ENERGY AND SOCIETY: UNDERSTANDING THE COST OF CONSUMPTION

Thesis submitted in accordance with the requirements of the University of Liverpool for the Degree of Doctor in Philosophy.

ELLEN TALBOT

Department of Geography and Planning

School of Environmental Sciences

University of Liverpool

August 2021

Terence Talbot

1954 – 2017

"Outside of a dog, a book is a man's best friend. Inside of a dog, it's too dark to read"

- ***Groucho Marx***

Declaration

I, Ellen Talbot, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated within.

Signed:

Abstract

Energy and Society: Understanding the Cost of Consumption

Ellen Talbot

Fuel poverty definitions have previously been limited by a monotopical monetary indicator and annualised statistics and UK policy directives are complicated by multiple fragmented and misaligned stakeholders. This thesis presents a holistic view of the current geographies of energy consumption, aiding the reimagining of the fuel poverty vernacular through the inclusion of socio-demographic indicators and novel consumer datasets in order that it be broadened to encompass the lived experience.

The predominant aims of this work were to firstly provide a thorough exploration of the geography of energy consumption, and those factors that contextualise differentiated access to, and consumption of both gas and electricity in England and Wales. This work endeavoured to provide a substantive contribution to the integration of consumer datasets into social science research, by proving the utility in effective data linkage across novel commercially generated big data and other ancillary traditional data sources, as well as acting as a catalyst for increased collaboration between academic and commercial partners by highlighting the value of releasing consumer data for social science research.

Acknowledgements

I would like to thank my supervisors Professor Alex Singleton and Dr Dani Arribas-Bel for their support and advice throughout the duration of my studies, both academically and personally. The support of the Geographical Data Science Lab has been invaluable, and I would like to extend my thanks to each and every member for the companionship, learning, laughs and lunches that have been shared by us all. Special thanks to Alec Davies for introducing me to Quick Chef, and for talking me down off the ceiling more than once. I would also like to take this opportunity to thank the ESRC, the North West Doctoral Training Centre and the CRDC, whose funding made this research and all the invaluable experiences that accompanied it possible.

To my close family and friends, my Mum, Pete and Tim especially who showed support, love and patience and offered motivation, pep talks, proofreading and pints when I needed it most – thank you! To Sarah, who got so sick of hearing about it she moved to New Zealand. Thanks also to friends far and wide who have called, texted, written, bought coffee or sent anonymous Dominos deliveries as a surprise – it didn't go unnoticed and it really has meant the world! To Jonny Marshall, the Piggies, the Mad Wednesday Society, PRNDP, and the Hot Girls – your WhatsApp groups are my safe space – thank you for responding at all hours of the day and night. My experience of the University of Liverpool and the city itself was made incredibly special by course mates, housemates and old mates alike; thank you to 'The Ivanhoes', who continue to be lifelong friends today, and Booth, who took me to every pub in Liverpool and made sure I never went hungry.

Thanks also to Alfie and Merlin even though they can't read, my PhDogs – the walks, cuddles and play times were a welcome break during long days of coding.

Finally some honourable mentions; COVID-19 – for closing the pubs, cancelling my plans and gifting me six weeks furlough; John and Elis for keeping it session; Manchester R-Ladies for their knowledge, enthusiasm and encouragement, my colleagues at Pets at Home for their encouragement and understanding and Dr Nicki Mazey, who wrote me a prescription and turned the lights back on in my life.

Table of Contents

Figures

Tables

Acronyms and Abbreviations

AHC	After Housing Cost
BEIS	Department for Business, Energy and Industrial Strategy
BHC	Before Housing Cost
CDRC	Consumer Data Research Centre
DEC	Display Energy Certificate
DECC	Department for Energy and Climate Change
DEP	Domestic Energy Provider
EDA	Exploratory Data Analysis
EHS	English Housing Survey
EPC	Energy Performance Certificate
ESRC	Economic and Social Research Council
EUC	Energy User Classification
FiT	Feed in Tariff
GDPR	General Data Protection Regulations
GHG(s)	Greenhouse Gases
GIS	Geographic Information Systems
GUI	Graphic User Interface
HIHC	High Income High Cost
HIP	Home Improvement Pack
IEA	The International Energy Agency
IHD(s)	In Home Display
IMD	Indices of Multiple Deprivation
IUC	Internet User Classification
kWh	Kilowatt Hours
LIHC	Low Income High Cost
LILC	Low Income Low Cost
LSOA	Local Super Output Area
MEES	Minimum energy efficiency standards
MSOA	Middle Super Output Area
ONSPD	Office of National Statistics Postcode Database
ONSPL	Office of National Statistics Postcode Lookup
OA	Output Area
OAC	Output Area Classification
OECD	Organisation of Economic Co-operation and Development
Ofgem	Office of Gas and Electricity Markets
ONS	Office of National Statistics
PCS	Postcode Sector
SAP	Standard Assessment Procedure
SMAR	Smart Meter Adoption Rate
SOA	Super Output Area
TOE	Tonne of Equivalent Oil
TOU	Time Of Use (tariff)
WCSS	Within Cluster Sum of Squares

1 Introduction

1.1 Background

Energy consumption is an integral part of society and everyday life. Its governance has historically been disjointed, with multiple agencies acting independently of one another with limited capacity. As the issues of climate change become increasingly prevalent, governments in developed nations have begun to address these disparities and form cohesive bodies, yet issues of historic responsibility, burden sharing, and financing investments still make this global agreement an enormous challenge.

In the UK, the reduction of carbon emissions and Greenhouse Gases is at the core of the Government's energy policy, in which the residential sector plays a key role. Central policy measures have been introduced to support and encourage reductions in domestic energy consumption through efficiency improvements. In tandem to these policy objectives, the UK Government is also committed to reducing the number of households which find themselves struggling to meet their energy demands through cost-effective improvements such as insulation, which lasts for many years and has a positive impact on the UKs overall emissions target.

Another element of this efficiency strategy is the upgrade of the existing energy infrastructure to a 'smart grid', which encompasses introducing smart meters into domestic properties, replacing traditional meters with ones which allow two way connectivity as well as increased visibility over a household's consumption at a highly granular level for both supplier and consumer. This lends itself to the collection of large scale digitised consumption datasets, housing information and metadata on people's temporal patterns of consumption and geographic location. Commercially this can be utilised to a competitive advantage, but in a research context can provide new opportunities to enhance the valuable geodemographic dimension and better comprehend the interactions, preferences and constraints exhibited by society.

This emergence of 'big data' as a general trend across many domains has shaped what has been coined as the 'fourth paradigm of science', representing a fundamental shift to data driven research. Whilst the applications for such innovative datasets are numerous and wide ranging, epistemological, ethical and methodological questions have emerged which have implications with regards to the data's content and coverage when reused in a research environment. For example, consumer generated 'big data' often exists as a by-product of an alternate process, leading to self-selected populations, built in bias and lack of quality control. However, given that the data is rarely available outside of their commercial environments, these issues remain an important consideration for academics.

Given that such data has very rarely been shared for academic research, this thesis presents a unique opportunity to study a commercial and nationally expansive dataset. It is anticipated that through the combination of these big data with traditional data sources, these will present a holistic view of the contemporary geographies of energy consumption. This contributes to the current field of research and engenders benefits for applications where the characteristics of energy consumption are of high importance. Examples of this include the understanding and reimagining of fuel deprivation, where energy consumption is constrained to the extent that it negatively impacts upon households lives. Such applications would likely include the expansion of definitions in the fuel poverty vernacular away from a static, monetary measure to a multi-faceted indicator. Furthermore to provide guidance on smart meter roll-out programmes which ensure that the Domestic Energy Providers are achieving the greatest social improvement whilst also meeting the government mandated installation rates.

1.2 Research Aims

As summarised above, the purpose of this research is to provide a thorough exploration of energy consumption, using innovative 'big data' alongside the close coupling of more traditional population and built environment data. The more specific aims of the analyses can be understood across two main themes. Firstly, to address the challenges encountered when considering innovative big data in an academic context, second to address the limitations of the current fuel poverty vernacular by utilising data tools to enrich the current understanding. As such, the objectives of this thesis are:

- An evaluation of the utility of big data, exploring the ability to extract relevant insight when making inferences about the general population.

- The critical evaluation of traditional fuel poverty definitions, that incorporates demographic characteristics and highlights the multifaceted nature of the lived experience of fuel poverty.

Which are set out to be achieved through the following research questions:

1. What data quality issues are unique to consumer energy data and how can they be addressed pragmatically to enable the generation of useful insights?

2. What are the limitations of the current UK fuel poverty definition?

3. To what extent can smart meter, energy and demographic data inform the fuel poverty vernacular to address the limitations of the current fuel poverty definition?

An underlying theme of all objectives of this work is evaluating the representativeness of these new forms of energy data, and thus appraising their potential application for matters of public and social good. This represents a stark contrast to commercial endeavours, which typically focus on understanding their consumer base to maximise profits. It was hypothesised that energy consumption data, both singularly and through linkage with traditional ancillary data, may advance our knowledge of the interactions of people and place by providing metrics which quantify the functional relationship between them.

1.3 Thesis Structure

This thesis comprises seven chapters which begin with a review of relevant literature, and an overview of the innovative big data utilised by this study and associated methodologies that are core to this thesis. The analytical chapters that follow provide validation of the energy data utilised within the thesis, followed by their integration within a geodemographic framework to provide new insights into fuel poverty. The thesis concludes with a discussion of the findings within the national policy context and makes some suggestions for future work. The following sub-sections provide a more detailed overview of each chapter as presented in the thesis.

1.3.1 Chapter 2 – Literature Review

This chapter provides an overview of concepts and literature that frame the analyses conducted throughout this thesis. This includes, firstly, an overview of the current global energy landscape and key limitations of policy that need addressing if this domain is to progress. The European Union (EU) and UK policy contexts are also considered, before a detailed examination of the implications of these policies in the UKs residential sector. This is followed by an overview of traditional and current consumption practises in the UK; the evolution of energy technologies from static, non-responsive infrastructure to an upgraded 'smart grid' and its associated benefits and limitations. This encompasses the emergence of first and second generation smart meters and the accompanying technologies of 'in home display units', again noting the benefits and impacts on suppliers, consumers and policy objectives.

Discussion diversifies to consider the societal implications of these policy objectives, and the multifaceted discourse of material deprivation, and its relationship to fuel poverty is introduced before a detailed exploration of the validity of the current fuel poverty definition. Finally, an overview is provided of the big data, consumer data and energy data landscapes in regard to the study of people and populations, thus framing the social and spatial data sources drawn upon in the empirical chapters of the thesis.

1.3.2 Chapter 3 – Data and Methodological Framework

Chapter 3 provides a contextual introduction to the attributes and characteristics of those datasets utilised within the thesis. Various data quality issues are discussed alongside data manipulation that was necessary in order to enable linkage and extract meaningful insight. In addition to energy data, this also includes discussion of other ancillary relevant sources including census variables, and a series of pre-compiled indicators that include the geodemographic Internet User Classification and the Indices of Multiple Deprivation.

Following the presentation of the data sources, geographical scale is considered and the Modifiable Areal Unit Problem introduced. A discussion on postcode geographies and the justification for employing a bespoke dasymetric reweighting methodology is presented, evaluating the relative benefits of the selected technique in comparison to alternative methods.

The final section of the chapter details a framework for building new geodemographic classifications, giving an overview of the typical stages of construction.

The purpose of this chapter was to inform the interpretations of the analysis presented in chapters 4, 5 and 6, but also to provide data driven evidence of some of the issues outlined in the literature review, such as the bias and veracity of the smart meter data and the pragmatic steps required when aiming to reliably integrate them into academic research.

1.3.3 Chapter 4 – The Geographies of Smart Meter Users

Chapter 4 is the first of three empirical chapters and examines the geography of smart meter adoption rates as well as address data quality issues inherent in the underlying consumer energy data. Analyses begin by investigating the extent of the geographic variations as a feature of the self-selective nature of those consumers recorded by the data before addressing those substantial cleaning procedures required before useful insight could be extracted. Secondly, aggregated energy consumption patterns are investigated at varying temporal granularities, before finally endeavouring to highlight the demographic trends present within the smart meter data. This is done firstly to reiterate the biases present within the dataset, and secondly to investigate the intersection between smart meter adoption rates and socio-demographic characteristics of deprivation.

1.3.4 Chapter 5 - Representing Fuel Poverty with an Energy User Classification

Chapter 5 is the second empirical study that examines how Energy Performance Certificate (EPC) data can provide evidence of socio-demographic disparities in fuel poverty. This strengthens a narrative of fuel poverty being a multifaceted problem which is developed through a segmentation based on energy efficiency and small area demographic characteristics. This typology presents a critique of those methods implemented to currently define fuel poverty. Methodological steps are outlined pertaining to the segmentation of small areas, followed by an exploration of each resulting clusters characteristics in terms of their geodemographic and consumption characteristics.

1.3.5 Chapter 6 – Evaluating the Energy User Classification's Utility; Suggesting Areas of Improvement for the Domestic Energy Provider to Achieve the Greatest Social Good.

Chapter 6 presents an evaluation of the multivariate outputs from the previous chapter. Internal validation interrogates fit statistics for each resulting cluster, alongside an external validation of the utility of the EUC to conceptualise a more nuanced definition of fuel poverty. A final practical validation in the form of a case study highlights the intersection between smart meters and this new measure of fuel poverty. By combining these outputs from the previous chapters, it is possible to provide recommendation in order that the on-going rollout of smart metering technology achieves a greater social good, whilst also remaining mindful of the constraints faced by both the DEP and the consumer.

1.3.6 Chapter 7 – Discussion, Application and Future Works

Chapter 7 consolidates the principle findings from each chapter and their overarching contribution to this thesis. Key methodological and knowledge contributions are highlighted in the context of both smart meter and EPC data, but also more widely for the extrapolation of consumer and energy data, particularly when identifying vulnerable populations. Key issues and limitations of the work are addressed, and the chapter concludes by highlighting paths for future development.

1.4 Notes on Software and Code

The majority of analyses in this thesis were undertaken in R Open Source Software for Statistical Computing; an open source software freely downloadable from WWW.R-PROJECT.ORG. Associated codes are available on request or through Github[1]. Other software utilised included QGIS – an open source geographic information systems software.

1.5 Ethics

This research was deemed exempt by a University of Liverpool Research Ethics and Integrity Officer. Proof is provided in Appendix 9.1

[1] HTTPS://GITHUB.COM/SGETALBO/THESIS_CODE - this is a private repository due to the nature of the work conducted on the DEP dataset – please send a request for access to ETALBOT1291@GMAIL.COM

2 Literature Review

This chapter provides an overview of the concepts and literature relevant to the exploration and analysis conducted in this thesis. Section 2.1 provides an introduction to the global energy landscape and its necessary governance, followed by the implications across various geographic scales, notably: international, within the European Union and domestically in the UK. Section 2.2 gives an overview of the deprivation and fuel poverty vernacular; firstly understanding deprivation as a whole, before paying particular attention to the changing definitions of fuel poverty and their relative merits and pitfalls as well as the fundamental challenges that need addressing if we are to fully appreciate and begin to tackle the scope of the problem. Sections 2.3 then provides an overview of energy data and technologies as a changing landscape – from traditional to innovative and how the improved technologies can be utilised for population insight amongst other things, both commercially and in the social sciences. Section 2.4 concludes the chapter with an overview of the current data landscape and big data and addresses the practicalities of integrating this data into social science research. This includes a consideration of consumer data as a facet of big data, and smart meter data's position within the consumer realm. Finally, geodemographics are introduced, and their utility in summarizing complex relationships across space.

2.1 Energy and Energy Policy in a Global Context

Global energy governance is driven by three distinct areas of policy; climate change, energy security and energy access (Dubash and Florini, 2011). There is also a strong development component, sometimes regarded as the fourth area of influence (Goldthau and Witte, 2010).

Climate change is a foregrounding issue for governments and policy makers as well as academia and the media; a rise in global temperatures of 0.8°C since 1880 - the majority of this warming occurring since 1975 due to the increased release of greenhouse gases (GHGs) into the Earth's atmosphere - shows a significant upward trend with potentially catastrophic results (Nejat *et al.*, 2015). Climate policy carries significant consequences for the future of energy globally – energy activities in the three main domains of industry, transportation and residential consumption contribute disproportionately to local, regional and global pollution problems – two thirds of all GHGs are generated by these domains, leaving little doubt that governance must concentrate on them (Dubash and Florini, 2011).

The International Energy Agency (The IEA) suggests that the impact of climate change can only be minimised if the global temperature increase is limited to less than 2°C by 2050 (The International Energy Agency, 2017). The most likely way of achieving this result is thought to be through energy efficiency, which could contribute up to 49% of the energy related CO_2 emission reductions that are needed. Given that energy efficiency "constitutes the optimum utilisation of energy resources" it is therefore considered one of the most important mechanisms through which countries can act to mitigate the effects of climate change in both the short and long term (G20, 2016, p.4). As such, energy efficiency and energy conservation are a long-term priority for G20 members (Figure 2-1) (who currently account for over 80% of both global energy consumption and greenhouse gas emissions worldwide). Work has already begun to this end and despite the massive emissions contributions they make, the G20 members have proven experience in achieving energy reduction and implementing energy efficiency measures; from 1990 to 2013, the G20s total energy consumption saving reached around 4.3 billion tonnes of oil equivalent and about 10.4 billion tonnes of carbon dioxide emissions were avoided.

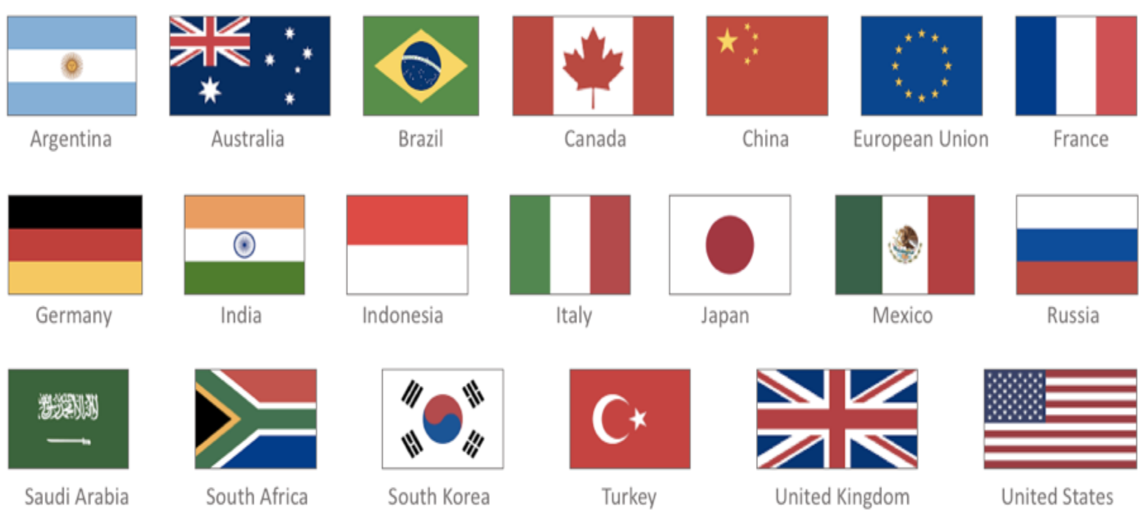


FIGURE 2-1 THE G20 COUNTRIES

Within the G20 countries, America and China account for about 40% of global greenhouse gas emissions, generating a staggering 50% and 80% of their respective electricity from coal (the most carbon intensive of the fossil fuels) and the role of China is to become even more pronounced given the rapid acceleration in its energy demand (Ekins *et al.*, 2015, chap.2). Internationally, the UK comes tenth in terms of global CO_2 emissions at 1.6%. This comes after China, the US, India, Russia, Japan, Germany, South Korea, Canada and Iran (Nejat *et al.*, 2015).

China provides a perfect example of the impact rapid urbanisation can have on a country's energy demands, contributing significantly to a country's consumption profile, especially regarding consumption of carbon intensive fossil fuels. It has witnessed rapid economic growth since their government implemented the 'Reform and opening up policy' in the late 1980s, with an average increase in yearly gross domestic product of more than 9% (Wang *et al.*, 2016). Given that China has not yet completed the historical task of industrialisation and urbanisation, this level of growth poses a number of challenges for the nation; a large amount of empirical research confirms the existence of a correlation between economic growth and energy consumption, and thusly also an increase in CO_2 emissions (Wang *et al.*, 2016; Hu *et al.*, 2017). Furthermore, the rapid development of China's economy has led to high concentrations of urban populations; China's urbanisation rate increased from 37.7% in 2001 to 55% in 2014, promoting the growth in energy demand in a domestic setting, where consumption of electricity has tripled between 2001 and 2014 (Hu *et al.*, 2017). As such, it remains that energy policy is central to climate change policy and vice versa for the simple reason that the combustion of fossil fuels is the single most important source of all the emissions responsible for anthropogenic climate change.

