NRDC 2016. Study Highlights Energy Burden for Households and How Energy Efficiency Can Help | NRDC. [Accessed 15 May 2017]. Available from: https://www.nrdc.org/experts/khalil-shahyd/study-highlights-energy-burden-households-and-how-energy-efficiency-can-help.

Ofgem 2014. " Beyond average consumption " Development of a framework for assessing impact of policy proposals on different consumer groups Updated report to Ofgem. [Accessed 7 May 2017]. Available from: www.cse.org.uk.

Ofgem 2015. *Insights paper on households with electric and other non-gas heating* [Online]. [Accessed 13 May 2017]. Available from: https://www.ofgem.gov.uk/ofgem-publications/98027/insightspaperonhouseholdswithelectricandothernon-gasheating-pdf.

Oliveira Panão, M.J.N., Gonçalves, H.J.P. and Ferrão, P.M.C. 2008. Optimization of the urban building efficiency potential for mid-latitude climates using a genetic algorithm approach. *Renewable Energy.* [Online]. **33**(5),pp.887–896. Available from: http://www.sciencedirect.com/science/article/pii/S0960148107001358.

ONS 2013a. *170 Years of Industrial Change across England and Wales* [Online]. [Accessed 9 April 2017]. Available from: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/census/2011-census-analysis/170-years-of-industry/170-years-of-industrial-changeponent.html.

ONS 2015a. *Births by Parents' Characteristics in England and Wales: 2014* [Online]. Available

from:
http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsbyparentscharacteristicsinenglandandwales/2014.

ONS 2015b. *Chapter 5: Financial wealth, Wealth in Great Britain, 2012 to 2014* [Online]. Available from:
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/dcp171776_428639.pdf.

ONS 2013b. *Divorces in England and Wales: 2013* [Online]. Available from:
http://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/divorce/bulletins/divorcesinenglandandwales/2013#children-of-divorced-couples.

ONS 2015c. *English housing survey 2013 to 2014: household report.*

ONS 2014a. *Families and Households - Office for National Statistics* [Online]. [Accessed 19 March 2017]. Available from:
https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2015-01-28.

ONS 2013c. *Families by socio-economic classification of the Family Reference Person (FRP) and Family Type, UK.*

ONS 2014b. *Families in the Labour Market* [Online]. [Accessed 12 March 2017]. Available from:
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/dcp171776_388440.pdf.

ONS 2012. *Family Resources Survey: financial year 2010/11 - GOV.UK* [Online]. [Accessed 8 March 2017]. Available from: https://www.gov.uk/government/statistics/family-resources-survey-201011.

ONS 2013d. *Home ownership and renting in England and Wales – Detailed Characteristics.*

ONS 2011a. Household Energy Consumption in England and Wales. [Accessed 12 May 2017]. Available from:
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/dcp171766_321960.pdf.

ONS 2011b. *Housing* [Online]. Available from:
http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/social-trends-rd/social-trends/social-trends-41/housing-chapter.pdf.

ONS 2015d. *Labour Force Survey (LFS) - Office for National Statistics* [Online]. [Accessed 19 February 2017]. Available from:
https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/labourforcesurveylfs.

ONS 2011c. *LC6115EW NS-SeC of Household Reference Person by household composition* [Online]. [Accessed 19 March 2017]. Available from:
http://web.ons.gov.uk/ons/data/dataset-finder?p_auth=dKoXvGg1&p_p_id=FOLeftPanelSpring_WAR_FOLeftPanelportlet&p_p_lifecycle=1&p_p_state=normal&p_p_mode=view&p_p_col_id=column-

2&p_p_col_count=1&_FOLeftPanelSpring_WAR_FOLeftPanelportlet_context=Census&_FOLe.

ONS 2015e. *Lone parents with dependent children by marital status of parent, by sex, UK: 1996 to 2015.*

ONS 2015f. *Lone parents with dependent children by marital status of parent, by sex, UK: 1996 to 2015 - Office for National Statistics* [Online]. [Accessed 19 March 2017]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/adhocs/005660loneparentswithdependentchildrenbymaritalstatusofparentbysexuk1996to2015.

ONS 2013e. *Marriages in England and Wales: 2013* [Online]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/marriagecohabitationandcivilpartnerships/bulletins/marriagesinenglandandwalesprovisional/2013.

ONS 2011d. *Tenure by household size by number of bedrooms* [Online]. [Accessed 27 February 2017]. Available from: http://web.ons.gov.uk/ons/data/dataset-finder/-/q/datasetDownload/Census/DC4405EW?p_auth=kvuNs5Iz&p_p_auth=uN9hhmrY&p_p_lifecycle=1&_FOFlow1_WAR_FOFlow1portlet_geoTypeId=2011HTWARDH&_FOFlow1_WAR_FOFlow1portlet_UUID=0.

ONS 2015g. *The effects of taxes and benefits on household income Statistical bulletins - Office for National Statistics* [Online]. [Accessed 19 February 2017]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/theeffectsoftaxesandbenefitsonhouseholdincome/previousReleases.

ONS 2014c. *The single population increased between 2001 and 2011* [Online]. [Accessed 6 March 2017]. Available from: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/census/2011-census-analysis/how-have-living-arrangements-and-marital-status-in-england-and-wales-changed-since-2001-/summary.html.

ONS 2016. *Why are more young people living with their parents? | Visual.ONS* [Online]. [Accessed 4 April 2017]. Available from: http://visual.ons.gov.uk/living-with-parents/.

ONS 2015h. *Working and Workless Households: 2015.*

OrdnanceSurvey 2017. Raster and Vector data | Support | Ordnance Survey. [Accessed 27 May 2017]. Available from: https://www.ordnancesurvey.co.uk/support/understanding-gis/raster-vector.html.

Osborne, J.W. 2010. Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation.* **15**(12),pp.1–9.

OurNottinghamshire 2015. History, landmarks and famous people | Sneinton: A general history | Sneinton | Nottingham City: Specific areas of the City | Nottingham City | Notts Places | Our Nottinghamshire. [Accessed 22 May 2017]. Available from: http://www.ournottinghamshire.org.uk/page_id__733.aspx.

OVOenergy 2014. How much electricity does a home use? | OVO Energy. [Accessed 17 September 2017]. Available from: https://www.ovoenergy.com/guides/energy-guides/how-much-electricity-does-a-home-use.html.

Ozturk, H.K., Canyurt, O.E., Hepbasli, A. and Utlu, Z. 2004. Residential-commercial energy input estimation based on genetic algorithm (GA) approaches: an application of Turkey. *Energy and Buildings*. **36**(2),pp.175–183.

Pagano, M.E., Hirsch, B.J., Deutsch, N.L. and McAdams, D.P. 2003. The transmission of values to school-age and young adult offspring: Race and gender differences in parenting. *Journal of feminist family therapy*. **14**(3–4),pp.13–36.

Parker, D.S. 2003. Research highlights from a large scale residential monitoring study in a hot climate. *Energy and Buildings*. **35**(9),pp.863–876.

Patton, M.Q. 1990. *Qualitative evaluation and research methods* . SAGE Publications, inc.

Pedersen, L. 2007. Use of different methodologies for thermal load and energy estimations in buildings including meteorological and sociological input parameters. *Renewable and Sustainable Energy Reviews*. [Online]. **11**(5),pp.998–1007. Available from: http://www.sciencedirect.com/science/article/pii/S1364032105000924.

Platchkov, L.M. and Pollitt, M.G. 2011. The economics of energy (and electricity) demand. *The Future of Electricity Demand: Customers, Citizens and Loads*. **69**,p.17.

Reinhart, C.M. and Rogoff, K.S. 2009. *This time is different: eight centuries of financial folly*. Princeton University Press.

Remourban 2017. City overview. [Accessed 21 May 2017]. Available from: http://www.remourban.eu/Cities/Lighthouse-Cities/Nottingham/City-Overview.kl.

Rezgui, Y., Wilson, I. and Li, H. 2010. Promoting sustainability awareness through energy engaged virtual communities of construction stakeholders. *Collaborative Networks for a Sustainable World*.,pp.142–148.