As discussed, there is considerable evidence that access to energy and the quantum of its use is closely correlated with both economic growth and advances in human development (Dubash and Florini, 2011). In low income and emerging economies especially, the replacement of low quality fuels such as biomass (organic material, usually burned to generate heat) with high quality ones such as oil and coal is central to their ability to participate in economic modernization (Smil, 2010; Grubler, 2012). Worldwide, 1.4 billion people lack access to electricity and 2.7 billion rely on biomass for daily tasks, "depriving them of any opportunity to participate in energy dependent processes of economic modernization" (Dubash and Florini, 2011, p.9)

As countries develop from rural communities into globalized economies, their changing consumption leads to altered priorities regarding their energy security. Despite there being no clear definition of energy security and the concept being 'rather blurred', for some it is taken to mean the reliable provisions of fuel - be that through the reduction of short term 'shocks' (political disruptions, technical failures or intermittency) or tackling long term stresses (depletion of fossil fuels, the accumulation of greenhouse gases or growing demand). For others the definition is linked to the protection of the poor against commodity price volatility (Löschel *et al.*, 2010; Stern, 2011; Winzer, 2012). Throughout the urbanisation process, investments are made in infrastructure and technology, improving a countries short term energy security. In the longer term, the Organisation for Economic Co-operation and Development (OECD) promotes a global shift toward a greener future to tackle the climate crisis, whilst also recognising that policies are not simply a one size fits all and must be implemented at a national level to take account of local environmental, economic and developmental settings (OECD and The International Energy Agency, 2012).

2.1.1 The Governors of Global Energy

The successful governance of global energy requires the blurring of several boundaries – between global and national scales, state and non-state actors and between fuel sources and markets (Dubash and Florini, 2011). The significant and urgent realities of the 21st century are highly politicised and understanding the role that markets and institutions play in determining outcomes of global energy relations are crucial (Goldthau, 2016).

Given that decentralised transnational agencies such as The IEA, The Organisation of Petroleum Exporting Countries and the Energy Charter Treaty as well as the G8 and G20 "have not yet shown the interest or ability to grapple with the full range of needs to address the trade-offs between them", problems frequently arise and there is a movement to reduce the reliance on this fragmented infrastructure. Any newly developing international framework for climate change mitigation must include all stakeholders, including the developing countries and evolve with a mutual interdependence (Dubash and Florini, 2011, p.7).

Stakeholders are often limited in scope and capacity, and exist only to deal with a specific problem or crises and have evolved over time into a piecemeal network of overlapping and partial frameworks of principles, rules, norms and processes and are inadequate to address many of the market or governance failures (Dubash and Florini, 2011). This leads to a lack of effectiveness and swiftness in policy making at the transnational level, because of poor integration with policy making at the national level. These difficulties often mean that fundamental issues remain unresolved, and a coherent global agreement continues to be an enormous challenge. There needs to be a move away from the traditional notions of energy security in order to break the deadlock between market players (such as energy companies and commercial banks) and the public sector (governments, international organisation and policy makers) over burden sharing when it comes to developing a new energy architecture (Goldthau, 2016). Furthermore, globalised energy policies are particularly difficult to implement given the arguments over how the responsibility of emission reduction should be distributed, lack of agreement over historic responsibility and financing the investment in developing countries (Ekins *et al.*, 2015). For the developing and newly developed there are mounting concerns over their levels of growth and the impact they will have on the acceleration of climate change as well as the sustainability of their usage.

One example in particular pertains to the burden sharing associated with carbon leakage. This phenomenon occurs as a result of the disjointed global environmental policies and is the product of two conflicting economic factors. Firstly, a country with strict environmental policies may focus on reducing high emission production resulting in the price rises of these products, thereby stimulating another (less strict) country to increase production and export of such goods to meet the demand, leading to the increase in the CO_2 emissions of the export country. Conversely the price of fossil fuels may also cause carbon leakage. Strict environmental regulation of one country will lead to decline in demand of fossil fuels, leading to price decline. As a result, countries with less strict environmental regulation may use fossil fuels as a substitute for other inputs in the industrial process, thus increasing greenhouse gas emissions (Guo *et al.*, 2010)

What can be agreed on however, is that these levels of growth are unsustainable and so a coherent set of standards must be agreed on. Especially pertinent is the access to oil and gas, which is certain to remain a key policy objective for governments around the world regardless (Goldthau and Witte, 2010). In 2016, the Paris Agreement was ratified, replacing the Kyoto Agreement and bringing together 175 states and the European Union for the first time under a common clause to mitigate global climate change, thus beginning to tackle the obstacles outlined here. It relies on National Defined Contributions but also provides 'enhanced support to assist developing nations' (Jonas *et al.*, 2019).

2.1.2 International Drivers of Change and Response to Global Energy Policy

It is pertinent to look at the role of the many EU (European Union) institutions and their influence over national and international energy policies given that since the 1990s it has been a prominent actor in global climate change negotiations, with ambitious targets granting it world leading status (Skovgaard, 2014). Over recent years, the EU has agreed several new documents that promise to strengthen Europe's presence in international energy policies (Goldthau, 2016). During the first decade of the 21st century, the EU underwent a profound change in its attitude towards energy policy and ended that decade with formalising its commitment to energy policy when it included a chapter in the 2009 Lisbon Treaty with the specific aim of fostering a more cohesive relationship between member states with regard to policy implementation (Birchfield and Duffield, 2011). Other documents which cement this commitment include the EU Energy 2020 Strategy and the Energy Roadmap 2050 which illustrate energy scenarios for the next four decades (European Commission, 2010; Langsdorf, 2011).

The EU 2020 Energy Strategy stipulates that all EU countries must aim to reduce GHG emissions by 20%, increase the share of renewable energy by 20% and to make a 20% improvement in energy efficiency on 1990 levels by the year 2020 (Bradshaw, 2013). However, sceptics doubted the logic behind these targets given the convenience of the 20/20/20 – 2020 title and the fact that only the emissions target was legally binding meant that some felt the targets arbitrary (Goldthau, 2016). At the time of writing it is too early to know whether the EU 2020 targets have been reached, but the latest report from the European Environment Agency states that it is on track to meet it's 20% emissions reduction target for 2020. It does however also acknowledge that the global pandemic spanning the entirety of 2020 will have had a significant impact on GHG emissions and levels of consumption. The revised targets for 2030 and the long term targets for 2050 are set out, again acknowledging that changes may have to be made to reflect post-Covid recovery plans (European Environment Agency, 2020). It is fair to say that there is still scepticism around the 'political will' of the EU and it's nations and their ability to reach these targets without legal obligation to do so, especially given the impact of the UK contribution being removed from these targets following it's exit from the EU (Sanchez Nicolas, 2020).

It is agreed that despite the EU being by far one of the largest importers of energy, (buying nearly twice as much as the US and five times that of China) it redeems itself by having the lowest energy intensity of all regions and the highest demand for renewables (Smart Energy GB, 2018a). The UK benefits from considerable fossil fuel resources ranking 4th among IEA members and 17th globally, however, a 50% reduction in domestic production in 2000 meant that the UK was forced to import 17% of its oil needs and 38% of its natural gas needs (Nejat *et al.*, 2015). However, this demand for renewables may in future come to cause problems; the EU is comparatively small compared to other individual nations and landmass is a valuable asset where renewable energy sources are concerned. As such, the EUs high targets for the development and deployment of these energy sources is cause for concern. Despite this, and of particular relevance to this thesis, the report does point to smart technology as the key to fully exploiting the potential for energy savings, the reasons for which are explored more fully in proceeding sections (European Commission, 2010).

2.1.3 The UK Energy Policy Landscape

The UK has become one of the most committed EU states to combating climate change and has a key role in demonstrating international leadership on the issue as well as being central to securing the previously mentioned 2015 Paris agreement. The core of UK policy is one of CO_2 emission mitigation. The 2008 Climate Change Act committed the UK to a 34% reduction in greenhouse gas emissions in comparison to 1990 levels by 2020 and an 80% reduction by 2050. As a result, central policy measures have been introduced to support and encourage reductions in energy usage. This is to be achieved through a process of setting 5 year caps on GHG emissions termed 'Carbon Budgets' (HM Government, 2009). The UK has launched a Green Investment Bank, with £3 billion capital and has given £125 million towards research and development in carbon capture and storage (Goldthau, 2016) It has also successfully reduced its GHG emissions over several decades, with a 44% fall from 1990 levels in 2008 (Nejat *et al.*, 2015; The Committee on Climate Change, 2019). This was largely driven by the decreasing combustion of coal for electricity generation, as well as reduced levels of fuel consumption by businesses and the industrial sector, and more efficient vehicles resulting in lower transport emissions (BEIS, 2019; Hausfather, 2019). In May 2019, The Committee on Climate Change reaffirmed the UKs commitment to the Paris Agreement by pledging to achieve a net zero GHG emission target for 2050 (The Committee on Climate Change, 2019).

As well as being a global player in the introduction of GHG reduction policy, the UK government are also committed to reducing energy consumption closer to home. The residential sector is a high priority when tackling overall CO_2 emission reduction. The UK is above the EU average in terms of domestic energy consumption at 29% of overall usage. In 2016 the EU averaged 25.4% (Eurostat, 2019). This variation could be explained by the relative age of the UK housing stock, which is the oldest in the EU (Nicol *et al.*, 2016). Many UK homes date from the Victorian era and are as such, less well insulated and ultimately consuming more energy to maintain the same level of thermal comfort, especially given the UKs temperate climate and residential consumption and efficiency is discussed in more detail in the following section (Liddell and Morris, 2010). As one of the first countries to industrialise, the UK offers the longest observed record of energy transitions in the modern era (Ekins *et al.*, 2015). An analysis of energy transitions since before the industrial revolution in the UK explains the dynamics of long-run change as a positive economic and welfare feedback loop and as previously discussed is now being replicated in developing countries such as China (Fouquet, 2010).

Both the domestic and commercial energy sectors in the UK are regulated by the Department for Business, Energy and Industrial Strategy (BEIS) and the Office of Gas and Electricity Markets (Ofgem) who are in place to regulate the monopoly companies which run the gas and electricity networks. Increased energy prices globally have driven up costs for consumers and so these regulatory bodies act in the interest of consumers to ensure their energy security by taking decisions on fair pricing, facilitating decarbonisation and enabling competition and innovation (Ofgem, 2019a). In the domestic market, there are currently 12 large and 46 small energy providers. The market share is monitored by Ofgem and assessed based on how many electricity meters are installed on the distributional network by a supplier. As of 2016, British Gas were the largest provider with a 23% market share (Longley *et al.*, 2018). Many of these providers also have a commercial offering although the market is very different.

2.1.4 Residential Energy Efficiency and Energy Consumption

Given consistent increases in domestic energy consumption (the average household reported increases of 1.2% between 2000 and 2008) because of inefficiencies, increased appliance usage, and higher standards of comfort and convenience, dwellings have become an important target area for the UK government in terms of energy efficiency improvements and emissions savings (Firth *et al.*, 2008). The previously mentioned 2008 Climate Change Act included domestic reductions in its 20% target, and a government white paper published in 2009 stipulated a 29% decrease in domestic emissions on 2008 levels by 2020 yet it is likely that these targets will also be missed (HM Government, 2009). Thusly in the UK, central policy measures have been introduced that support and encourage reductions in domestic energy use through efficiency improvements, (HM Government, 2009), and alongside these the UK government is committed to reducing the number of households who find themselves struggling to meet the energy demands and their associated costs whilst also having a positive impact on the UKs overall emissions target.

To respond specifically to the problem of excess energy consumption through inefficient housing, regulations introduced for new buildings in 2006 mean that improvements have to be made to levels of insulation, air tightness and the efficiency of space and water heating and lighting, all of which aim to improve the energy efficiency of the building and reduce its emissions; and from 2016 the UK government introduced legislation to improve the efficiency of all newly built housing that ensured they were all zero carbon (Nejat *et al.*, 2015). However, this focus on new build houses was challenged; it is estimated that up to 80% of existing housing stock will still be in use in 2050 when the targets need to be met and so it is argued that the legacy of older, hard to treat buildings, characterised by poor insulation and high consumption should be at the centre of the policy debate (Swan *et al.*, 2013; Robinson *et al.*, 2018a). Developing policies which invest in improving energy efficiency appear on the surface to make a lot of sense given that buildings account for up to 40% of total fuel consumption and a third of total emissions; it is a cost-effective approach to making long term improvements and savings; insulation on a home lasts for many years, not to mention the environmental benefits and the economic growth generated by this relatively new industry (Middlemiss, 2017).

With regards to managing domestic energy consumption, historically, traditional energy meters - which are only variations of those present since the early 20th century - provided limited visibility to households of their energy consumption. They are still popular because of their low production price and excellent reliability but are often installed in difficult to reach locations displaying usage only in terms of Kilowatt Hours rather than cost. This makes it very difficult for residents to get an overview of particularly inefficient practises within the household. The government proposed a full transition away from these first generation meters to smart metering for both gas and electricity by 2020 (Haben *et al.*, 2016). This target was revised in 2019 to 2024, and then again in 2021 to 'mid-2025' to reflect the impact that the pandemic has had on 'staff, customers and the supply chain' (Ofgem, 2021). Upgrading of manually read gas and electricity meters to include smart grids and smart meters is considered instrumental to achieving emissions reduction and meeting energy efficiency targets. The associated technologies and their cost and benefits are discussed in greater detail in a following section.

On the supply side, there is an operational setting in which Domestic Energy Providers (DEPs) are under increasing pressure to make improvements. For them, traditional meters limit their ability to accurately predict demand, whereas smart grids would allow more efficient planning in both the short and long term; real time responses to outages and emergencies, the ability to detect theft and to capitalise on dynamic pricing, implementing time of use tariffs to shift demand away from peak times, leading to more efficient generation and storage of energy and in turn reducing wastage and improving reliability (Guerreiro *et al.*, 2015). For both suppliers and consumers, the Internet connected technology of the smart grid, smart meter and in home displays (IHDs) provide reliable real time readings on the consumption of energy at an unprecedented cadence. The overarching objective of such technology is that these lead to a shift in demand whereby consumers take more active interest in their consumption, ultimately leading to decreased consumption and therefore emissions.

2.1.5 Summary

It is clear that the disjoined nature of global energy governance is impacting on individual nations abilities to achieve proposed climate change mitigation targets; distribution of responsibilities; support for developing nations and issues such as carbon leakage all create tensions between global actors. It is intended that the Paris Agreement will go some way to bringing all these actors together under a common clause by implementing national defined contributions to allow each member state to report their best efforts rather than defining a 'one size fits all' policy whilst also providing enhanced support to developing nations.

The European Union presents a world leading stance on global climate change mitigation, and The UK is one of the most committed members. The intention is to reduce consumption by 80% of 1990 levels by 2050, whilst also achieving a net zero carbon status by the same date (The Committee on Climate Change, 2019). With respect to domestic energy consumption, legislation is already in place to ensure that newly built homes are efficient and energy efficiency improvement policies are being enacted, including the implementation of smart technologies on both the supply and demand side, but this is not without criticism and is thus covered in greater detail in a proceeding section.

Domestic energy consumption is increasing, as are its costs, as households strive for comfort and convenience realised through increased appliance usage and indoor temperatures. This increasing usage is having a direct impact on the global climate, and governments are posturing to consistently reduce carbon emissions through a number of legislative measures. In the UK these include domestic energy reforms through improvements in energy efficiency level; a part of which is the installation of smart metering technology within homes.

2.2 Separating Deprivation and Fuel Poverty

2.2.1 Characterising Deprivation

In human geography and studies of demography, deprivation is considered to be a multi-dimensional phenomenon, characterised by a range of domains encompassing finance, health, education and crime amongst others and is a consequence of a lack of income and other resources, which cumulatively can be seen as living in poverty (Payne and Abel, 2012). Material deprivation refers to an individual's inability to afford or access basic resources such as food, heating or educational materials to such an extent that they find themselves excluded from the society in which they live (OECD, 2014). This definition moves away from an income based measure of poverty and increases the attention on non-monetary indicators, adding important information which permits a greater understanding of the causal mechanisms at work (Boarini and D'Ercole, 2006).

The probability of an individual experiencing material deprivation is dependent on a range of characteristics of themselves and the household where they live. Lower income individuals are more likely to experience material deprivation than higher income ones, and low education in the household head results in a higher probability of the household being deprived. Other factors which influence deprivation are; household structure; household tenure; employment status and to a lesser extent age and ethnicity (Boarini and D'Ercole, 2006).

The population living with deprivation and fuel poverty are likely overlapping. Of the commonly cited drivers of deprivation, as discussed above, many also relate to fuel poverty. As a result, understanding those predicting factors of overall deprivation is likely to reveal which are closely linked to fuel poverty. However, it is important to reiterate that fuel poverty is a distinct form of hardship, separate from material deprivation for reasons such as the fact that the rate of fuel poverty is linked to changes in energy prices and the energy efficiency of dwellings and appliances (Watson and Maitre, 2015). In the UK in particular, fuel poverty and poverty are divided by the politicisation of energy and welfare policies; chiefly that you can treat one with energy efficiency only but not the other. This emphasis means that it is possible to ignore the impact that austerity measures have had on the fuel poor and similarly, that other exploratory factors linked to fuel poverty are reduced, leading to an ignorance of the interrelated drivers of the problem and the lived experience of fuel poverty (Middlemiss, 2017). Furthermore, this entrenchment of the new austerity politics diminishes the responsibility to develop associated policy measures such as increasing income for the poorest household and controlling rising energy costs.

2.2.2 Energy Poverty and Fuel Poverty

Globally, the literature attends to energy poverty and fuel poverty with one overarching definition, however, the terms are divergent in some important respects. In countries both developed and developing, the overarching condition of both energy and fuel poverty is "the inability to attain a socially and materially necessitated level of domestic energy services" (Bouzarovski and Petrova, 2015).

Importantly, fuel poverty is induced through high or rising energy prices, low incomes and inefficient housing stock, whereas energy poverty is driven by a lack of networked energy provision due to economic under-development. Energy poverty is expressed through a lack of access to adequate facilities and is consequently linked to negative impacts on health, equality, education and economic development (Pachauri and Spreng, 2003). Fuel poverty, however, manifests itself through inadequate heating in the home and the lack of important services such as lighting and appliances, leading to both short and long term mental and physical health problems as well as exclusions in wider society. In short, fuel poverty is the term widely used to refer to the societal inequalities rising from a person's lack of ability to consume energy, largely due to the cost, and it is energy poverty that encompasses the environmental injustices that lead to chronic under consumption of fuel in a domestic setting (Bouzarovski and Petrova, 2015)

Across Europe, fuel prices have increased steadily, putting greater pressure on governments and policy makers to better define, measure and work to alleviate the phenomenon. The European Commission has suggested that a pan-European definition would be inappropriate given diversity of socio-political and energy contexts found across the EU, however, fuel poverty has recently gained attention in national political, practitional and academic agendas within France, Spain, Germany and Belgium amongst others, who have been engaging in this widely recognised societal challenge (Thomson and Snell, 2013; Bouzarovski and Petrova, 2015).

2.2.3 UK Fuel Poverty

This section considers residential fuel poverty in a UK context and the ways in which it can be defined, understood, addressed and legislated for.

Residential fuel poverty has been historically difficult to define, and as discussed, there is no internationally unified example. The broadly accepted definition is that of Brenda Boardman (1991). She defines fuel poverty as:

"The inability to afford adequate heat because of energy efficiency in the home."

It exists as the product of three aggravating factors: low income, high energy prices and energy inefficient housing stock. It is the last of these which is critical in differentiating fuel poverty from other types of deprivation as certain types of dwellings will undeniably cost more to heat than others, as a function of their physical configuration and specification. Fuel poverty exists where low income houses pay high energy costs because they live in inefficient dwellings. This is a very real concern for many households in the UK due to the comparatively low quality of the national housing stock when compared to the rest of the EU (Royston and Guertler, 2013; Nicol *et al.*, 2016). This inefficiency coupled with a temperate climate which regularly causes internal temperatures to dip below those required for healthy living (21°C in the living room and 18°C in all other rooms (Simcock *et al.*, 2016) presents a very real risk of households suffering from associated physical and mental health concerns, ranging from asthma caused by damp and mould, to excess winter deaths (Liddell and Morris, 2010).

Isolation of fuel poverty as a distinct form of deprivation is usually traced back to the 1973 oil crisis, when soaring domestic fuel prices resulted in many households facing difficulties affording fuel (Bradshaw and Hutton, 1983). The issue began to garner wide public attention and in 1975 the National Right to Fuel Campaign was formed with the objective of ending fuel poverty in the UK and securing a warm, dry and well-lit home for all, regardless of income and location (National Right to Fuel Campaign, 2013).