Riadi, I., Istiyanto, J.E. and Ashari, A. 2013. Log Analysis Techniques using Clustering in Network Forensics. *arXiv preprint arXiv:1307.0072*.

Richalet, V., Neirac, F.P., Tellez, F., Marco, J. and Bloem, J.J. 2001. HELP (house energy labeling procedure): methodology and present results. *Energy and Buildings*. **33**(3),pp.229–233.

Roberts, S. 2008. Energy, equity and the future of the fuel poor. *Energy Policy*. **36**(12),pp.4471–4474.

Robson, C. and McCartan, K. 2016. *Real world research*. John Wiley & Sons.

Rose, D. and Pevalin, D.J. 2000. Social class differences in mortality using the National Statistics Socio-economic Classification—too little, too soon: a reply to Chandola. *Social science & medicine*. **51**(7),pp.1121–1127.

Rosenfeld, M.J. 2014. Couple Longevity in the Era of Same-Sex Marriage in the United States. *Journal of Marriage and Family*. **76**(5),pp.905–918.

Rossman, G.B. and Wilson, B.L. 1985. Numbers and words combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation review.* **9**(5),pp.627–643.

Rubin, D.B. 2004. *Multiple imputation for nonresponse in surveys.* John Wiley & Sons.

Rylatt, R.M., Gadsden, S.J. and Lomas, K.J. 2003. Methods of predicting urban domestic energy demand with reduced datasets: a review and a new GIS-based approach. *Building Services Engineering Research and Technology.* **24**(2),pp.93–102.

de Sa, P. 1999. Mapping the energy future: Energy modelling and climate change policy. *Energy Policy.* **27**(7),pp.433–434.

Saha, G.P. and Stephenson, J. 1980. A model of residential energy use in New Zealand. *Energy.* **5**(2),pp.167–175.

Sanquist, T.F., Orr, H., Shui, B. and Bittner, A.C. 2012. Lifestyle factors in US residential electricity consumption. *Energy Policy.* **42**,pp.354–364.

Santamouris, M., Kapsis, K., Korres, D., Livada, I., Pavlou, C. and Assimakopoulos, M.N. 2007. On the relation between the energy and social characteristics of the residential sector. *Energy and Buildings.* **39**(8),pp.893–905.

Sapsford, R. and Jupp, V. 2006. *Data collection and analysis.* Sage.

Saran, S., Wate, P., Srivastav, S.K. and Krishna Murthy, Y.V.N. 2015. CityGML at semantic level for urban energy conservation strategies. *Annals of GIS.* **21**(1),pp.27–41.

Sardianou, E. 2005. Household energy conservation patterns: evidence from Greece *In*: *London School of Economics and Political Science Symposium London Current Social Research in Greece.*, pp. 1–26.

Schot, J., Kanger, L. and Verbong, G. 2016. The roles of users in shaping transitions to new energy systems. *Nature Energy.* **1**,p.16054.

Schröder, C., Rehdanz, K., Narita, D. and Okubo, T. 2015. The decline in average family size and its implications for the average benefits of within-household sharing. *Oxford Economic Papers.* [Online]. **67**(3),p.760. Available from: +.

Sciencing 2014. What Is Solar Altitude? [Accessed 30 July 2018]. Available from: https://sciencing.com/solar-altitude-23364.html.

Scott, A., Evandrou, M., Falkingham, J. and Rake, K. 2001. Going Into residential care: evidence from the BHPS 1991-1998.

SEMAGIX 2009. What is Semantic Data. [Accessed 1 August 2018]. Available from: http://www.semagix.com/what-is-semantic-data.htm.

Shirani, F., Parkhill, K., Butler, C., Groves, C., Pidgeon, N. and Henwood, K. 2016. Asking about the future: methodological insights from energy biographies. *International Journal of Social Research Methodology.* **19**(4),pp.429–444.

Shmueli, G., Patel, N.R. and Bruce, P.C. 2016. *Data Mining for Business Analytics: Concepts, Techniques, and Applications with XLMiner.* John Wiley & Sons.

SIG3D 2014. Modeling Guide for 3D Objects - Part 2: Modeling of Buildings (LoD1, LoD2, LoD3) - SIG3D Quality Wiki EN. [Accessed 30 May 2017]. Available from: http://en.wiki.quality.sig3d.org/index.php/Modeling_Guide_for_3D_Objects_-_Part_2:_Modeling_of_Buildings_(LoD1,_LoD2,_LoD3).

SIG3D 2015. SIG 3D der GDI-DE Geodateninfrastruktur Deutschland . . **2015**(8/24/2015). Available from: http://www.sig3d.org/index.php?catid=2&language=en.

Skelton, C. 2013. *Soft City Culture and Technology: The Betaville Project.* Springer Publishing Company, Incorporated.

Slife, B.D. and Williams, R.N. 1995. *What's behind the research?: Discovering hidden assumptions in the behavioral sciences.* Sage publications.

Somekh, B. and Lewin, C. 2005. *Research methods in the social sciences.* Sage.

Son, H., Kim, C., Kim, C. and Kang, Y. 2015. Prediction of government-owned building energy consumption based on an RReliefF and support vector machine model. *Journal of Civil Engineering and Management.* [Online]. **21**(6),pp.748–760. Available from: http://dx.doi.org/10.3846/13923730.2014.893908.

Sorokin, P.A., Zimmerman, C.C. and Galpin, C.J. 1930. *A systematic source book in rural sociology,.* Minneapolis: University of Minnesota Press.

Sorrell, S. 2015. Reducing energy demand: A review of issues, challenges and approaches. *Renewable and Sustainable Energy Reviews.* **47**,pp.74–82.

Sözen, A. and Arcaklioglu, E. 2007. Prediction of net energy consumption based on economic indicators (GNP and GDP) in Turkey. *Energy policy.* **35**(10),pp.4981–4992.

Stewart, D.W. and Shamdasani, P.N. 2014. *Focus groups: Theory and practice.* Sage publications.

Stone, J., Berrington, A. and Falkingham, J. 2011. The changing determinants of UK young adults' living arrangements. *Demographic Research.* **25**,pp.629–666.

Strand, R.K., Pedersen, C.O. and Crawley, D.B. 2001. Modularization and Simulation Techniques for Heat Balance Based Energy and Load Calculation Programs: The Experience of the ASHRAE Toolkit and EnergyPlus *In*: *Building Simulation 2001.*

Strzalka, A., Bogdahn, J., Coors, V. and Eicker, U. 2011. 3D City modeling for urban scale heating energy demand forecasting. *HVAC&R Research.* **17**(4),pp.526–539.

Sun, Y., Wang, R., Li, J. and Liu, J. 2017. GIS-based multiregional potential evaluation and strategies selection framework for various renewable energy sources: a case study of eastern coastal regions of China. *Energy Science & Engineering.* **5**(3),pp.123–140.

SUST-it 2016. Calculating the cost of running old appliances, electricity usage. [Accessed 12 June 2017]. Available from: https://www.sust-it.net/running-costs-of-old-appliances.php.

Swan, L.G. and Ugursal, V.I. 2009. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews.* [Online]. **13**(8),pp.1819–1835. Available from: http://www.sciencedirect.com/science/article/pii/S1364032108001949.

Tashakkori, A. and Teddlie, C. 2010. *Sage handbook of mixed methods in social & behavioral research.* Sage.

Taylor, M. and Change, E.R.C. on M.-S. 1995. British Household Panel Survey user manual vol. A: Introduction, technical report and appendices.

Teddlie, C. and Tashakkori, A. 2003. Major issues and controveries inthe use of mixed methods in the social and behvioral sciences. *Handbook of mixed methods in social & behavioral research.*,pp.3–50.

The poverty site 2016. United Kingdom: Low income by family type. Available from: http://www.poverty.org.uk/05/index.shtml.

theguardian 2016. Percentage of global population living in cities, by continent | Environment | theguardian.com. [Accessed 19 August 2017]. Available from: https://www.theguardian.com/news/datablog/2009/aug/18/percentage-population-living-cities.

Thomas, R., Have, T., Kunselman, A.R., Pulkstenis, E.P. and Landis, J.R. 1998. Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics.*,pp.367–383.

thomthom 2015. CleanUp 3.3 for SketchUp. [Accessed 13 June 2017]. Available from: http://www.thomthom.net/thoughts/2015/01/cleanup-3-3-for-sketchup-respecting-your-locks/.