Despite this promise, major advancements in the fuel poverty vernacular were not made until the publishing of Brenda Boardman's 'Fuel Poverty' in 1991, which offered the first quantitative definition and multi-disciplinary account of the problem. She introduced a 10% threshold definition, whereby fuel poverty was the situation where expenditure on energy services was equal to or greater than 10 percent of income (Boardman, 1991, p.201). This figure was derived from then contemporary data as to the energy expenditure across the lowest three income deciles.

Even still, fuel poverty did not become a formal concern of the UK government until 2000, when the Warm Homes and Energy Conservation Act 2000 required that the Government "specify a target date for achieving the objective of ensuring that as far as is reasonably practicable, persons in England and Wales do not live in fuel poverty" (UK Parliament, 2000). Subsequently, a target was established that fuel poverty should be eradicated in England by 2016, and in vulnerable households by 2010 (a vulnerable household is defined as one which contains infants, the elderly, or those who are disabled or suffering from a long term illness) (Department of Energy and Climate Change, 2014b). A complimentary strategy was born, and a version of Boardman's fuel poverty definition written into policy for monitoring purposes.

In the subsequent decade, a range of policies both economic and technical were implemented with the goal of tackling fuel poverty. However, on the face of it, these were a resounding failure. Fuel poverty steadily rose year on year and both the 2010 and 2016 poverty eradication targets were missed, which can be construed as evidence of an ineffective policy approach on the part of multiple incumbent Governments. In 2010, the UK coalition government commissioned a review of fuel poverty definitions, and the October Spending Review included a commitment to re-evaluate the use of the 10% definition as part of a drive to reduce state expenditure, and a subsequent report which later became widely known as the Hills Review reaffirmed fuel poverty as a serious problem distinct from income poverty (Hills, 2012). It marked a change in the UK fuel poverty vernacular, from a condition that should and can be eradicated (as in the previous fuel poverty target by 2016), to a condition that can at best be alleviated (Middlemiss *et al.*, 2019), replacing the 10% indicator with the Low Income High Cost (LIHC) indicator which considers a household to be fuel poor if: "They have required fuel costs that are above national median levels and were they to spend that amount they would be left with a residual income below the official poverty line" (Department of Energy and Climate Change, 2017).

2.2.4 The Hills Report

In his review, Hills reconsidered the difference between abject poverty and fuel poverty; it is not a new distinction, but the Hills review represented a further entrenchment of this 'dividing practise' as discussed in the previous section. The separation has highly important policy implications, chiefly because it distances discussions of fuel poverty from those of overarching poverty and was identified with reference to the interaction between low incomes and high required spending. In doing so it foregrounds energy efficiency measures as an appropriate response to fuel poverty above measures that address low incomes or cost of living (Middlemiss, 2017). Whilst Hills praised the 10% definition for its ability to capture the interactions of the drivers of fuel poverty, he found fault with its ability to effectively represent the nature of that problem and identified a multitude of weaknesses, some of which are highlighted in Table 2-1 overleaf.

TABLE 2-1 ISSUES AND PROBLEMATISATIONS WITH THE 10% DEFINITION

Issue	Problematization
The fixed threshold	A fixed threshold means that the definition of fuel poverty is extremely sensitive to that choice.
High Income High Cost	Under the 10% definition, those with high incomes and high fuel costs can be considered fuel poor if their energy costs are sufficient.
Treatment of housing cost	For the purposes of measurement, incomes have been considered before housing costs are subtracted, i.e., inclusive of income that is not truly disposable as it is apportioned to a specific, unavoidable purpose.

As an alternative, Hills proposed a conceptualisation of fuel poverty which reconfigures it in relative terms; the Low Income High Cost definition (LIHC). It differs from the 10% definition which is based on an absolute threshold for fuel costs and is instead relative; a household is fuel poor if its fuel expenditure is comparatively high, and its income is comparatively low. Figure 2-2 illustrates the LIHC fuel poverty definition.

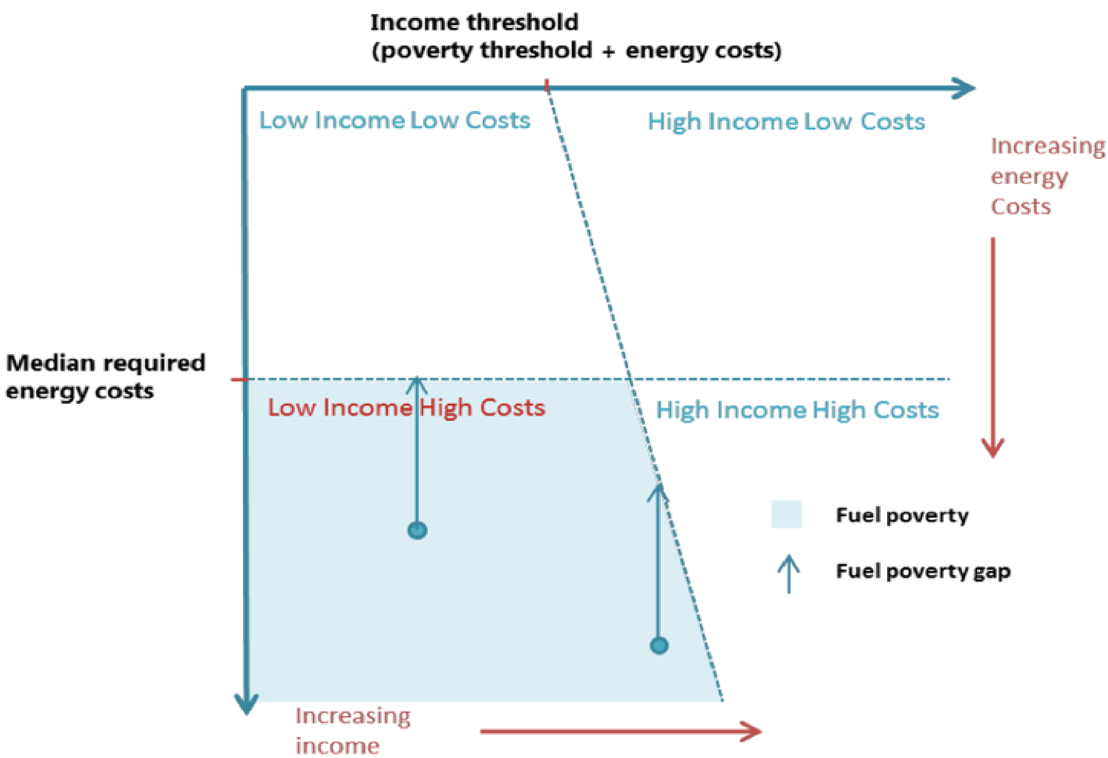


FIGURE 2-2 THE LOW INCOME HIGH COST FUEL POVERTY INDICATOR (BEIS 2019)

The thresholds used are as follows; the income threshold falls "where subtraction of required equivalised energy costs from income leaves the household at the Department for Work and Pensions' official poverty line, after housing costs" (Hills, 2012, p.53). Effectively, this defines a low income household as one that, having paid required energy costs, is below the official poverty line. "The energy cost threshold lies at the point whereby equivalised household bills equal the national median" and whilst this does not entirely eliminate the failure of prior policy it does go some way to mitigating the distortionary impact of price rises upon official figures (Hills, 2012, p.59). Hills also defined another measure of fuel poverty known as the fuel poverty gap - the reduction in required spending which would take a household out of fuel poverty as can be seen in Figure 2-2 (Department for Business Energy and Industrial Strategy, 2019).

Under the 10% definition, the Department for Energy and Climate Change measured fuel poverty under both before housing costs (BHC) and after housing costs (AHC) but used BHC for official statistics. Considering AHC results in a reduction in considered income for those with higher housing costs, which manifests itself in a shift away from pensioners who are more likely to own their homes outright towards working age adults, including families with children. This was a popular move as it was argued that AHC more accurately reflects the composition of the fuel poor group, where housing costs are high.

When the LIHC definition was written into official policy in 2013 (Department of Energy and Climate Change, 2013) the number of fuel poor households did decrease from 4 million to 2.7 million but the number of fuel poor individuals increased from 7.4 million to 7.8 million (Middlemiss and Gillard, 2015; Robinson *et al.*, 2018b). This came as a result of the elimination of some Low Income Low Cost (LILC) and High Income High Cost (HIHC) households, but equivalised energy usage meant more larger households with higher occupation were considered fuel poor. It was chosen partly for this very reason; it has a tendency to show a consistent population of fuel poor households over time due to its equivalisation of fuel costs, further entrenching the notion that fuel poverty is a condition that can at best be alleviated, whilst the introduction of the fuel poverty gap indicator also placed the emphasis on cost-effective spending to target and prioritise only the most vulnerable, as a result of the government's austerity driven policies (Middlemiss, 2017)

The official target was to 'ensure that as many fuel poor homes as is reasonably practicable achieve a minimum energy efficiency rating of Band C by 2030', with interim milestones of 'Band E by 2020' and 'Band D by 2025', therefore placing the entire focus of the strategy on energy efficiency improvements (Department of Energy and Climate Change, 2014c). The Committee on Fuel Poverty whose key role is to monitor and report on progress towards these milestones commented that despite the average fuel poverty gap closing by 14% over the last 4 years, progress towards achieving even the smallest improvements is "slow and flat-lining", as can be seen in Figure 2-3 overleaf (Committee on Fuel Poverty, 2018).

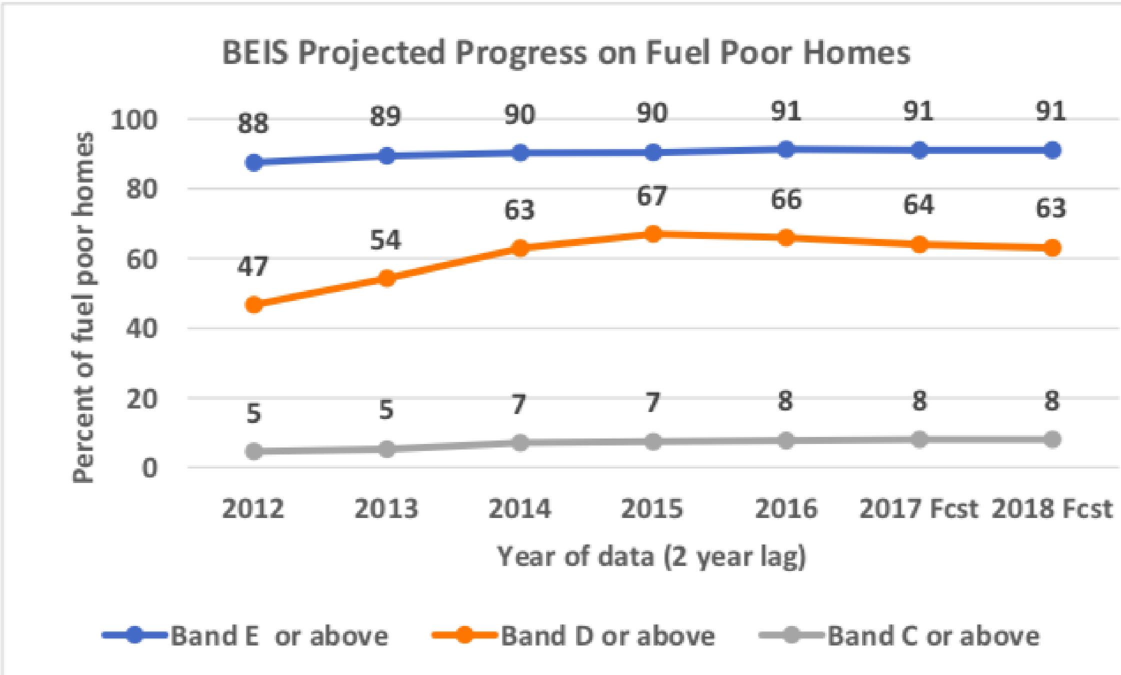


FIGURE 2-3 BEIS PROJECTED PROGRESS ON ENERGY EFFICIENCY IMPROVEMENTS IN FUEL POOR HOMES (Committee on Fuel Poverty, 2018)

This austerity led redefinition also meant a shifting of responsibility. Government schemes such as 'Warm Front' were concluded, leaving only supplier-led improvement schemes, making them entirely accountable for the delivery of energy efficiency measures to fuel poor households. This reduced the role of the state in supporting the fuel poor in favour of a model based on supplier obligations, funded via energy bills. The fuel poor become the subjects of the energy market, and certainly from a lived perspective, any gains through energy efficiency have been easily overshadowed by changing welfare policies and energy prices (Middlemiss and Gillard, 2015). The Committee on Fuel Poverty reported that not all households in fuel poverty will take any cost reduction from energy efficiency improvements as monetary gain and will instead trade-off for increased thermal comfort. This is typically dependent on their annual net household income and the pattern of comfort taking is described as "thermostat settings increasing until either the thermostat reaches 21°C or half of the financial gain is spent on additional heating cost, whichever occurs first" (Bridgeman *et al.*, 2018). The fact that some in fuel poverty systematically underheat their homes is hidden in the current definition of fuel poverty, as it uses a predefined required heating pattern to obtain an adequate level of warmth. Those vulnerable to welfare policy changes or in receipt of benefits are also likely to be reluctant to increase their consumption for fear of increased and unexpected costs which could lead to them returning to a state of fuel poverty, or faced with the "heat or eat" dilemma, where cold weather shocks are equivalent to income shocks and see low income households cut back on other necessities such as food in order to finance the additional cost of keeping warm (Beatty *et al.*, 2014).

As previously discussed, these austerity lead policies mean that only the most vulnerable or most impacted fuel poor subjects can have help meeting their needs, whether that means those with the largest fuel poverty gaps or those with physical vulnerabilities. This means that LILC households are not a priority, the income poor living in energy efficient housing are often overlooked by policy makers where there is no further benefit in pursuing energy efficiency measures despite the fact that Ofgem's 'Energy Supply Probe' identified that low income households were less likely to change tariffs, switch suppliers, compare offers, have the ability to access on-line offers and be more likely to be prevented from switching by existing debt, regardless of their energy costs (Middlemiss, 2017; Ofgem, 2009, pp.11, 59). This lack of engagement with fuel costs could lead to LILC houses being reclassified as LIHC and therefore fuel poor if energy prices rise significantly in the future. Low income homes are also more likely to have been given pre-payment meters or are unable to pay via direct debit; both of which incur higher costs, all of which amount to more undetected inequality in the LIHC indicator and need clarification.

The positioning of energy efficiency improvements as a key technology under the LIHC definition has implications for what is possible in fuel poverty policy and beyond. A focus on energy efficiency reduces attention to other structural problems which exacerbate fuel poverty, particularly fuel costs and pricing, and income inequality and suggests that reforms of this nature are beyond the realm of possibility (Middlemiss, 2017). This redefinition lends itself to a technical reassessment of the need for help, one which is related to the efficiency of the housing stock, and is far removed from the lived experience, leading to some households who do not fit this prescriptive definition remaining hidden (Middlemiss, 2017).

2.2.5 Socio-Economic Indicators of Consumption

Although the definition of 'energy' is variable, there remains broad agreement over those demographic factors influencing overall consumption (Frederiks *et al.*, 2015). These explanatory variables encompass three main fields: socio-demographic (e.g., income, education, household size, dwelling type and tenure), psychological factors (e.g., knowledge, values, attitudes) and external factors (e.g., economics, political and legal). The scope of this study means that whilst the external influences and drivers of domestic energy consumption have been acknowledged, only socio-demographic indicators are investigated.

Household income is often highly correlated with energy consumption (Wyatt, 2013). It is intrinsically linked to factors such as employment status, education and household size, all of which may facilitate or constrain energy related behaviours (Abrahamse and Steg, 2011; Rhodes *et al.*, 2014; Jones and Lomas, 2015). Those in full time employment and earning higher, more constant income being more likely to spend less time at home than those who are unemployed, and thus require lower levels of heat and light, but are also more likely to own and use more appliances than lower income households, as well as typically occupy larger properties with more rooms. Those who are unemployed or at home the day due to illness or caring responsibilities may be forced to reduce their consumption to a level of thermal discomfort in order to offset the extended hours of usage. Low income houses may find it harder to recover from unexpected energy expenses such as higher winter bills than those with a high income who typically have the disposable income to absorb an unexpected cost. Income also affects a consumer's ability to manage their energy accounts. Low income households are more likely to have been placed on a pre-payment tariff to prevent their account accruing debt, or may be explicitly restricted to pre-payment tariffs due to debt problems whereas higher earners are more likely to pay a fixed, predictable amount each month via direct debit; under-consuming in the summer months and building up credit to offset higher winter costs (Middlemiss and Gillard, 2015). The link between deprivation and pre-payment tariffs is widely reported and plays a significant role in consigning some people to perpetual fuel poverty (Middlemiss and Gillard, 2015; Longhurst and Hargreaves, 2019). They are relatively overpriced per kWh and do not allow for a credit and debt balance; if a household on a pre-payment tariff do fall into debt, in the short term they may have to live without heat, light and appliances, and in the longer term are likely to find themselves unable to move away from their provider to a cheaper, more affordable tariff due to their account balance (Middlemiss and Gillard, 2015).

Tenure has been shown to have an indirect effect on energy consumption. The landlord/tenant dichotomy, or the 'split-incentive' arises when the interests of both parties misalign (Bird and Hernández, 2012; Ástmarsson *et al.*, 2013). When the landlord provides the housing and the tenant pays the energy bills neither sees a benefit in making improvements to the energy efficiency of the home due to realising little return on their investment, making it difficult for tenants to have any autonomy over their energy usage and Hope and Booth (2014) found that 40% of the landlords in their study were deterred from making efficiency improvements to their rental properties because they saw no personal benefit (Ástmarsson *et al.*, 2013). Private rentals also represent the worst performing tenure type, with only 8% of homes obtaining an A-C energy efficiency rating (Hope and Booth, 2014). These represent the main constraints to the reduction of fuel poverty in privately rented accommodation in the UK and private renters are the most likely to be in the deepest fuel poverty (Ástmarsson *et al.*, 2013; Department of Energy and Climate Change, 2014c). Furthermore, due to the precarious nature of rental contracts, tenants find themselves disempowered and unable to request improvements to their living conditions for fear of reprisal or losing their accommodation – "Should a tenant be unhappy, a landlord can simply end the tenancy and install new tenants. There is a need for greater and clearer powers for tenants to request such improvements and mechanisms to ensure that landlords follow through without prejudice" (Hope and Booth, 2014, p.377). It is also the case that often, tenants do not know their rights or what they can expect from their landlord (Petrova, 2018). This is beginning to change, as with the "Minimum Energy Efficiency Standards" (MEES) which have been in place since April 2016, where a landlord can no longer refuse a reasonable request for improvement by a tenant, but progress appears on the face of it to be slow; properties need to be rated only a band E in order to be let to a tenant and the definition of 'reasonable' is highly subjective. Properties can also be exempt if a landlord refuses consent, and in reality MEES only apply to about 20% of rental properties due to the legislation not applying to social and local authority owned housing (Hope and Booth, 2014; French *et al.*, 2018). Conversely, home-owners are more likely to invest in energy improvements as they are less transient, more financially secure and more likely to benefit from long term savings; around 15% of owner-occupied homes have an A-C efficiency rating (Hope and Booth, 2014). They also possess the autonomy to make decisions about and make changes to their properties.

Characteristics of the physical dwelling have been linked to variations in occupant energy consumption. Dwelling type, age and size are influential, as well as the fixtures and fittings within the home (Wyatt, 2013). Depending on the number of features such as floors, rooms and windows as well as levels of insulation, central heating and ventilation, up to half of total household energy could be accounted for (Schipper *et al.*, 1989). Accommodation type in particular has been used as a proxy for usage habits as it can be representative of a family's life stage; a large family are more likely to occupy a home with greater floor area and more rooms and so will have a greater need for heating and lighting. They are also likely to require and use more appliances more frequently, for instance the washing machine and dishwasher, when compared to a single person living alone, but the cost of not having or not being able to afford efficient appliances should also be considered (Chapter 2) (Holloway and Bunker, 2006; Jones and Lomas, 2015).

2.2.6 The Effects of Fuel Poverty on Individuals

As discussed throughout, these socio-spatial characteristics are inherently linked to fuel poverty, yet the lived experience of being in fuel poverty is wholly overlooked by the government's strategies. The following section highlights the reported impacts of living in fuel poverty as a daily experience.