Tiwari, P. 2000. Architectural, demographic, and economic causes of electricity consumption in Bombay. *Journal of Policy Modeling.* **22**(1),pp.81–98.

Torres-Reyna, O. 2010. Getting started in factor analysis (using Stata 10). *Princeton University, New Jersey.*

Trimble 2017. Importing and Exporting COLLADA Files | SketchUp Knowledge Base. [Accessed 13 June 2017]. Available from: https://help.sketchup.com/en/article/3000168.

tripod 2015. Support Vector Machine Regression . . **2015**(10/30/2015). Available from: http://kernelsvm.tripod.com/.

Trochim, W.M.K. and Donnelly, J.P. 2006. Research methods knowledge base.

Troy, A. 2008. Geodemographic segmentation *In*: *Encyclopedia of GIS.* Springer, pp. 347–355.

Tso, G.K.F. and Guan, J. 2014. A multilevel regression approach to understand effects of environment indicators and household features on residential energy consumption. *Energy.* [Online]. **66**,pp.722–731. [Accessed 20 May 2017]. Available from: http://www.sciencedirect.com/science/article/pii/S0360544214000772.

Tso, G.K.F. and Yau, K.K.W. 2007. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy.* **32**(9),pp.1761–1768.

TUBerlin 2017. Energy Atlas Berlin. [Accessed 16 July 2017]. Available from: http://energyatlas.energie.tu-berlin.de/en/.

Tuhus-Dubrow, D. and Krarti, M. 2010. Genetic-algorithm based approach to optimize building envelope design for residential buildings. *Building and Environment.* [Online]. **45**(7),pp.1574–1581. Available from: http://www.sciencedirect.com/science/article/pii/S0360132310000077.

Tunstall, R., Bevan, M., Bradshaw, J., Croucher, K., Duffy, S., Hunter, C., Jones, A., Rugg, J., Wallace, A. and Wilcox, S. 2013. The links between housing and poverty: an evidence review. *JRF Report, York: Joseph Rowntree Foundation.*

UKDataService 2016. UK Data Service. [Accessed 27 May 2016]. Available from: https://www.ukdataservice.ac.uk/.

UKGeographics 2016. Uk Geodemographic Classifications Review - UK Geographics. [Accessed 18 December 2016]. Available from: http://www.ukgeographics.co.uk/blog/uk-geodemographic-classifications-review.

UKGOV 2016. UK Population Density. [Accessed 8 February 2017]. Available from: http://www.neighbourhood.statistics.gov.uk/HTMLDocs/dvc134_c/index.html.

UN 2014. *World Urbanization Prospects: The 2014 Revision, Highlights.* New York, USA: United Nations.

Varma, S. 2006. Preliminary item statistics using point-biserial correlation and p-values. *Educational Data Systems Inc.: Morgan Hill CA. Retrieved.* **16**(07).

Villmann, T. 1999. Benefits and Limits of the Self Organizing Map and its Variants in the Area of Satellite Remote Sensoring Processing. *Topology.* **2**,p.2.

Vringer, K., Aalbers, T. and Blok, K. 2007. Household energy requirement and value patterns. *Energy Policy.* **35**(1),pp.553–566.

Waller, W. and Hill, R. 1951. The family: a dynamic interpretation.

Webb, D.A., Sweet, D. and Pretty, I.A. 2002. The emotional and psychological impact of mass casualty incidents on forensic odontologists. *Journal of Forensic Science.* **47**(3),pp.539–541.

Wells, W.D. and Gubar, G. 1966. Life cycle concept in marketing research. *Journal of Marketing Research.*,pp.355–363.

Wezenberg, H. and Dewe, M.B. 1995. Adaptive neural networks for tariff forecasting and energy management *In: Neural Networks, 1995. Proceedings., IEEE International Conference on.* IEEE, pp. 877–881.

WhatPrice 2017. Advantages and Disadvantages of UK House Types. [Accessed 12 March 2017]. Available from: http://www.whatprice.co.uk/building/uk-house-types.html.