House quality, poverty, physical health and mental wellbeing are all outcomes of the condition of fuel poverty, and there is a cyclical risk associated with living in fuel poverty. Worsened physical health such as respiratory illness linked to dampness and mould are associated with sub-optimal mental health and the increased likelihood of stressors associated with being unable to afford solutions, which then lead to an increase in coping behaviours such as smoking and overeating (Mould and Baker, 2017). Breaking such a cycle and separating the ill health caused by the living conditions from the health conditions that are instrumental in the individual finding themselves in fuel poverty is complex. However, it is clear that long term physical disability can severely restrict the earning power of an individual and result in them living in perpetual fuel poverty as well as being disengaged from society more generally. Increased rates of mortality during cold weather (known as excess winter deaths or EWD) were first noted many years ago, and occur mainly due to changes in blood pressure and chemistry during cold weather, which in turn increase the risk of fatal cardio or cerebra-vascular events such as strokes or pulmonary embolisms (Liddell and Morris, 2010). The immune system is also suppressed, increasing the incidence of infections. Furthermore, studies have recently begun to examine the enduring and cumulative health impacts associated with living in sub-optimal conditions. These include increased risk of influenza, pneumonia and asthma (Liddell and Morris, 2010). As this broader spectrum of health impacts becomes more evident, preventable health impacts increasingly become the primary rationale for tackling fuel poverty in many parts of the world (Wilkinson *et al.*, 2007).

In the UK in particular, human health is construed as the main beneficiary of the Governments fuel poverty strategies, but the question remains as to whether or not policies which invest in actions so indirectly related to human health be expected to deliver significant health impacts through what is in essence a housing regeneration policy (Liddell and Morris, 2010).

2.2.7 Changes in the Fuel Poverty Vernacular

As discussed in the Literature Review, fuel poverty is currently defined by the Low Income High Cost (LIHC) indicator, introduced by John Hill in what is widely referred to as 'The Hills Review' (2012) (Section 2.2.4). This replaces the 10% indicator popularised by Brenda Boardman (1991) and makes several improvements on it; by measuring a household's income after housing costs have been considered and by making the cost of fuel a relative measure rather than an absolute threshold (Section 2.2.8). However, it still ignores the lived experience as a purely monetary based measure and is still too linear at a national policy level (Moore, 2012). The energy efficiency methods to reduce it also lend themselves to a technical problematisation, which foreclose alternative strategies and forms of intervention, entrenching the notion that fuel poverty is a linear problem affected only by low income, high energy prices and inefficient housing stock (Longhurst and Hargreaves, 2019).

Intuitively, a focus on energy efficiency and reducing carbon emissions from homes makes a great deal of sense, especially given the need in the UK to upgrade the housing stock to make it fit for the 21st century (Rosenow *et al.*, 2013). Many fuel poor homes are poorly insulated and investment in energy efficiency is a cost effective approach in both the long and short term as the benefits of home improvements remain for many years (Boardman, 2013). But as previously mentioned, it can be driven by many factors other than the traditional fuel poverty triad (Middlemiss, 2017). Acknowledging this multifaceted issue will lead to an improvement in how fuel poverty is understood and thus can be legislated for. To reiterate, Bouzarovski and Petrova (2015) and Middlemiss and Gillard (2015) identify the key indicators as; access, affordability, flexibility, energy efficiency, needs and practises, quality of building fabric, tenancy relations, energy cost and supply, stability of household income, social relations and ill health, given that in both cases "a change in any one of these elements or the relationship between them can materially affect a households access to affordable energy" (Longhurst and Hargreaves, 2019, p.2). Furthermore, the political landscape of austerity driven policy making increases peoples vulnerability to change; policies such as universal credit which on the surface are unrelated to fuel poverty compound the effects for those experiencing it by affecting a household's available income.

One area where the need for change, especially with regard to policy making, generates a highly charged discussion is the rental sector. When such fuel poverty characteristics are coupled with rental rather than owner occupied properties, there is a dichotomy between the obligation of the landlord and the tenant where neither will see any benefit to making substantial improvements to the home (Section 2.2.5); and is cited as one of the biggest barriers to improving energy efficiency in the rental sector (Hope and Booth, 2014). Landlords see little incentive to invest as it is their tenant who will benefit from the lower bills, and the tenant is neither inclined to invest in improvements as they won't live in the property long enough to see real financial reward, or are prevented from doing minor improvements through lack of consent from the landlord. Privately rented properties are some of the most likely to find themselves in fuel poverty and suffer from the worst energy efficiency (Hope and Booth, 2014).

2.2.8 Why Further Change is Needed

This change of definition and the implementation of its associated strategies, targets and indicators with a strong focus on energy efficiency creates a narrow interpretation, which is not reflective of the complex and multifaceted nature of the lived experience (Middlemiss, 2017; Middlemiss *et al.*, 2018; Longhurst and Hargreaves, 2019). The annualised statistics that guide the current policy frame fuel poverty as a problem of aggregate rates and trends rather than as a daily lived experience (Department for Business Energy and Industrial Strategy, 2018). This technical problematisation of fuel poverty entrenches it as one that can be solved by energy efficiency measures alone.

This exclusionary framework ignores other ways of 'knowing' fuel poverty, particularly those which relate to the household experience (Longhurst and Hargreaves, 2019). There is a growing body of work seeking to broaden the focus and draw attention to a wider set of socio-spatial factors than the traditional fuel poverty triad. Bouzarovski and Petrova (2015) identify access, affordability, flexibility, energy efficiency, needs and practises as key, whilst Middlemiss and Gillard (2015, p.147) focus on "quality of building fabric, tenancy relations, energy cost and supply, stability of household income, social relations and ill health", given that in both cases "a change in any one of these elements or the relationship between them can materially affect a households access to affordable energy" (Longhurst and Hargreaves, 2019). An improved understanding of the dynamic elements of fuel poverty serves to highlight the precarious nature associated with many experiences of fuel poverty, which are often obscured by the narrow, macro-level statistics. (Longhurst and Hargreaves, 2019).

2.2.9 Summary

To summarise, overall deprivation and fuel poverty are two distinctly different forms of hardship, however, it is the political definitions which are entrenching this divide. Material deprivation is a multifaceted phenomenon referring to an individual's inability to afford basic resources; a definition which moves away from an income based measure of poverty which permits a greater understanding of the causal mechanisms at work. Yet despite the fact that the populations living with deprivation and fuel poverty are very likely to be overlapping, the current definition of fuel poverty is very much income led, reducing the attention on non-monetary indicators.

The current, austerity led definition of fuel poverty changes the vernacular from a problem which can be eradicated into a problem of targeting the priority households. Framing fuel poverty as a technical problem linked to energy efficiency which can only be addressed by investments excludes the multidimensional and interrelated behavioural factors. In order to ensure that households who exhibit vulnerabilities do not become hidden because of this definition, changes are required which, as has happened with material deprivation, move away from an income based measure to encompass the many facets of fuel poverty and begin to address them.

2.3 Energy Data

2.3.1 Traditional Technology

Gas and electricity meters have been an essential but modest element of the energy infrastructure in the UK since the early 20th century - this arose from the advent of gas and electricity becoming available to the masses on a large, saleable scale (Darby, 2010). The most common modern electro-mechanical induction meters are, as previously discussed, mostly a variation of those and are still widely produced today. They are popular due to their low production price and excellent reliability; counting the revolutions of an electrically conductive metal disc (Ma *et al.*, 2017). Yet their lack of flexibility and responsiveness means that they are falling out of favour with policy makers, energy providers and consumers as they all look for a more efficient and affordable way to monitor energy consumption. There is also an increasing awareness of the 'creep phenomenon'; due to the mechanical nature of the conventional meters, physical wear and tear occurs which is difficult to avoid and leads to inaccuracies in consumption levels (Ma *et al.*, 2017).

Traditional meters require a large amount of manual calibration. To receive an accurate bill for usage, a householder must be available for a meter reader to come and take a meter reading periodically, which is costly for the supplier and time consuming for both (Darby, 2010). If this cannot be done, because for example the meter reader typically calls during the day when people are at work, or isn't done frequently enough, the user will receive a bill for estimated usage based on an average for the house type and previous months (Logica, 2007; Darby, 2008). This kind of billing could lead to unexpected expense (higher bills than usual if usage is overestimated, or a requirement to clear debts if underestimated over a long period of time) and fluctuating demand due to seasonality may push people who are ordinarily not fuel poor into short term fuel poverty while their finances recover (Ofgem, 2009; Robinson *et al.*, 2018b). Another billing method is to average out expected annual usage and bill the same amount monthly, allowing consumers to get into debt over the more expensive winter months, but holding credit that could be utilised elsewhere once the balance has been paid off in the cheaper summer months (Hazas *et al.*, 2011). This is however generally still preferable to a pre-payment tariff, as these can create genuine difficulty for people during the winter months when the ratio of fuel cost to income is generally much higher and there hasn't been an option to overpay (Moore, 2012).

As already discussed under the energy efficiency sphere, these meters are often installed in difficult to reach locations displaying usage only in terms of Kilowatt Hours rather than cost. From a fuel poverty perspective these traditional meters make it very difficult for households to interpret their usage in terms of cost, making identifying and altering particularly expensive practises very difficult. It was the government's intention that a full transition away from these first generation meters to smart metering would be complete for both gas and electricity by 2020 but this has since been extended several times to 2025 (Haben *et al.*, 2016). The benefits of upgrading these meters to new smart meters are discussed more fully in the next section, but it is considered instrumental for improvements not only in high level emissions targets but also fuel poverty ones.

2.3.2 Traditional Classifications

Traditionally households are simply classified by the energy suppliers as low, medium or high consumers, with only their average consumption levels to categorise them, no recommendations for improvements and no further guidance on their energy practices. This may be down to the suppliers, who wish to keep their marketing strategies private to retain a competitive advantage but is ultimately unhelpful in helping increase peoples understanding of their energy consumption as the ranges in the groups are so wide, as is shown in Table 2-2.

TABLE 2-2 STANDARD CATEGORISATION OF ENERGY USERS AND THEIR AVERAGE ANNUAL CONSUMPTION

Energy user group	Electricity (kWh)	Gas (kWh)
High	4,600	17,000
Medium	3,100	12,000
Low	1,900	8,000

According to USwitch (2018), an average 'low user' typically has a small one or two-bedroom property, where a single person or couple spend little time. They use their washing machine around once a week but do not own many other large inefficient appliances such as a tumble dryer or dishwasher. A medium user will live in a typical three bedroom house with their small family or partner and one or two children. They are likely to spend time at home in the evenings and at weekends but are generally out during the day. They have more appliances than low users for convenience, which are used a couple of times a week (washer dryer, dishwasher etc.) and they also have slightly higher usage based on the number of electrical devices in the house. A high-energy user lives in a large property, probably with more than 4 bedrooms. This can either be a large family home or a shared house of multiple occupancy, and the house is likely to be occupied by at least some of the occupants most of the time. Multiple appliances are in daily use. Figure 2-4 provides this information as a digestible infographic, aiming to educate users to ensure they choose a tariff which is right for their level of usage. (USwitch, 2018).

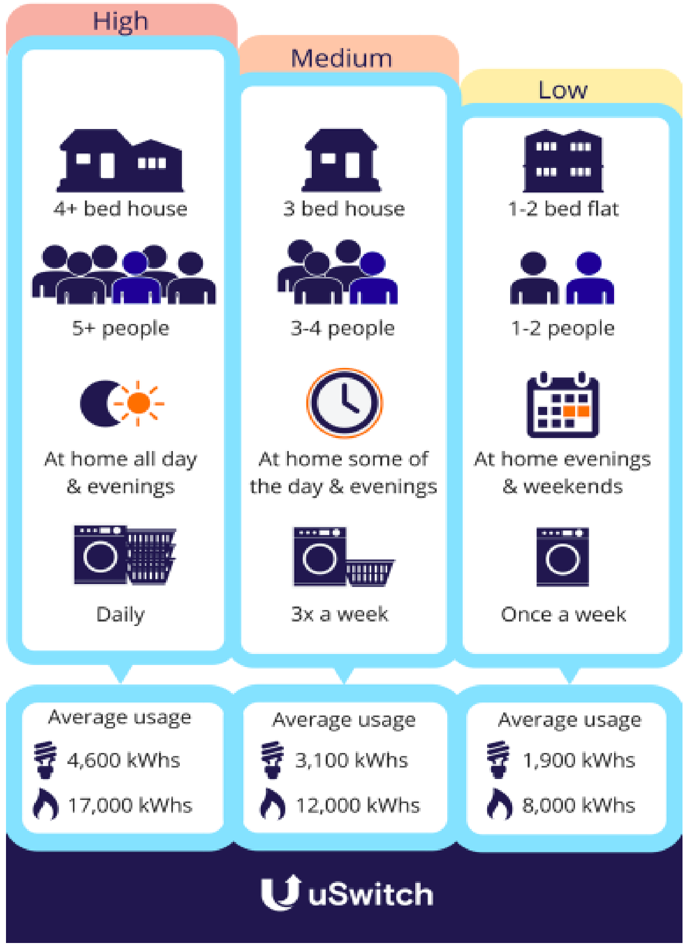


FIGURE 2-4 ENERGY CONSUMPTION CATEGORISATION INFOGRAPHIC (USwitch, 2018)

2.3.3 Improved Technologies

Improved technologies go beyond the meter that is present within the residential property and encompass an entire cultural shift to an Internet connected smart grid. A smart grid can be defined as an electric system that uses two way, secure communication technologies to provide near real-time information on every aspect of energy generation, delivery and consumption to achieve a system that minimizes environmental impacts, enhances markets, improves reliability and service, whilst also reducing costs and improving efficiency (Gharavi and Ghafurian, 2011; El-Hawary, 2014). The previous section on energy efficiency cited some of the benefits of an interconnected smart grid to the supplier in terms of energy security; better planning of resources; faster resolution to outages and improved reliability (Guerreiro *et al.*, 2015; Haben *et al.*, 2016). Table 2-3 describes in greater detail the benefits to suppliers, consumers and wider society.

TABLE 2-3 BENEFITS OF AN INTEGRATED SMART GRID (HOUSE OF COMMONS SCIENCE AND TECHNOLOGY COMMITTEE, 2016)

Stakeholder	Benefit
Consumer	
	Easier switching between suppliers
	More accurate billing; the avoidance of billing issues and the need for meter readings
	Avoidance of debt accumulation through access to accurate, near real time information
Utilities and Energy Providers	
	Removes the need for site visit meter readings
	Reduces call centre traffic through reduced queries
	Improved theft detection and debt management
Society	
	Benefits of optimised electricity generation and network management
	Network reinforcement and electricity storage
	Technical innovation and new economic opportunities
	Reduced carbon generation and meeting climate change targets

Smart grids mark an enormous cultural shift in the UK energy sphere; the largest change to the UK energy market since the shift to North Sea gas (Darby, 2010). They are characterised by improved communications and two way feedback throughout the generation, distribution and consumption processes, all of which must take place in real time. Figure 2-5 illustrates the feedback loop that a fully operational smart grids could enable and gives an overview of the extensive technologies required.

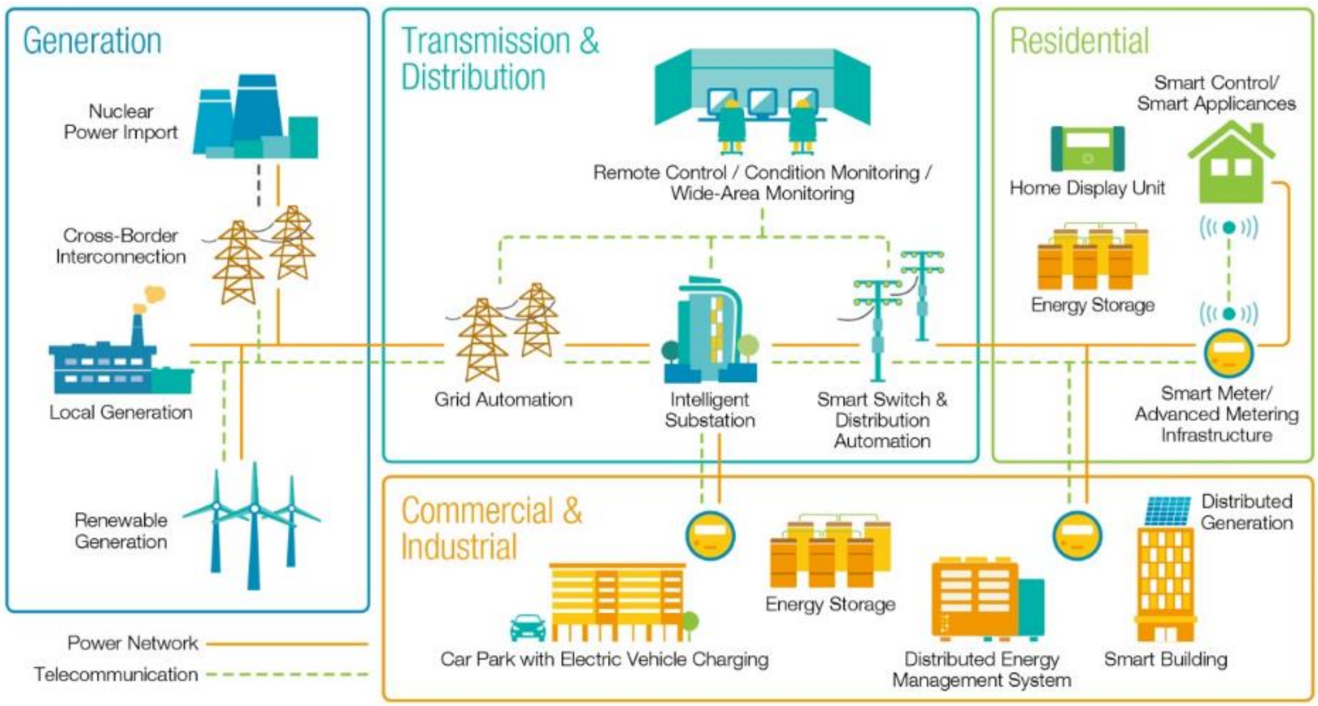


FIGURE 2-5 A FULLY OPERATIONAL SMART GRID SYSTEM (El-Pro-Cus, 2019)

From a policy perspective tackling 'energy poverty', upgrading the UK infrastructure to a 'smart grid' gives the potential for solving many energy problems (Stern, 2011; Smart Energy GB, 2018b). This upgraded infrastructure leads to more efficiency in the short term (better matched supply and demand and less wastage) and planning for long term futures (increasing energy security, planning for unexpected power outages and informing the number of power stations the UK is likely to need in the future) (UK Committee on Climate Change, 2010; Guerreiro *et al.*, 2015; Smart Energy GB, 2018b). In the UK, smart meters have come to mean meters that can both measure and store data at regular intervals and act as the point of communication between supplier and consumer in the domestic setting, using an inbuilt wi-fi signal (Guerreiro *et al.*, 2015; Smart Energy GB, 2018b). Smart meters have been defined as "advanced meters that identify consumption in more detail than conventional meters and communicate via a network back to the utility for monitoring and billing purposes" and link to a wider scheme of upgrades throughout the UK allowing for energy to be more efficiently produced, stored and planned for to meet demand (Faiers *et al.*, 2007; UK Committee on Climate Change, 2010; Smart Energy GB, 2018b, p.85). During the initial stages of the UK rollout programme meters were installed into homes but were not without technical difficulty. The SMET1 meter was only ever supposed to be installed during a testing or foundation stage and would frequently "go dumb" in areas of poor mobile reception and failed to work in 30% of properties, typically those with thick walls or in high rise buildings. A smart meter is installed with an accompanying In Home Display (IHD) which refers to the device or monitor that connects to the smart meter and provides the consumer with visual information about their consumption levels and associated costs (Sovacool, 2015).

The roll out programme in the UK is viewed by policymakers as integral to encouraging greater efficiency inside homes and achieving the proposed emission reduction targets. For consumers this is because IHDs will enable education and awareness of consumption, which are considered some of the greatest barriers to sustained energy reduction, and on the supply side, as mentioned above, for the integrated feedback loop. However, despite the benefits; there are technical challenges associated with getting smart meters into every household in the UK, involving installing (at least) a combined 104 million pieces of new equipment when counting gas and electricity meters, IHDs and wireless communication networks (Lewis and Kerr, 2014). It is expected to cost more than £200 per household and at least £11 billion in total, representing a complex and costly scheme, much of which will be paid for through customer's bills.

There are also societal barriers to uptake which must be considered if the programme is to succeed; the scheme is argued to be the biggest "behavioural change that this country has ever seen" (House of Commons Science and Technology Committee, 2016). It is widely regarded as "an incredibly tough job convincing every household in England, Wales and Scotland to install a smart energy meter" (Barnett, 2015, p.3). Buchanan et al (2016) provide a detailed overview of the perceptions of the British public and found that in general they appear to be "apathetic or ambivalent", with 53% indicating that they are undecided about whether they should be installed in every UK home (Department of Energy and Climate Change, 2014b). Some key themes surrounding both the perceived risks and opportunities were identified by consumers; risks and worries were overwhelmingly associated with mistrust of the energy suppliers and suspicions that the consumer was unlikely to be the main beneficiaries of the smart meter scheme. This is likely to act as a barrier to their willingness to adopt these new technologies as they struggle to understand the (profit making) suppliers rationale for reducing consumption (Fell *et al.*, 2014; Lopes *et al.*, 2014). Privacy was also a major concern; respondents to studies on smart meters frequently feel they are being watched over by an invasive presence (Fell *et al.*, 2014; Sovacool, 2015). People are however convinced of the benefits offered by dynamic billing, the opportunity to save money and the possible improvements to personal comfort (Buchanan *et al.*, 2016; Darby and Pisica, 2013; Smart Energy GB, 2014).

2.3.4 Improved Classifications

The advent of improved metering technologies means that many new tariffs are already on the market, which are based on personalised energy use to a lesser degree than what smart metering can offer. New generations of meters preceding smart metering include white meters, more commonly known as 'Economy 7' in the UK (USwitch, 2017). These are already breaking consumers down into more precise and therefore targetable groups than simply "high, medium or low" as far as energy suppliers are concerned. This is beneficial to supplier marketing - making it easier for them to aim their tariffs at specific types of users and provide tariffs that meet the needs of many more people.

The 'time of use' tariffs available to those who have a white meter installed in their home offer a different way of consuming energy, with different tariffs tailored for usage, based on each supplier's peak and off peak times. They encourage users to shift their consumption to times when energy is available more cheaply but charge a higher than average price for use at peak times (USwitch, 2017). These tariffs are positive from an environmental and energy security point of view, reducing peak load on the grid, and therefore power plant capacity (Energy Saving Trust, 2017). It is fair to say though that these tariffs would not be suitable for everybody and some may even find themselves worse off - it is important that both family life and the infrastructure of the house are right for the tariff; most of the energy usage should already take place at off peak times, i.e. those who are retired or home workers would find it easy to shift their usage away from the evening peak, but a young working family may not. The house should also be equipped with a storage tank to store hot water that is heated cheaply, and appliances that can be set on a timer to come on at night to take the most advantage of the cheapest overnight rates.

There are also now 'green supply tariffs' which ensure that at least some of the energy that you buy is 'matched' by the purchase of renewable energy on your behalf (Energy Saving Trust, 2017). There are also 'feed in tariffs' available to anyone who has installed or is looking to install a variety of small-scale renewable and low carbon electricity generation technologies (such as solar panels) where a payment is made on a quarterly basis for any unused energy fed back into the grid to encourage their uptake (Ofgem, 2018). None of these innovative tariffs would be possible without a network that provides a two-way relationship, feeding information to both the consumer (to tell them when energy is cheaply available or what they stand to gain by providing energy) and to the supplier, to inform them when people are able to sell them their surplus energy. Once again, these offer a huge environmental and energy security benefit, reducing grid load and power plant capacity (Energy Saving Trust, 2017).

It is anticipated that once the smart metering roll-out is complete, there will be an end to many of the traditional estimated tariffs, as everybody will be able to monitor their consumption in almost real time and revise their energy usage habits to not just a time when it is cheaper, but ultimately become more aware of their consumption and reduce their usage in the long term. However, it is important that there is enough education and access to information to ensure that people are still getting the best deal for them as every household is different.

The advent of these new tariffs is also beneficial to the suppliers; incentivising customers to switch their energy usage times will minimise the peak usage increase and decrease power plant and grid load. This detailed feedback to the supplier will also enable the improved detection of fraudulent activity or tampering (Darby, 2010; Ma *et al.*, 2017)

2.3.5 Improved Understanding

It has been shown in studies such as Faruqui, Sergici, and Sharif (2010) and Ehrhardt-Martinez and John (2010) that providing visible feedback to smart meter users through an IHD gives tangible results with regard to reducing energy consumption with an average of 7% across studies. IHDs transform a once static and incomprehensible energy bill into a dynamic, transparent and most importantly controllable process by translating kilowatt hours (kWh) into pounds and pence (Faruqui *et al.*, 2010). This feedback in familiar monetary terms can be seen as an educational tool, allowing homeowners to experiment with appliance usage and alter their most expensive practices, for example, switching their washing machine from a 60° cycle to a 30° cycle or adjusting their thermostat by one or two degrees (Darby, 2010). The level of effect the feedback has on consumption is however linked to a number of pre-existing factors such as users' values, beliefs, norms and capabilities; for example, there may be a level of comfort that those who can afford to will choose not to go below, and so in the long term there may be a levelling out of energy savings as people make the choice between consumption and comfort or convenience. This direct feedback has been shown to reduce demand by almost double the 7% saving when a pre-payment meter is installed, suggesting that monetary savings are the biggest motivation to those on the most expensive tariffs (Carroll *et al.*, 2014).

Conversely it has also been suggested, that this feedback can be overwhelming if the correct pre-requisite understanding with regard to both the new technology and consumption practises is not there, and so there is a need for education alongside new technology, not just on how to reduce energy but also on the technology itself (Darby, 2010). Lack of understanding leads to mistrust, which in turn translates into disengagement with the scheme and no real savings in the long term as people feel discouraged from changing their habits (Oltra *et al.*, 2013).

This information also allows the supplier to deliver more accurate billing and a long term cost saving in terms of a reduction in staffing costs as the job of 'meter reader' becomes obsolete, although it is open to debate how much of the saving will be passed onto the consumer, especially given the increased burden of employing skilled tradespeople to undertake the installations for the foreseeable future (Roberts and Redgrove, 2011).

From an academic perspective this new source of highly granular data has utility in dissecting the diurnal patterns of people and their lived environment and experience. Energy is closely linked to daily routine, most acts within the home impact on energy usage in some way, from making a morning coffee, to relaxing with a games console or taking a shower (Buchmann *et al.*, 2013). Given that the stages of a family's life cycle (a combination of criteria such as family members' age, marital status, and size/type) appears to be one of the strongest predictors of household energy consumption, this passive logging of energy consumption could provide valuable insight into household demographics; this idea is discussed at greater length in Section 2.4.4 (Frederiks *et al.*, 2015).

The dataset utilised in this thesis is the most comprehensive available; studies exist which utilise Smart Meter data, however they are often limited by one or more of the following; sample size, timeframe, meter type or cadence, usually as a factor of accessing commercial data sources. Sample sizes in existing literature range from 180 households in research by Buchmann *et al.*, (2013) through to 225,000 households examined by (Kwac *et al.*, 2013), however even a study of this size was limited by only having access to data with a two week timeframe. Carrol, Lyons and Denny (2014) also conducted research using smart meter data with a sample of 5,000 home s over 18 months but were limited to electricity meters only. Further research exist which attempt to estimate energy consumption without the use of smart meter data, which necessitates a reliance on annualised statistics. For example Druckman and Jackson (2008) have used 'spend on fuel' from the Expenditure and Food Survey as a proxy for consumption. Whilst Jones and Lomas (2015) do use primary consumption data, it is collected from traditional meters 3 times a year from 315 homes, which takes significant time and effort.

2.3.6 How Smart Meters May Benefit Those in Fuel Poverty

The interconnected nature of the smart grid will benefit those households in fuel poverty in a multitude of ways. Firstly, the direct feedback to the energy provider will see the end of estimated billing and monthly aggregate usage; for the consumer this means no unexpected bills as a result of the provider incorrectly estimating a household's usage, making it much easier to plan and budget for energy costs throughout the year. They also facilitate the introduction of time of use and demand-response tariffs, both of which help smooth consumption throughout the day and reduce peak demand, promoting the use of off peak consumption by offering preferential rates (Carroll *et al.*, 2014). This increased visibility over their consumption is compounded by the adoption of IHDs; the UK is the only EU country that has stipulated that a smart meter must also be fitted with an IHD (Sovacool, 2015). The IHD translates consumption into monetary terms, highlighting particularly inefficient practises and acts as a reminder to pay more attention to consumption levels. As discussed prior, the addition of a smart meter, coupled with an IHD can result in an average reduction in electricity consumption of up to 11% depending on the time of day (Faruqui *et al.*, 2010; Lynham *et al.*, 2016), thereby helping to narrow the fuel poverty gap. However, studies find the reduction declines over time, suggesting that users see the IHD as a novelty at first and interact with it much more frequently, leading to short term changes in behaviour, which do not necessarily translate into long lasting habits (Lynham *et al.*, 2016).

As discussed though, it is also important to note that those of lower socio-economic status, those with no formal qualifications and households of more vulnerable groups (who are also the most likely to find themselves in fuel poverty) are the least likely to engage with IHDs therefore realising the least benefit and so there must be a concerted effort in outreach and education within these marginalised groups. Due to their increased propensity to also be housed in rented households and be placed on pre-payment tariffs, the installation of a new physical meter is an additional barrier to their access.

2.3.7 Summary

The commercially sensitive nature of smart meter data means that many existing energy studies are based on small samples, or utilise surveys focusing on the end user experience, apart from one notable exception of Brounen, Kok and Quigley's (2012) study of 300,000 Dutch homes. Results from attitude surveys are limited in their ability to discern the complex correlations between dwelling characteristics, occupancy behaviours and consumption (Yohanis *et al.*, 2008). The innovative nature of the technology and limited schemes in place also invariably dictate and limit the scope of existing studies.

The technological changes within the energy sector are enabling improvements across the board. It is the belief of governments and policy makers as well as academics studying the behaviours linked to smart meters and IHDs that they can have an effect on reducing residential energy demand with benefits in the form of financial reductions and greater control to the household and reduced carbon emissions and increased environmental improvements to society. Studies such as Carroll et al (2014) suggest that smart metering is effective as it acts as a motivator and a reminder to the consumer by increasing the visibility of consumption within the household. It is clear however, that for short term changes to become habits, smart meters and IHDs must overcome those barriers the consumers have vocalised; lack of information, mistrust and challenges in accessing the technology to name a few.

2.4 Big Data

'Big Data' have received much attention commercially, in the media and in academia, yet formal definitions in literature differ wildly. The most common is Laney's Three Vs - Volume, Velocity and Variety (Laney, 2001). Other 'Vs' have been added into this definition over time with the most common (value and veracity) described below.

- **Volume** - consisting of terabytes or petabytes, yet also references the often vast dimensionality of the data.
- **Velocity** - data are often collected continuously and have a high, often second by second, temporal resolution.
- **Variety** - can reference intra-data variety (the diversity of information in a single given dataset), or inter-data variety (the vast number of datasets that fall under the big data umbrella).
- **Value** - refers to the value the collected data can bring to the intended process.
- **Veracity** - reference to the uncertainties surrounding data quality, which can be influenced by several factors including data origin and collection or processing methods.

It is well known that academia, government and industry have long been collecting large amounts of population data such as censuses, and so it is not necessarily the size of the datasets which primarily defines big data but the way they are generated. It is 'velocity' that sets big data apart from conventional data repositories and infrastructure as they are produced through automated continuous systems with a high refresh rate as against the tightly controlled, manual and sampled data we are used to (Miller, 2010; Kitchin, 2014b). Because of the way in which these data are produced, large portions are georeferenced, giving insight into spatial trends (Goodchild, 2013). This is another way in which big data can be conceptualized and includes directed, automated and volunteered spatial data:

- **Directed** - generated from digital forms of surveillance on a person or place by a human operator, such as passport control or CCTV.
- **Automated** - generated as an automatic function of a device or system such as scanning travel passes, interaction with websites, retail transactions and weather sensor data.

- **Volunteered** - generated by volunteered interactions, such as from social media or crowd sourced data. Examples include Flickr, OpenStreetMap and Twitter.

Currently, most georeferenced big data are being generated through location based services such as mobile devices but there are also spatial referencing systems such as residential postcodes and georeferenced sensors (Laurila *et al.*, 2013) . As a by-product of this, the representation of daily interactions such as work, leisure, communication, consumption and travel are now unprecedented.

All of this has coincided with computational and technological advancements, which have led to a vast transformation of the data landscape in recent years. Of interest to this thesis is the potential to develop a deeper understanding of the population, given that it is now a common attestation that information derived from big data is one of the foundational elements for understanding future societies, across a broad spectrum of social, political, economic and environmental processes (Einav and Levin, 2014; Graham and Shelton, 2013).

The emergence of big data has facilitated a paradigm shift to what has been termed 'data-driven science'. It is changing how knowledge is produced, business conducted, and governance enacted (Bollier and Firestone, 2010). Successful analysis of big data requires a realist approach, which allows for a greater degree of flexibility in the interpretation of results. These are more likely to be extracted through exploratory, rather than confirmatory techniques, generating insights which are 'born from the data' as against 'born from the theory' and has been coined the 'fourth paradigm of science" (Kelling *et al.*, 2009; Kitchin, 2014a).

Graham and Shelton (2013, p.259) review in greater detail what big data means for geography, and apply the ideas of 'data-driven science' to 'data driven geography' and suggest that the fears about spatial inequity in representation and self-selection biases, as well as "barriers to research and their implications for governance, privacy and our way of knowing the world" will outweigh the hopes that geographers will be able to utilize big data to influence and address "long standing questions of social injustice, inequality, and our relationship with the environment". Even though there is little theoretical groundwork (and many complexities to using big data), academics and commercial businesses are beginning to realise the value that big data holds for new ways of explaining the world, especially since the future of some traditionally used data sources cannot be guaranteed (Singleton and Spielman, 2014).

2.4.1 Bias, Practicalities and Governance

Interpretation is at the centre of data analysis. Regardless of the size of a data set, it is subject to limitation and bias. Without those being understood and outlined prior to analysis, misinterpretation is the result. Predefining these caveats also provides an opportunity for those consuming the analysis to make an informed decision about the trustworthiness of the results. Big data is at its most effective when researchers take account of the complex methodological processes that underlie the analysis of social data (Boyd and Crawford, 2012).

Because of the emerging nature of the big data field, there is a degree of theoretical uncertainty. Representativeness is one of the fundamental areas of this uncertainty, referring to how well the data capture the case they seek to represent and how well that represents the overall population (Kitchin, 2014a). Despite their volume, these data are still a sample. If traditional data can be said to suffer from sampling error, then big data equally suffers from sampling bias. Big data are inherently biased due to the nature of their production; to be included in a dataset, regardless of its size, one must fall within the target population of whatever it is that is being tracked and collected. Particular effort must be made, sometimes through efforts of triangulating novel data with more conventional methods like the Census, to ascertain the representativeness of the big datasets to the behaviours of the general public to make findings trustworthy, and also to avoid the pitfalls of generalisation (Lansley, 2014; Longley *et al.*, 2015). Despite these new biases to consider, it is worth noting that data collected by machine methods are generally automatically and passively collected, which can prevent survey biases like those generated from response and non-response effects (Lenormand and Ramasco, 2016, p.362).

The detailed interactions across space and time captured at such a high granularity in big data pose a substantial ethical and legal consideration. Given the often sensitive nature of these datasets including information at an individual level, access must be tightly controlled. There are fundamental challenges to be faced when considering the integration of these data into social science research, in the forms of legal, ethical and data uncertainty.

There are various definitions regarding data and information governance. For instance, the Data Governance Institute defines data governance as a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, when, under what circumstances and using what methods (Data Governance Institute, 2014). Soares (2012) succinctly recasts this to accommodate for big data:

"Big data governance is part of a broader data governance program that formulates policy relating to the optimization, privacy, and monetization of big data by aligning the objectives of multiple functions."

To unpack this definition, big data governance needs policy which finds the balance between competing objectives to determine whether the potential gains from new findings outweigh the associated risks to both regulation and reputation (Morabito, 2015). Correct data governance policies should provide a framework for setting data usage rules as well as implementing controls designed to ensure that information remains accurate and consistent (Morabito, 2015). By controlling the creation, sharing, cleaning, consolidation, protection, maintenance and integration of information, data which previously would have been uncertain and meaningless become valuable and insightful. This does, however, need to be backed up by its underlying metadata, giving context to content and building useful inventories of big data to ensure the contained variables are correctly interpreted (Morabito, 2015).

The recently imposed General Data Protection Legislation (GDPR) is likely to have a profound impact on the future of big data in the UK, EU and beyond and there are many ways in which it could be breached, with legal implications and sometimes severe monetary sanctions. It is considered the most comprehensive and forward looking piece of legislation addressing the challenges facing data protection during this data shift. It replaces the previous legislation - the 1995 Data Protection Directive - and could substantially alter the way big data is collected and analysed (Zarsky, 2017). It is well reported and reflected in central policy that big data analyses of population based datasets generate substantial societal benefits, and as a meaningful framework, it is here to stay, but affects and is affected by the extent of data protection policy. Privacy legislation in particular is increasingly debated; the new data era has shifted the goal posts with regard to exploitation. It is commonly reported that people, when signing a privacy policy on the ways in which their data will be generated, stored or disclosed, very often do not understand or even read the policies before signing them and serve as little more than a liability disclaimer for the companies than as an assurance of privacy (Polonetsky, 2012).

Ethically, privacy is considered a basic human right and there is a need to recognise acceptable practices regarding the access and disclosure of personal information (Elwood and Leszczynski, 2011). Under the GDPR this includes anything related to an identified or identifiable individual and could be as simple as a name or number, but also include other identifiers such as IP or email address (Information Commisioners Office, 2018). Age, gender, political opinions and criminal activity also constitute personally sensitive data and should be treated as such. The way this is handled currently is through a series of anonymisation processes such as de-identification, aggregation and physical computational techniques such as encryption, secure storage and restricted access (Kitchin, 2014b).

From a research perspective, it is important to consider that many commercial entities want to keep their data restricted to retain a competitive advantage, which has a significant impact on the ability of academic researchers to realise the data's full potential. The appetite for research partnerships is slowly changing though, as both parties realise the benefit of combining analytical expertise for public gain. These are commonly done between a corporation with data they are willing to share and universities, sometimes directly through Application Programme Interfaces (APIs) or through a trusted intermediary such as the ESRC (Economic and Social Research Council) funded Consumer Data Research Centre (CDRC).

2.4.2 The Practicalities of Working with Big Data

While big data can yield exceptionally useful and valuable information, they also present new challenges with regard to how much data to store, how much this will cost, whether the data will be secure, and how long it must be maintained (Soares, 2013). For example, inaccurate, incomplete or fraudulently manipulated data pose increasing risk as enterprises become more dependent on the data to drive decision-making and assess results. Considerations are thus:

- **Quality** – This includes considerations such as accuracy, completeness, vagueness, ambiguity and precision as well as consistency, scale, coverage, sample size and bias. (Wang *et al.*, 2005; Harris and Jarvis, 2014)

- **Quality Control** – Because of the novel production process of these new datasets, aspects of this are often unknown and data treatment must be undertaken before their dynamics are understood. A lack of data to reference against can impede this process and make it difficult to prove the accuracy of the data and therefore the results generated from it.
- **Errors** – Some errors are easily identifiable; spelling or syntax mistakes for example, but others may be harder to detect, especially if the dataset is not corroborated with sufficient metadata, giving the expected content of each variable. Spatial and temporal errors may be hard to detect, as admissible but incorrect data is provided possibly through an entry, coding or assignment error, or in terms of temporal data when objects change between the time of data collection and data utilisation. These errors have the potential to obscure pattern and processes rather than reveal them (Graham and Shelton, 2013).
- **Analytical** – Practical challenges are also faced in the age of big data. Traditional analysis techniques are no longer adequate as the volume of data leads to them being prohibitive in both time and expense (Levy and Lemeshow, 2013). Inferential statistics have become irrelevant tools in the 'populations not samples' debate and there is therefore a need for novel, exploratory methodologies to address these data challenges. Commonly used algorithms must also be scalable in order to operate on big data efficiently.

From a geographical perspective, it is important to remain critical of the patterns shown in big data. For example, Miller and Goodchild (2015) discuss three major challenges for data-driven geography; 'populations not samples'; 'messy not clean data' and the issue of 'correlation not causality'.

- Populations are problematic from an analytical viewpoint as traditional techniques tend to be confirmatory rather than exploratory, used with a specific question in mind, where the samples were under the control of the researcher (Miller, 2010). However, data-driven methods call for exploratory techniques to gain descriptive insight into large populations.
- As previously stated, big data are not usually collected for a specific purpose, rather are a by-product of various processes and generally are used to examine topics not connected to their original purpose. Their inherently messy nature means that new methods must be used to clean and understand them before they are used for analysis.

- The new data-driven science advocates correlation over causality, looking to identify and observe relationships rather than look into the causes of the phenomenon. It is important to remember though that correlations can be random in nature, especially in wide datasets containing lots of variables (Kitchin, 2014a).