White, J.M. and Klein, D.M. 2008. *Family theories.* Sage.

Whiting, S. 2010. Socio-demographic comparison between those UK families with up to two children and those with three or more. *Population Matters. Available online at: http://populationmatters. org/documents/family_sizes. pdf (accessed April 3, 2014).*

Wiesmann, D., Lima Azevedo, I., Ferr??o, P. and Fern??ndez, J.E. 2011. Residential electricity consumption in Portugal: Findings from top-down and bottom-up models. *Energy Policy.*

**39**(5),pp.2772–2779.

Williams, L.V. 2011. *Prediction markets: Theory and applications.* Routledge.

Williams, R. 2013. Panel data 4: Fixed effects vs random effects models.

Wooldridge, J.M. 2010. *Econometric analysis of cross section and panel data.* MIT press.

Wooldridge, J.M. 2002. Econometric Analysis of Cross Section and Panel Data. *Booksgooglecom.* **58**(2),p.752.

Wróżyński, R., Sojka, M. and Pyszny, K. 2016. The application of GIS and 3D graphic software to visual impact assessment of wind turbines. *Renewable Energy.* **96**,pp.625–635.

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

Xu, H. and Spitler, J.D. 2014. The relative importance of moisture transfer, soil freezing and snow cover on ground temperature predictions. *Renewable Energy.*

Xu, X., Ding, L., Luo, H. and Ma, L. 2014. From building information modeling to city information modeling. *Journal of Information Technology in Construction.*,pp.292–307.

Yilmaz, L., Chan, W.K. V, Moon, I., Roeder, T.M.K., Macal, C. and Rossetti, M.D. 2015. A REVIEW OF ARTIFICIAL INTELLIGENCE BASED BUILDING ENERGY PREDICTION WITH A FOCUS ON ENSEMBLE PREDICTION MODELS

Yin, R.K. 2013. *Case study research: Design and methods.* Sage publications.

Yohanis, Y.G., Mondol, J.D., Wright, A. and Norton, B. 2008. Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use. *Energy and Buildings.* **40**(6),pp.1053–1059.

Young, D. 2008. When do energy-efficient appliances generate energy savings? Some evidence from Canada. *Energy Policy.* **36**(1),pp.34–46.

Yuce, B., Mourshed, M. and Rezgui, Y. 2017. An ANN-Based Energy Forecasting Framework for the District Level Smart Grids *In*: *Smart Grid Inspired Future Technologies: First International Conference, SmartGIFT 2016, Liverpool, UK, May 19-20, 2016, Revised Selected Papers.* Springer, pp. 107–117.

Yuce, B. and Rezgui, Y. 2017. An ANN-GA semantic rule-based system to reduce the gap between predicted and actual energy consumption in buildings. *IEEE Transactions on Automation Science and Engineering.* **14**(3),pp.1351–1363.

Zhang, N., Xu, W. and Cao, P. 2011. Analysis of the factors that influenced rural households' residential energy consumption—Based on micro data of nine provinces. *Chin. J. Popul. Sci.* **3**,pp.73–82.

Zhang, Q. 2004. Residential energy consumption in China and its comparison with Japan, Canada, and USA. *Energy and buildings.* **36**(12),pp.1217–1225.

Zhang, T., Siebers, P.-O. and Aickelin, U. 2012. A three-dimensional model of residential energy consumer archetypes for local energy policy design in the UK. *Energy Policy.*

**47**,pp.102–110.

Zhao, H. and Magoulès, F. 2012. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*. [Online]. **16**(6),pp.3586–3592. Available from: http://www.sciencedirect.com/science/article/pii/S1364032112001438.

Zhou, S. and Teng, F. 2013. Estimation of urban residential electricity demand in China using household survey data. *Energy Policy*. **61**,pp.394–402.

Zhu, L. and Mao, H. 2015. Building Interior Model Interpolation between IFC and CityGML. *The Open Construction and Building Technology Journal*. **9**(1).

Znouda, E., Ghrab-Morcos, N. and Hadj-Alouane, A. 2007. Optimization of Mediterranean building design using genetic algorithms. *Energy and Buildings*. **39**(2),pp.148–153.