Chen, Mao, and Liu (2014) and Gandomi and Haider (2015) list further practical challenges including the fact that data which refreshes so quickly typically has a very short 'use by date', making it irrelevant almost before it can be processed, the energy management relating to physical storage, processing and computational power required to cost-effectively gain insights; scalability of the algorithms and co-operation across many different disciplinary fields. From an analytical perspective, Miller and Goodchild (2015) present concerns that by moving towards a data driven geography, the lack of appropriate confirmatory techniques limit the viability and confidence in the results gleaned from exploratory methods. It is fair to say that these exploratory methods are still emerging and lack a cohesive framework, meaning the researcher must be clearer in their rationale for employing these techniques in order to retain legitimacy in their findings.

2.4.3 Consumer Generated Big Data

There is no doubt that the data generated by consumers fits the definition of 'big data'. Consumer data is that which arises out of transactions between individuals and service organisations. However, there are tensions between the corporations which hold the data, and the non-commercial researchers who want to add value to it. As mentioned, data such as these have long held a competitive advantage for the organisations producing them, giving detailed insights into what, where and how often people consume or interact with their product or service. Examples of consumer data include, but aren't limited to; online ordering, store transactions (usually collected through loyalty cards), public transport usage, and as is the case here, energy consumption in the home. This variety of data and the fact that they are continuously collected, capture entire consumer populations and include temporal, longitudinal and geographical dimensions make them particularly attractive propositions for looking at societal patterns, but can also be highly sensitive and be personally identifiable.

2.4.4 Energy Data as a New Form of Big Data

Smart meter data can mostly be considered as 'big data'; characterised by providing detailed and disaggregate information without the need for routine survey collection, one meter reading every half hour generates in excess of 48 million readings a day, per million customers (Alahakoon and Yu, 2013; Longley *et al.*, 2018). The 'velocity' aspect of big data is of particular interest to energy researchers, because of the need to respond quickly to real time events such as equipment reliability or security monitoring, outages and surges in demand. Of the analytical algorithms available to process this huge quantity of data, many are unable to complete the tasks in a time span to make them practicable to implement the results. But, insights can be used to predict future events big or small from power outages to surges in demand caused by the fluctuations of everyday usage; for instance a novel example known as 'TV pickup' where a break in a popular television program causes a huge power surge as people collectively move away from the television to do other things[2] and whilst this is minor, it is important that events such as this can be predicted as the UK aims to move toward a responsive smart grid, generating only what is required.

Ardakanian et al. (2014) argues that generating consumption profiles is one of the fundamental data mining operations achievable through smart meter data - using household features captured through the profiles to understand different categories of consumers. Despite all the obvious benefits to both suppliers and consumers of being able to monitor energy usage in real time, there are also some significant arguments to be made in the case of user privacy and the safety of their data. There are studies which look at the potential for re-identification of anonymized individuals, such as Buchmann et al. (2013), who finds that smart meter data is inherently identifying, and there are elements of external and internal invasions of privacy that need to be considered and handled with sensitivity as the technology becomes more widespread and improves in quality and granularity (McKenna *et al.*,

---

[2] When Lisa admitted to shooting Phil in EastEnders 2001, an estimated 22 million viewers tuned in. When it was over, they caused a surge of 2,290MW (916,000 kettles worth), more than five times the normal pickup of 400MW seen at the end of an average EastEnders episode. During the England vs Brazil World Cup Quarter Final in 2002, despite the early morning, half time saw a surge of 2,570 MW (1.1 million kettles) (Drax, 2016)

2012; Guerreiro *et al.*, 2015). There have been media outcries regarding privacy concerns ranging from the illegal to the controlling, which in turn has affected the consumers trust of smart metering, hampering the roll out (Jawurek *et al.*, 2011). In Holland the mandate to require every home to have a smart meter was retracted due to it violating citizens' rights to privacy. It is now being done on a voluntary basis, with a much-decreased uptake (McKenna *et al.*, 2012). With data from smart meters available at such high granularity, some users are rightly concerned about the ability of multiple actors to access and misuse their personal data, being able to find out in great detail about their daily habits. For example, a criminal may infer when they are likely to be sleeping, out at work or have gone on a holiday and left the house unoccupied (McKenna *et al.*, 2012). There are concerns over suppliers profiting from the sale of customer information and consumption profiles to third party marketers, for insurance purposes; premiums could increase if the insurer feels you are underrepresenting the number of appliances in a household and so forth (Jawurek *et al.*, 2011).

From an internal perspective, increased visibility in the home and a clearer record of energy usage could lead to tensions between tenants and landlords for example. If a landlord investigates the tenant's usage and deems them to be consuming energy excessively, then they may decide to investigate whether or not they are subletting, or over-occupation is occurring in the property. In recent news, a landlord in London covered a thermostat in their property with a locked cage to stop tenants adjusting it to what they deemed to be an excessive temperature, and currently no laws exist to prohibit this behaviour (BBC, 2019). It may also lead to less obvious but still very real levels of personal intrusion; for instance, a domineering family member "spying" on others activity in the home. Other examples include stalkers tracking movements of victims or the police using energy consumption during law enforcement; verifying claims that people were where they stated they were, or leaving a child at home alone (Hargreaves *et al.*, 2010; Lisovich *et al.*, 2010).

However, it also affords the supplier greater control over their customers in a beneficial way; the smart meters become more resistant to fraudulent behaviour such as suspiciously high or low consumption, and evidence of tampering becomes more visible and these cost savings can be reinvested or passed on to the customer (Darby, 2010; Ma *et al.*, 2017). It may also aid the detection of illegal activities such as drug production and sweatshops (Lisovich *et al.*, 2010).

2.4.5 Consumer Data as Indicators

The desire not to reveal customer profiles in order to retain a competitive advantage is however at odds with the aspiration to access data for public good and make valuable contributions to the understanding of society. Many techniques already exist in the commercial world to gain insight on people's lived experience, but the overarching focus of these analyses is to produce indicators which improve profitability rather than to provide an improved, generalisable insight into society. Consumer data also offers a means of generating bespoke indicators based on daytime consumption patterns, thus creating representations of society that are not solely based on residential characteristics (Longley *et al.*, 2018). As discussed in Section 2.3.4 energy data offers commercial benefits for personalised tariffs but may also provide opportunity in regard to bespoke indicators.

It is the volume of information which makes smart meter data an exciting prospect in data driven research. The usage data contains information that is both spatially and temporally referenced (through customer addresses and their time of use), and unlike other forms of consumer generated big data such as loyalty cards, they do not rely on the customer performing a transaction once a meter is installed. This frequency may prove particularly useful for developing individual trajectories, where households can be seen waking, leaving, returning to and interacting with their homes. The longitudinal nature of this data makes it particularly valuable in providing insight into general spatio-temporal trends of individuals over various granularities, from by the hour, diurnally, weekly and seasonally. This also allows for the quantification of change between two static periods in time, capturing both short and long term dynamics. Furthermore, the georeferenced element of these data provide utility for inferring relationships with existing definitions of neighbourhood types and characteristics by linking to existing national statistics (Webber *et al.*, 2015). This innovative research may enable the creation of bespoke geodemographic classifications, linking consumption characteristics with existing geodemographic classifications.

2.4.6 Geodemographics

Area classification is the classifying of areas into groups on the basis of the similarity of characteristics of selected features within them (Vickers and Rees, 2007). One of the most commonly used is the geodemographic classification, which Sleight (2007, p.16) defines as;

“The analysis of people by where they live”

They provide a unique way of bringing together spatial patterns from a range of variables and identify similarities and dissimilarities between areas and can be said to work because of a fundamental notion in social structures, homophily, or ‘birds of a feather flock together’ i.e. if similar people live in similar places then knowing information about one person enables information about others in that locality to be broadly inferred (Sleight, 2007; Weiss *et al.*, 2012). This is consistent with Tobler’s first law of geography, that is, that “everything is related to everything else, but near things are more related than distant things” (Tobler, 1970, p.236)

There is a long legacy of producing geodemographic classifications in both the UK and the US, beginning with Charles Booths poverty maps in the early 1900s and Burgess and Park’s concentric zone model slightly later in the 1920s (Park and Burgess, 1925). This work was developed further by Carpenter, Shevky and Bell (1955) to include ‘social area analysis’ and was later broadened to encapsulate a series of tools and techniques which became known as ‘factorial ecologies’ (Brunsdon and Singleton, 2015). Geodemographics emerged from this context in the 1970s, and the analysis of them was developed as a strategic way of identifying patterns from multidimensional census data, with demonstrable utility in both public and private sector applications (Webber, 1978).

Advances in data availability and data processing techniques mean geodemographic classifications have gained wider popularity and contemporary classifications are typically of a high geographical granularity; at small area or address level, spanning cities, regions and countries. In the UK context cluster units are usually calculated at the Output Area or postcode level; the smallest census and postcode geographies available respectively. The methodological processes employed in the generation of a geodemographic classification are detailed in Section 3.6, but the resulting outputs are represented through the study of their relative attributes, to be used in a variety of fields to infer behavioural, health or other specific characteristics of a population group (Alexiou, 2016). They offer huge advantages towards the analysis and recognition of geographical patterns and can help identify triggering factors or important associations in nearly the whole spectrum of social phenomena

(Alexiou, 2016). There exists a cornucopia of literature on their application in retail planning and market analysis, unsurprising given their popularity in the private sector to gain competitive advantage. Some of the most well-known general purpose classifications are those which have been privately developed; the ACORN (by CACI), MOSAIC (Experian) and Claritas (PRiZM). There has been a recent upsurge in geodemographic applications for public sector usage, in particular, policy analysis and regional planning due to the advent of new application areas, which bring with them a set of benefits that set them apart from the commercially developed classifications. Open classifications can be accessed and scrutinised by the public without cost and have transparent and published methodologies. They are also comprised of freely available input data, making them reproducible and easy to operationalise, update and repurpose (Brunsdon and Singleton, 2015).

Despite this, there are methodological shortcomings and procedural limitations that must be addressed; the difficulty in producing a geodemographic framework is that they are in fact, aspatial, and fail to integrate the nearness described by Tobler in a sophisticated way due to the way that the final classifications are assumed to have the same underlying characteristics within clusters (Brunsdon and Singleton, 2015). This aspatiality also lends itself to a disregard of the issues around scalability, which affects the ability to make comparisons between classifications built for varying extents (Openshaw *et al.*, 1980; Webber, 1980). The subscription of large numbers of individuals to generalised profiles leads to the assumption that the social profile assigned to an area is representative of all households, engendering the well-recognised ecological fallacy (the confounding of the characteristics of areas with the populace within them), as in reality, socially homogeneous areas are rare (Dalton and Thatcher, 2015). A final procedural limitation is that as the unit inputs into geodemographic classifications are not naturally occurring (i.e. postcodes), the geographical scale and boundaries between areas can affect analytical results, also known as the modifiable areal unit problem or MAUP, which can result in two fundamental issues; scale effects and zonation effects, which mean that caution should be exercised when conducting spatial analysis on aggregated data (Openshaw, 1984, Openshaw and Taylor, 1979). To expand on this, scale effects lead to statistical results appearing more pronounced the larger the scale and zonation effects may lead to results which could present differently had boundaries divided the areas up otherwise at the same scale (Flowerdew, 2011). Furthermore, an inherent commercial confidentiality within the private sector means that those methodologies remain a "black box", impairing both the critiquing of those methodologies as well as their reproduction (Longley, 2007; Singleton and Longley, 2009).

2.4.7 Summary

The ability to produce, capture and store information has been transformative for the current data landscape and big data is affording opportunities for research never previously possible. It is important that the considerations of data bias, self-selected populations and the longevity of the data are addressed as a caveat of any research carried out, in order to alleviate concerns related to the uncertainty associated with their fitness for purpose. Triangulation is an important step in this area, providing proper contextualisation within geographic theory and confirmation (or not) of the novel datasets suitability. As far as the specifics of smart meter data, volume is another considerable concern and only advancements in machine learning algorithms will allow for the data to provide useful insights in near real time, but the predictive power of historic data should not be undermined. Incorporating consumer data into a geodemographic classification may offer a more comprehensive view of a population, by linking it to existing statistics and inferring relationships between it and the predefined characteristics and definitions of neighbourhood types.

3 Data and Methodologies

3.1 Introduction

This chapter provides a high level overview of the various data sources utilised throughout this thesis, and discusses their spatiality, attributes and characteristics to provide a solid grounding before their implementation in analysis and evaluation in subsequent chapters. It also covers the high level methodologies implemented in later chapters, such as postcode reweighting and the methodological framework of geodemographic classifications. The volume, variety and veracity of both the Domestic Energy Provider (DEP) smart meter and Energy Performance Certificate (EPC) datasets are discussed, examining their utility and application within the research project. The smart meter dataset was provided by one of the leading DEPs, who provide energy and energy services to homes across the UK. Given the spatial extent, granularity and volume, these data are unparalleled in comparison to those used in previous research on the dynamics of smart meter usage, especially in a UK context. A further contribution of this thesis is therefore to illustrate the value of such commercially generated big data for research within an energy policy context as well as the utility of effective data linkage from this and other ancillary sources.

3.1.1 The Consumer Data Research Centre

Access to these data were made possible through the ESRC funded Consumer Data Research Centre (CDRC): a government funded big data initiative that aims to facilitate the access of commercially generated consumer datasets to academic researchers. In order to secure the data, a number of strict procedures were necessary to minimise the risk of disclosing commercially or personally sensitive information about the provider or its customers, for both individual privacy and reasons of competitive advantage. These data are personal in nature, describing residential locations and consumption patterns at an individual customer level. These are classified as 'controlled data' under the CDRC regulations - meaning data that needs to be held under the most secure conditions with stringent access restrictions. This thesis represents one of few investigations into a consumer dataset in this context and issues of access, data handling and presentation of results were important associated challenges. An overview of the process required to conduct analyses and obtain results is provided in Figure 3-1.

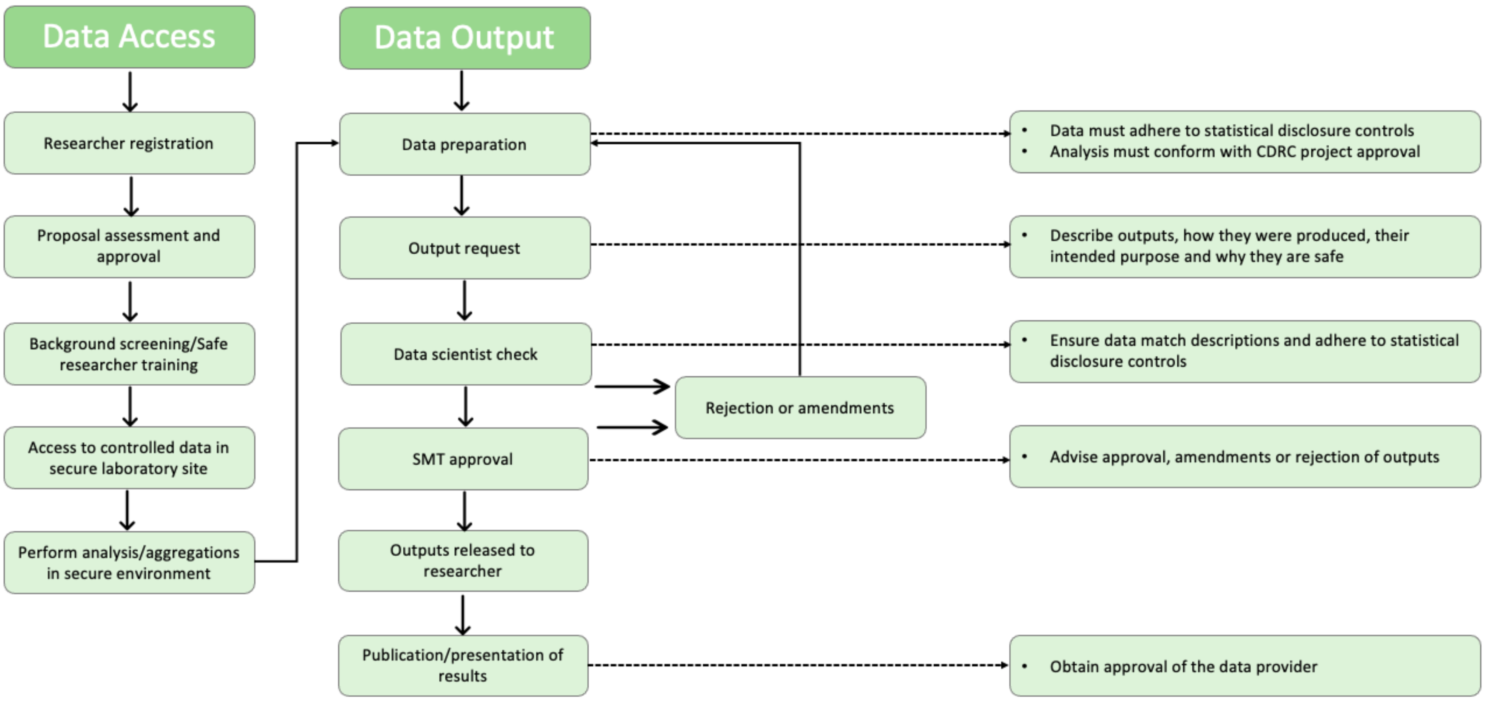


FIGURE 3-1 CDRC 'CONTROLLED DATA' PROCEDURES

Access was granted to these data via the CDRC's secure service held at the University of Liverpool's Computer Services Department secure facility. In the first instance, this requires preliminary vetting and training procedures that ensure access is only granted to trusted researchers (see the CDRC user guide 2018 in Appendix 9.2). Following this, researchers must complete a secure researcher training course and receive approval for proposed uses of the data and all analyses must be performed within the secure laboratory setting. To output data from the laboratory the data must firstly conform to a number of statistical disclosure controls. This includes but is not limited to; aggregation to large geographical areas, suppression of disclosive cells, ensuring percentages do not allow deduction of disclosive units and where counts are concerned, a threshold rule of no less than 10. These controls follow government specified rules and regulations on the handling of disclosive data (see the Government Statisical Services, 2014).

The second stage of data output then involves the assignment of two CDRC data scientists to carry out checks that ensure they match the output request descriptions and adhere to statistical disclosure controls. Finally, two members of the CDRC Senior Management Team (SMT) review and advise the approval, amendment or rejection of these outputs. Once obtained, the presentation and publication of analyses must also be approved by the data provider for commercial disclosure purposes.

As a result of these procedures, the presentation of analysis from these data have been necessarily constrained in order to adhere to both statistical and commercial disclosure controls. To achieve this adherence, the required data treatment is discussed in greater detail later in this chapter, pertaining to steps such as the spatial aggregation of the data.

3.2 The Energy Sector and Smart Meter Data

The UK energy sector is regulated by the Department for Business, Energy and Industrial Strategy (BEIS) and the Office of Gas and Electricity Markets (Ofgem). On the supply side, there are currently 62 active suppliers in the domestic gas and electricity retail market, consisting of 6 large suppliers and 56 small suppliers, most of whom are active in the supply of both energy sources (Correct as of Dec 2018, (Ofgem, 2018b)). The market share is monitored by Ofgem and assessed on the number of electricity meters on the distributional network attributed to each supplier. As of December 2018, British Gas was the largest provider with a 19% share, with Scottish and Southern Electric (SSE) and E.on the second and third largest providers with 13% each. At the time of the data collection, these providers held the same positions, but all have seen their market share drop as consumers move to smaller suppliers (Ofgem, 2019b).

3.2.1 What are Meter Readings?

Smart meters are the next generation of energy metering technology. Each meter is 'self-reading' and therefore records highly granular temporal energy consumption profiles for each installed address. These readings are provided to the supplier in watts (w) every half hour. Because watts are such a small unit, they are transformed into kilowatts (kW), equivalent to 1000w, and then into kilowatt hours (kWh). A watt indicates the power of an electrical appliance. A kilowatt hour is the amount of energy that an appliance consumes when it operates for one hour. Equation 1 shows the calculation applied to the DEP data to provide consumption levels in kWh.

$$kWh = \frac{watt \times 0.5}{1000}$$

EQUATION 1 CALCULATING KWH WHERE 0.5 EQUATES TO EACH HALF HOUR READING

As discussed in the literature review, smart meters connect wirelessly to send readings back to the energy providers - if this is lost because of faults or poor reception there can be missing data. Furthermore, if a meter reading is zero, it is difficult to know whether that is a genuine zero or a fault. It is possible, especially for gas that there may be a window between readings where none is used. It is less likely for electricity because of the 'standby rate' - where appliances and devices consume a nominal amount of energy when placed in standby mode - but if a house was left unoccupied, with everything switched off, a true zero could be possible (Wyatt, 2013). Given these data quality issues, Chapter 4 considers in greater detail the cleaning process for these zero values as part of the data preparation and minimisation methods to make it suitable for use in analysis.

3.2.2 Customer Trends

For each energy provider, a customer is defined by a unique account number attributed to the smart meters in their homes, but for national statistics on smart meter installations, the figures represent the number which are connected to the supply systems and the smart meter communications network. As such, the number of customers do not remain constant, due to the rollout programme increasing its membership from one month to the next and overall provider attrition. There is however a general trend of increasing participation. Throughout the national rollout scheme encompassing all suppliers, electricity meter installs remain consistently higher than gas. This variation could be explained by the fact that around 12% of households nationally are connected only to the electricity network. The most recent installation figures are summarized below in Figure 3-2, but do not include small providers with 250,000 customers or less.

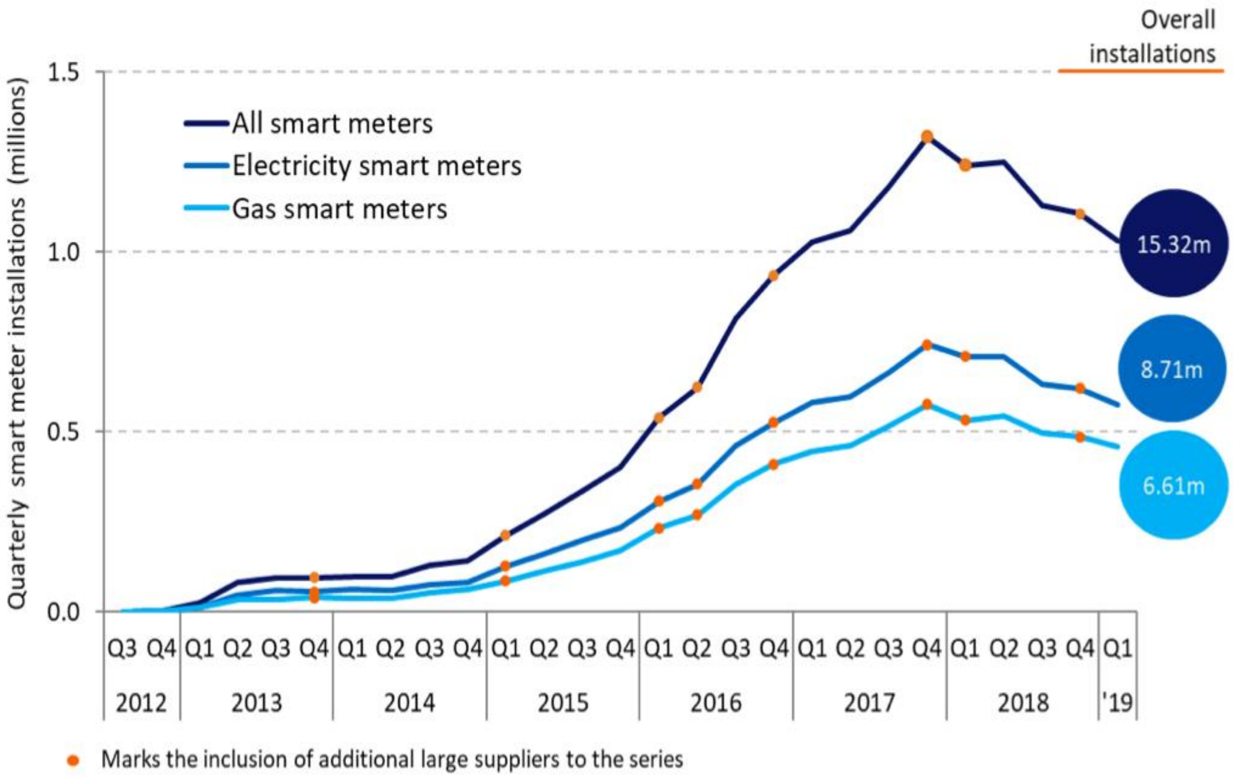


FIGURE 3-2 SMART METER INSTALLS FOR ALL MAJOR DOMESTIC ENERGY PROVIDERS (BEIS, 2019)

3.2.3 The DEP Data Overview

The DEP smart meter dataset follows a trend of new customers that is consistent with the national trends; for the timeframe for which data is provided (spanning 12 months for the 2015-16 financial year) electricity meters account for around 60% of the records and 40% come from gas. This variation could be explained by the fact that around 17% of UK households have multiple suppliers, meaning that one of their meters may not be present in the dataset despite having both gas and electricity supply.

Furthermore, in line with the national trend, the number of unique accounts does not remain constant as new smart meters are installed throughout the data collection period. In the case of electricity, 75% of users were already present in the dataset and so will have a full year's coverage. A breakdown of the rollout is shown in Figure 3-3 and from this we can surmise that during this period the rollout was gaining momentum (BEIS, 2017).

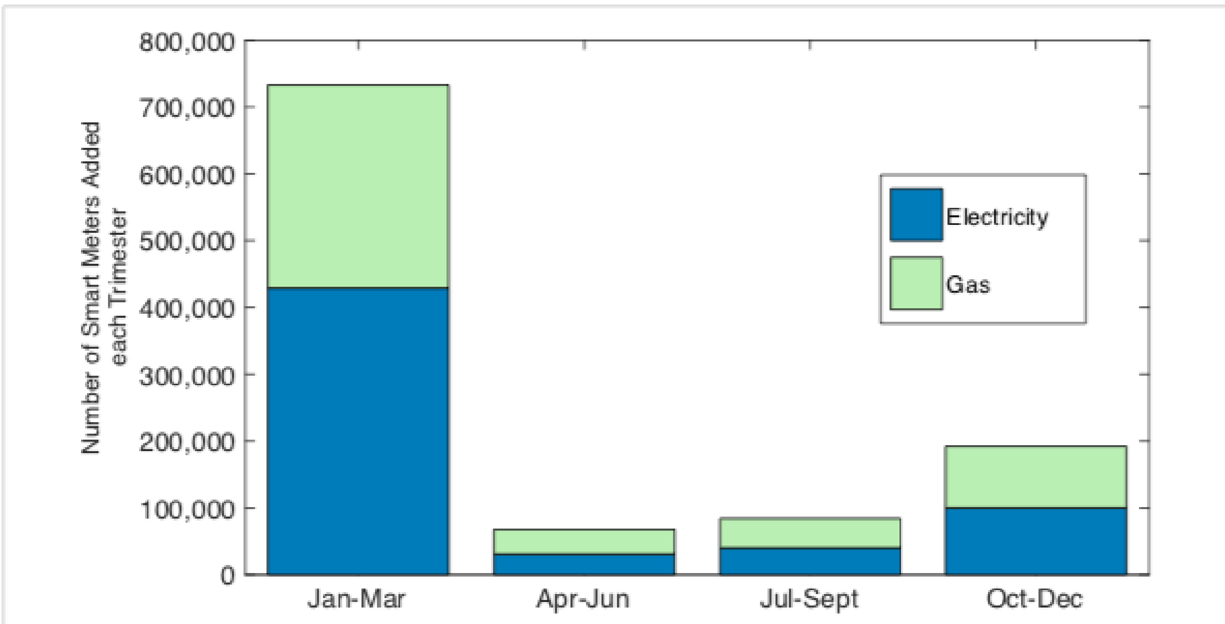
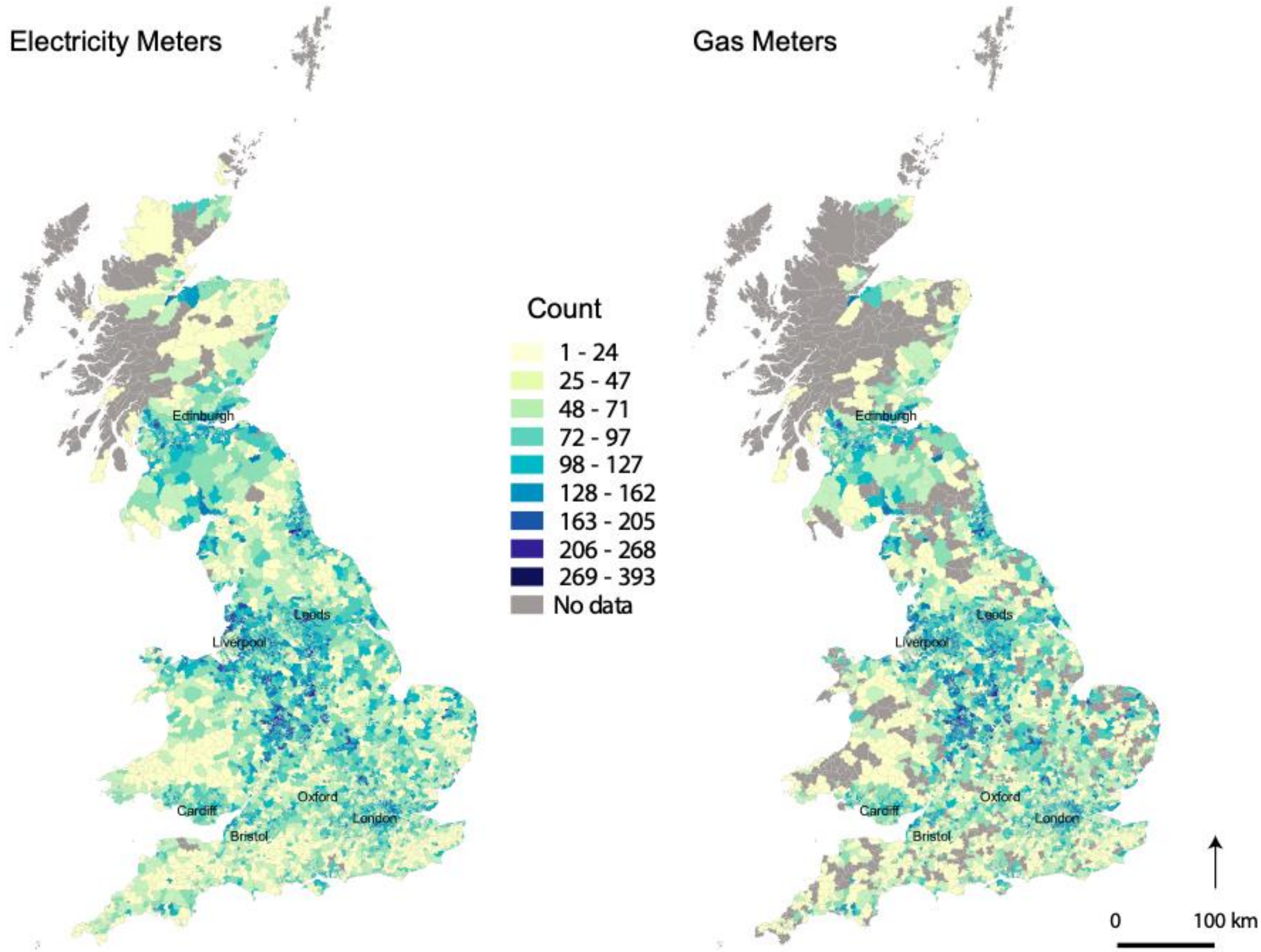


FIGURE 3-3 CHANGING SMART METER USERS OVER THE LIFESPAN OF THE DATASET (USHAKOVA ET AL., 2018)

IN ADDITION TO THE ROLLOUT GAINING MOMENTUM, IT WAS STILL IN ITS INFANCY AT THIS STAGE AND SEVERAL YEARS AWAY FROM COMPLETION, WITH NO FIXED ROLLOUT SCHEME IN PLAN BY THE DEP. AS THEY ARE INSTALLED AT THE ADDRESS LEVEL, INSTALLATION FREQUENCY CAN BE USED AS A PROXY FOR COVERAGE;

Figure 3-4 displays the coverage represented by the dataset as a whole, regardless of at what stage of the collection the household entered the dataset. It was found that generally, the percentage of households is no more than 3% for both gas and electricity. The highest percentages can be found in the West Midlands, North West and the North of Wales, which finds itself unusually over-represented.



From this it can be observed that the speed of installations may be greater in urban regions than rural, but it is also pertinent to note that this DEP, like most major suppliers have a legacy of regional bias in their customer base (Ushakova *et al.*, 2018).

FIGURE 3-4 PROPORTION OF GAS AND ELECTRICITY METERS RELATIVE TO THE TOTAL NUMBER OF HOUSEHOLDS BY POSTCODE SECTOR (USHAKOVA ET AL., 2018)

In the full dataset pre aggregation there were four descriptive variables and 48 consumption variables. It gives around 1,080,000 gas and electricity domestic smart meters, representing 43% of all installations by the end of the given year. Table 3-1 details the broader national figures from the BEIS (2017) (rounded to the nearest 00).

TABLE 3-1 TOTAL NUMBER OF METERS IN THE UK

Type	Number of Meters	Number of Postcode Sectors with at least 10 meters installed	Mean number of Meters per Postcode sector
Electricity	600,000	8000	70
Gas	480,000	7,500	60

The location identifier gave account holder information at Postcode Sector level (PCS) and was used to aggregate the records which resolves issues of reidentification, where individual consumers could be personally identifiable in the raw data through their address or consumption profile. At this level the data adhered to the CDRC disclosure controls outlined in the introduction and could therefore be extracted from the secure data environment for further analysis. Any PCS with a count of fewer than 10 households was removed from the dataset for their privacy.

3.3 Energy Performance Certificates

3.3.1 What is an Energy Performance Certificate?

Energy performance certificates (EPCs) were introduced in stages from 2007 and stem from the EU directive on the energy performance of buildings. They were first introduced as part of the now redundant Home Information Pack or "HIP" scheme required to sell a home. When these were phased out, the EPC element was retained as part of the Government's energy efficiency improvement strategy; it was intended that the energy efficiency of buildings was made transparent through these certificates, making comparing properties easier for buyers and making energy efficient homes a more attractive proposition. Research exists to suggest that an A rated home could sell for 14% more than an equivalent G rated property (Fuerst *et al.*, 2016). It was also intended that this would stimulate the market into making energy efficiency improvements before selling. The Standard Assessment

Procedure (SAP) algorithm takes information such as wall type and levels of glazing and then makes assumptions about a building's thermal properties. The procedure differs between new build and pre-existing homes; the former receive a full SAP calculation, whereas the latter are assessed using a reduced Standard Assessment Procedure (Hardy and Glew, 2019). The algorithm then generates an energy efficiency rating out of 100, and a linear rating from A (the most efficient) to G (the most inefficient), which is recorded in the EPC (Hardy and Glew, 2019). They also predict how costly it will be to heat and light, by calculating the expected total kWh per year for the building based on its characteristics, as well as its likely carbon dioxide emissions and stating what the energy efficiency rating could be if improvements are made whilst highlighting cost-effective ways to achieve a better rating (Energy Saving Trust, 2020). The full report given to a householder provides a high level of information, summarising the top actions that could be taken, the indicative cost and the typical saving, (Energy Saving Trust, 2020). Figure 3-5 Typical Output of an Energy Performance Certificate shows the EPC summary detailing the energy efficiency rating many will be familiar with.

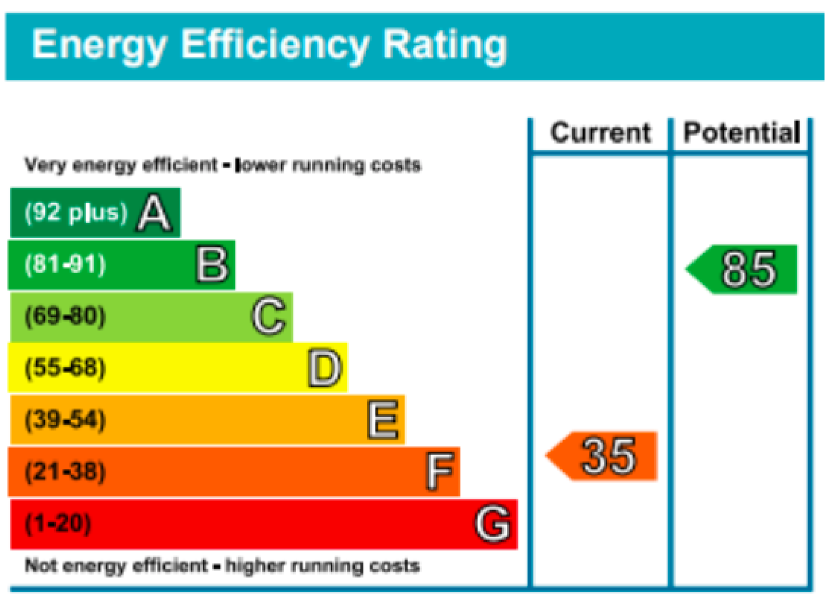


FIGURE 3-5 TYPICAL OUTPUT OF AN ENERGY PERFORMANCE CERTIFICATE (Energy Saving Trust, 2020)

Domestic certificates are valid for ten years but can be renewed sooner if, for example, the house has been put on the market for sale or rental and improvements have been made since the last certificate was issued thus making it a more attractive property. Other transactions which also require the generation of an EPC include certain types of government energy improvement funding schemes to prove eligibility such as an assessment for Green Deal energy efficiency improvements; following a Green Deal; Feed in Tariff application; Renewable Heat Incentive application or "ECO" (Energy Company Obligation) assessment.

Commercial properties also require an energy performance certificate, which is referred to as a Display Energy Certificate (often referred to as a DEC) in England and Wales and differ slightly from an EPC in that the energy assessor producing the certificate will also take meter readings and look at the actual energy consumption of the building when it is in use. The length of their validity is dependent on the size of the building; buildings over 1,000 square meters which are occupied by a public authority and frequently visited by members of the public are valid for one year. DECs for buildings of between 250 and 1,000 square meters are valid for 10 years (MHCLG, 2018). There is currently no regulation in place which states that commercial buildings under 250 square meters have to hold either EPCs or DECs. Places of Worship, warehouses and temporary structures are also exempt.

Energy Performance Certificate data are available in an open access dataset published on-line by the UK government, and details all energy certificates within their records. It can be freely downloaded in full by providing an email address or searched via an API. It was first published online in 2017 and is updated every six months. The full dataset used in this analysis was downloaded in November 2018, containing 176 variables and around 18 million records in total. The full definitions and metadata pertaining to each variable can be viewed in the Ministry of Housing, Communities and Local Government Guidance Notes[3] which accompany the data, an extract of which is included in Appendix 9.3, covering the subset of variables included in the final Energy User Classification in Chapter 5.

[3] https://epc.opendatacommunities.org/docs/guidance

3.3.2 Data Quality and Bias

EPCs assessments are done on site in a person's home – an accredited assessor attends the property to collect the data required for assessing energy features and generates an EPC once off-site using Government approved software. Whilst this software means that there is some level of standardisation across inputs, they are still error prone and rely on the conscientiousness and skill level of the assessor. Hardy and Glew (2016) estimate it is possible that the assessor is responsible for up to 62% of errors found within the EPC database. As it is only a visual assessment and non-invasive, if there is no evidence (physical or documentary) of fittings such as insulation for the assessor to observe then the level of fixtures and fittings are assumed based on the age of the relevant part of the dwelling. They also use assumptions in order to make properties directly comparable, such as standard occupancy when commenting on energy usage, both of which introduce uncertainty in the final outcome.

From the EPC data cleaning process (see Section 5.3) it was clear that there was a high margin for error when inputting the data into the software. Whilst some inputs were clearly standardised, or check boxes, the free text elements were difficult to decipher and led to a large amount of data having to be discounted as it was not possible to understand. For example, some of the EPCs in the register were completed in Welsh with spelling mistakes in the free text making them complex to translate. Their utility has been subjected to criticism from industry experts such as the Royal Institute of Chartered Surveyors due to their perceived poor quality and lack of regulation over the qualifications required to become an accredited assessor. The English Housing Survey (MHCLG, 2018), undertaken by the Ministry of Housing, Communities and Local Government (MHCLG) reported that of the 10.6m homes that had moved since 2008 (the introduction of EPCs) 76% of respondents stated that they were not influenced by the results of the EPC on their move and owners of non-standard housing such as listed buildings realise little value in the recommendations and find it very difficult to rectify their low ratings because the recommended improvements, such as double glazing, are often barred by the control on changes to such structures (MHCLG, 2019).

Chapter 5 reiterates this information and looks further into the literature regarding the usefulness of EPCs to owners and tenants as well as providing more detail into the full data cleaning process.

3.4 Supporting Data

Throughout the thesis a number of supporting datasets were utilised to contextualise the energy data and provide ancillary information, pertaining to both raw data and pre-compiled measures such as geodemographic classifications.

3.4.1 Census Data

Every decade since 1801, England has had a designated 'census day', whereby information is obtained on every member of the population. It is the most complete source of information about the population that contains details of family composition, health, employment and other socio-economic characteristics. This allows for targeted resourcing and policy planning, as well as academic research.

It is common across disciplines such as human geography and demography to use census data to study social trends. Census data are collected at an individual household level but are released for open use in aggregate form to avoid disclosure of personal information. This aggregation usually takes place on the basis of geographical location and in England and Wales, these small areas are referred to as 'Output Areas' (OAs). They nest neatly inside the larger geographies of Local Super Output Areas, Middle Super Output Areas, and Super Output Areas (LSOAs, MSOAs and SOAs respectively). This means that the same sets of data can be analysed at different scales, allowing for meaningful comparisons between census and other variables (Flowerdew, 2011).

The data is commonly provided in a series of pre-evaluated and highly requested tables, as well as commissioned tables for individual requests and for the first time in 2011 data was made available for download and is now available through multiple data services, such as the UK Data Service and Nomis web, where flexible user generated tables can be obtained to meet specific needs. The UK Data Service[4] was employed to gather the requisite data for this thesis which were integral to the clustering methodology presented in Chapter 5. Despite the census data being almost a decade old at the time

[4] https://www.ukdataservice.ac.uk

of writing, the fact that the data is the most complete set of statistics available for the country as a whole still holds a lot of value, especially in regard to the triangulation of the characteristics examined in the smart meter (from 2015) and EPC datasets, which contain data from between 2008 and November 2018. The existing literature aided the variable selection; taking data on the factors most commonly associated with the status of fuel poverty and those which are reported as having an impact on the ability to consume energy such as accommodation type, building type, tenure, as well as family life stage indicators; marital status, age of household members, and their economic activity and employment statuses. Counts for these variables were downloaded at OA level; the smallest geography is the easiest to scale up and aggregate and in order for the thesis as a whole to be consistent and provide meaningful comparisons between datasets and across chapters. The census data was reweighted to Postcode Sector Level throughout; a detailed methodology describing how this was achieved appears in Section 3.6.

TABLE 3-2 CENSUS TABLES CHOSEN FOR INCLUSION

Census Table Name	**Key Statistics**	**Scale**	**Coverage**
KS103EW	Marital and Civil Partnership status	OA	UK
KS401EW	Dwelling, Household Space and Accommodation Types	OA	UK
KS402EW	Tenure	OA	UK
KS403EW	Rooms, Bedrooms and Central Heating	OA	UK
KS601EW	Economic Activity	OA	UK
KS611EW	NS-SeC	OA	UK

3.4.2 The Internet User Classification

The 2018 Internet User Classification is a bespoke classification that describes how people living across Great Britain interact with the Internet (Alexiou and Singleton, 2018). Engagement with the Internet has an obvious impact in consumer behaviour; with regards to energy consumption it is the opportunity and willingness to engage with new Internet enabled devices within the home, including the smart meter technology and IHD, as well as associated household management tasks such as receiving, viewing and paying bills online, engaging with comparison and switching websites to get the cheapest tariffs, and accessing educational material with regard to smart meters and improved consumption practises. Using this dataset in the final chapter of this thesis gives considerable insight into the possible reasons for low smart meter engagement in our case study area, and the finding could be applied more widely.

The updated IUC classification was published in 2018 by the CDRC and uses consumer, survey and open data to produce the classification. It is openly available to download from the CDRC[5] and covers all of Great Britain.

TABLE 3-3 INTERNET USER CLASSIFICATION VARIABLES

Variables	Scale	Coverage
IUC Group	LSOA/Data Zone	Great Britain
IUC Group Label	LSOA/Data Zone	Great Britain

[5] https://data.cdrc.ac.uk/dataset/internet-user-classification

3.4.3 Small Area Incomes

Income data from the UK Government provided annual statistics on before and after housing costs at MSOA level across Great Britain in 2012, 2014 and 2016. This allowed for the visualisation of the population's percentage change in disposable income, relative to their income quintile over time. This data is derived from the Family Resource Survey (FRS) which takes into account all sources of income (such as self-employment or benefits)(Office for National Statistics, 2018).

Variables	Scale	Coverage	Sample Size
Net Annual Income Before Housing Cost	MSOA	England and Wales	19,200 households (FRS)
Net Annual Income After Housing Cost	MSOA	England and Wales	19,200 households (FRS)

3.4.4 Building Age Data

Building age data was used to aid the understanding of fixtures and fittings of homes in the EPC dataset and also to illustrate the relative age of the building stock in the UK; this goes some way to explaining the lack of energy efficient households in the UK, especially in northern areas, and reiterates the need for efficiency based policy interventions to consider existing housing stock as well as new build homes (Section 2.1.4). This data is made available for download through the CDRC, who have manipulated, cleaned and visualised the data but originates from the Valuation Office Agency in the 'dwelling age' dataset.

TABLE 3-4 BUILDING AGE DATA VARIABLES

Variables	Scale	Coverage
Dwelling Age band	LSOA	England and Wales

Dwelling count	LSOA	England and Wales

3.4.5 Indices of Multiple Deprivation

The Indices of Multiple Deprivation (IMD) is the official measure of relative deprivation at the small area level in England and follows an established methodology in broadly defining deprivation to encompass a wide range of factors affecting an individual's living condition and is used to facilitate the targeting of policies and resources within disadvantaged communities. It is based on 39 combined and weighted indicators across the following 7 domains; income; employment; health deprivation and disability; education, skills and training; crime; barriers to housing and services and living environment. It is calculated at LSOA level and covers the extent of England. In this thesis the 2015 iteration of the classification was utilised, and the Welsh Indices of Multiple Deprivation from the same year was appended to give improved coverage; it is constructed in the same way as the English IMD using the same domains and is produced by the Welsh Government.

TABLE 3-5 ENGLISH AND WELSH INDICES OF MULTIPLE DEPRIVATION VARIABLES

Dataset	Variables	Scale	Coverage
English Indices of Multiple Deprivation	IMD Rank	LSOA	England
English Indices of Multiple Deprivation	IMD Decile	LSOA	England
Welsh Indices of Multiple Deprivation	IMD Rank	LSOA	Wales
Welsh Indices of Multiple Deprivation	IMD Decile	LSOA	Wales

3.4.6 Output Area Classification

The Output Area Classification (OAC) summarises the social and physical structure of neighbourhoods using data from the 2011 UK Census, with its overarching aim being "to describe the salient and multidimensional characteristics of small areas across the UK" (Gale *et al.*, 2016, p. 3). It re-evaluates the 2001 Output Area Classification and places a greater focus on key elements of data selection and testing new methods, addressing some issues of its predecessor. It also adopted only open source software and released all code and metadata once the classification was completed, addressing those commonly cited issues of 'black box' classifications, transparency and reproducibility, thus increasing its accessibility and trustworthiness by allowing scrutiny. It describes geodemographic population characteristics across 8 Supergroups, 26 Groups and 76 Subgroups and is available at the OA level (Gale *et al.*, 2016). It is built using census variables from a number of domains that were said to best represent drivers of socio-spatial differentiation in the UK; demographic structure, household composition, housing, socio-economic and employment (Vickers and Rees, 2007). Table 3-6 below details an extract of the naming conventions used in the OAC cluster hierarchy.

TABLE 3-6 EXAMPLE OF OAC CLUSTER NAMES AND HIERARCHY

Supergroups	**Groups**	**Subgroups**
2 - Cosmopolitans	2a - Students Around Campus	2a1 - Student Communal Living 2a2 - Student Digs 2a3 - Students and Professionals
	2b - Inner-City Students	2b1 - Students and Commuters 2b2 - Multicultural Student Neighbourhoods
	2c - Comfortable Cosmopolitans	2c1 - Migrant Families 2c2 - Migrant Commuters 2c3 - Professional Service Cosmopolitans
	2d - Aspiring and Affluent	2d1 - Urban Cultural Mix 2d2 - Highly-Qualified Quaternary Workers 2d3 - EU White-Collar Workers

3.4.7 Urban Rural Classification

The Urban Rural Classification is a dataset produced by the Office of National Statistics (ONS) to distinguish urban and rural areas at the OA level, and at its most detailed pertains to four urban and six rural settlement/context combinations; Urban major conurbation; Urban minor conurbation, Urban city and town; Urban city and town in a sparse setting; Rural town and fringe; Rural village; Rural hamlet and isolated dwellings; Rural town and fringe in a sparse setting; Rural village in a sparse setting and Rural Hamlet and isolated dwellings in a sparse setting (Office for National Statistics, 2016a). Each OA is prescribed as urban or rural depending on its (population weighted) centre is within or outside a built up area of fewer or greater than 10,000 people. Its utility is based on the fact that socio-economic opportunities are likely to differ based on their make-up, and the barriers and challenges people face as well as the services people have access to will vary depending on their level of rurality (Office for National Statistics, 2016a). The utility of this data within this context aids understanding of access to services such as mains gas and standard fuel types, both of which are important factors in the affordability of fuel and the efficiency rating of the home.

TABLE 3-7 URBAN RURAL CLASSIFICATION VARIABLES

Variables	Scale	Coverage
Urban Rural Classification	OA	UK

3.4.8 Current Fuel Poverty Statistics

An estimate of fuel poverty is provided within data supplied by the ONS and is only available for the extent of England; it should however be noted that this dataset is caveated as being 'experimental statistics'. It is calculated using data from the English Housing Survey which collects information about people's housing circumstances, their condition and energy efficiency. Despite this, it is the best representation that exists and so was included in this thesis to provide spatial context to the current fuel poverty definition. By understanding the geographies of the current fuel poverty definition, it is possible to draw comparisons between it and the Energy User Classification, allows for identification of areas which present as 'not fuel poor' under the current definition but which display demographic attributes which indicate that this may not be true and vice versa, thus revealing the limitations of the current definition.

TABLE 3-8 FUEL POVERTY DATA VARIABLES

Variables	Scale	Coverage	Sample Size
Proportion of Households estimated to be in fuel poverty	LSOA	England	13,000 households in the English Housing Survey

3.5 Postcode Geographies

A challenge for this thesis was the spatial scale at which the DEP data were supplied. These were limited to postcode sectors (PCSs) as described earlier, however, as a geography is directly comparable to those units used to disseminate those other contextual data described in this chapter thus far. To avoid manipulation of the DEP data, a method was required to produce estimates from source data within the PCS zones. A dasymetric mapping technique was implemented to reweight attributes to PCS level. Although there are numerous methods such as built environment overlaps and area overlaps between zones that might be implemented to achieve this aim, a postcode matching method was implemented here.

A GIS 'area overlap between zones' methodology was trialled, using the Postcode and Census Geography shapefiles and a union algorithm, however, due to the granularity and large number of intersections it was extremely time consuming and when the outputs were inspected it was found to have led to slither polygons and inaccurate distribution of area, meaning results were less likely to be accurate than the reweighting methodology. It was also much more computationally expensive to repeat this process at all spatial granularities, and visually outputs were untidy due to the missing polygons. Because of these limitations, the GIS methodology was not progressed beyond preliminary trials.

At the most recent count (May 2019) there are 11,918 postcode sectors in the UK and Table 3-9 below gives a useful breakdown of postcode geographies and their comparative census geographies. A postcode sector averages around 3000 households but can be as low as 60 in deep rural areas and as high as 10,000 in densely populated areas.

TABLE 3-9 POSTCODE AND CENSUS GEOGRAPHIES (NOMIS AND ONS, 2017)

Geography	Number of Areas	Number of Households
Postcode Unit (CV37 6QW)	1,759,751	17
Postcode Sector (CV37 6)	11,199	3,040
Postcode District (CV37)	2,269	10,766
Postcode Area (CV)	127	232,663
Output Area	181,408	134
LSOA	34,753	702
MSOA	7,201	3,392
Unitary Authority	8,570	70,200

The Royal Mail postcode database on which the ONSPD is based is regularly updated to include new postcodes generated through new housing estates and business addresses and the removal of postcodes which have become redundant and so the numbers in the above table may become out of date over time.

The postcode matching methodology was possible with the use of the Office of National Statistics Postcode Database[6] (ONSPD) which provides a look-up between postcode boundaries and corresponding census boundaries for OA, LSOA, MSOA and SOA, as well as the Local Authority District (LAD) (Office for National Statistics, 2016b, 2019b, 2019a). Combining this with the Postcode Headcount and Household Estimates[7] table produced from 2011 census statistics, it is possible to re-calculate the proportion of postcode sectors within each of the output area levels. It is then possible to use this re-weighted population data to apportion data by using the output as a multiplying factor to reweight census data not ordinarily available for these postcode geographies.

Limits to this method include an assumption that the population spatial distribution is even across both the postcode and census geographies . It is also using 2011 census geographies; whereas Postcode names and boundaries are subject to continuous change (Office for National Statistics, 2016c). This may mean that the reweighted figures quickly become outdated and require recalculation on a regular basis. The methodology is reproducible to account for this, but it should be kept in mind; especially in urban areas where regeneration schemes and redevelopment mean that the housing, and therefore postcode landscape can change quickly.

For categorical data the ONS have a standardized practice used when Postcode and Census geographies do not align. Only one postcode per OA is assigned; and it is the one in which the majority of residents is contained, as per the census count. It is important to recognise that this method may lead to the loss of some detail in neighbouring postcodes but is a recognised and recommended methodology (Office for National Statistics, 2016b).

---

[6] https://geoportal.statistics.gov.uk/search?collection=Dataset&sort=name&tags=all(PRD_ONSPD%2CAUG_2020)

[7] https://www.nomisweb.co.uk/census/2011/postcode_headcounts_and_household_estimates

3.6 Geodemographic Classification

Geodemographic classification can be described as 'the analysis of people by where they live' (Sleight, 2007) and involve analysis of attributes relating to the socio-economic and built environment characteristics of small geographic areas. There are numerous approaches to their construction, balancing empirical analysis alongside classification builder experience (Alexiou, 2016). Such variation in the specification and creation of geodemographic classification should however be expected, given that particular configurations suitable for one classification may not be suitable for another. Harris, Sleight and Webber (2005) present a comprehensive overview of the various stages involved in constructing a geodemographic classification, and as such this section will utilise their framework to consider our approach, comprising of some of the following steps:

- Selecting potential measures
- Data evaluation
- Transformation and normalisation
- Weighting
- Standardisation
- Clustering
- Cluster hierarchy
- Textual and visual summaries

The first stage in the building of a bespoke geodemographic classification is to perform an evaluation of potential input measures. Most cluster analysis techniques implemented to build a geodemographic classification require measures to be continuous as opposed to discrete or categorical. Other clustering algorithms exist for these instances, such as Expectation-Maximisation or Hierarchical Clustering, but are less commonly used in this context (Harris *et al.*, 2005; Vickers and Rees, 2007; Singleton and Spielman, 2014). Data inputs are generally aggregated to a predefined geographic resolution, as dictated by the scales at which all data sources are available. Those resolutions in the UK on which a classification can be built vary, but common aggregated geographies include postcode (or aggregation of; postcode sector/postcode district) boundaries, OA, LSOA or ward, or in Scotland and Northern Ireland, Data Zones or Intermediate Zones.

Exploratory Data Analysis (EDA) is typically the next stage, evaluating input variables to examine issues such as missing values and correlation, assess distributions and skew, and more generally, gain an overview of relationships between variables. Available data can have skewed distributions, contain a high rate of missing values or originate from smaller sample sizes than is desirable, thus generating uncertainty (Alexiou, 2016). It is customary in geodemographic building to start with a larger pool of variables and progressively removing those that seem problematic or likely to skew results; for example, the aforementioned 2011 OAC classification considered 167 initial variables, of which only 60 were used to build the final classification (Gale *et al.*, 2016). At this stage a classification builder may choose which, if any, variables should be removed if they present duplicate information (typically those variables that show high correlations with others giving information which is already known) but there are no firm rules and choices can be largely subjective (Harris *et al.*, 2005). It is suggested that attributes with very high cross-correlations should be avoided as they effectively measure similar dimensions, although conversely may also capture variation across areas, which could be interpreted as pairs of variables having significant descriptive power (Voas and Williamson, 2001).

After gathering of input measures, it is then necessary to consider whether normalization should be applied, and if there is a need to transform the data onto a common scale. In an ideal scenario, all variables would exhibit normal distributions as some clustering algorithms (such as the commonly used K-means) are optimised to find spherical clusters, which can be problematic with skewed inputs. For many socio-economic data, there are very few situations where this holds true. There are a range of normalisation practises that can be implemented, including log10, Box-Cox and cube-root transformations (Gale *et al.*, 2016).

However, whether or not skewed data should be transformed is subject to debate (Singleton and Spielman, 2014). In some commercial classifications skew and other characteristics of measures deemed to be problematic are purported to be overcome by employing a weighting technique to reduce the impact of variables, but how the weights are derived is typically subjective and open to criticism (Harris *et al.*, 2005). In many open classifications there is often an explicit decision to not weight inputs given argued inherent subjectivity, although, it might be noted that such processes could be defensible if a full rationale and specification for these decisions were accompanying a classification.

The standardisation of data is applied in order to transform the data in order to equalise range and/or variance and has utility when input variables are from different sources and measurement scales. In order to assess how large or small a particular geographic area's variance is from the mean across a set of variables, and to draw comparisons between these measures, a common scale is required. Two of the most common functions applied are z-scores and range standardisation. Z-scores are commonly used, which are calculated by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation.

$$Z = \frac{x - \mu}{\sigma}$$

EQUATION 2 Z-SCORES - WHERE μ IS THE MEAN OF THE POPULATION AND σ IS THE STANDARD DEVIATION OF THE POPULATION.

This results in a set of scores that are positive if they fall above the mean and negative if they fall below, i.e. all standardized variables will have an adjusted population mean of 0. However, using z-scores can be problematic, for example if an input variable is highly skewed with many outliers, the resulting z-score can accentuate such effects and influence an area's cluster membership regardless of the area's other attributes. Range and inter-decile range transformations are also viable options; range standardisation compresses the values into the range of 0 – 1, but with different means and variances. It has been used successfully in geodemographic classifications such as the ONS 1991 classification and the 2001 Output Area Classification. Both range and inter-decile range standardisation reduce the impact of outliers by scaling the data into smaller intervals, resulting in a loss of information (Gale *et al.*, 2016). Again, weighting and variable normalisation techniques can be utilised to alleviate this issue.

Once a final set of cleaned and transformed variables are acquired, the next stage is to run a cluster analysis. Different combinations of algorithms are used, but typically involve the iterative allocation-reallocation method (K-means), and optionally, a hierarchical method such as Wards clustering (Vickers *et al.*, 2005).

The hierarchical method essentially treats each area as a separate cluster in the first instance and merges these 'clusters' based on measures of similarity. After similar clusters are merged, average values for the new clusters are computed and the process repeats until convergence, where an appropriate number of clusters (that exhibit minimum intra-cluster variance and maximum inter-cluster variance) are found. Although methodologically simplistic, this method can be expensive in terms of time and computational effort due to the assessment and reassignment of cluster pairs whilst holding the intermediary results in memory. This can be particularly problematic when datasets are extremely large

.

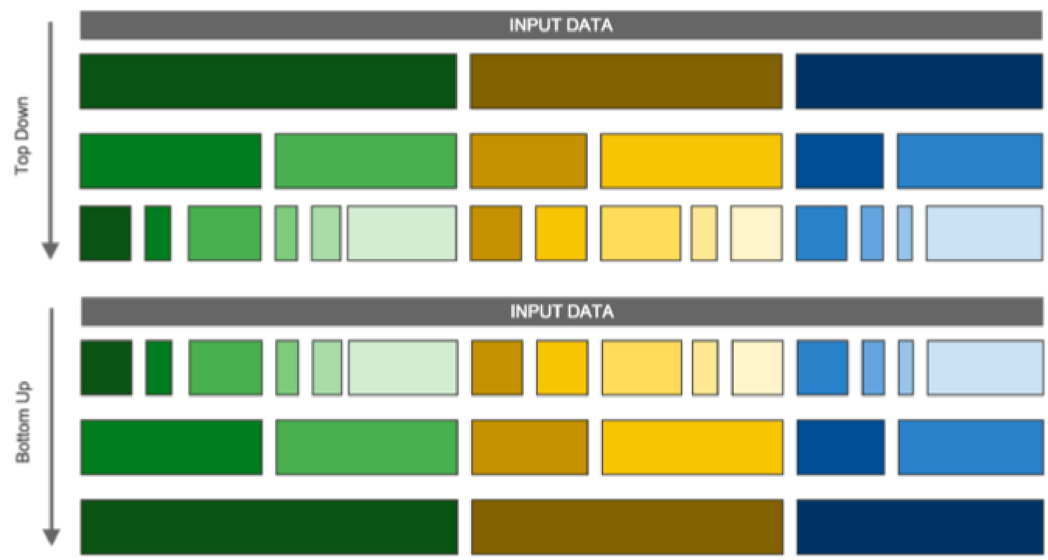


FIGURE 3-6 HIERARCHICAL CLUSTER DESIGN (Riddlesden, 2016).

Top-down hierarchies involve clustering input data into predefined numbers of clusters that will form the most aggregate tier of the resulting classification, which are then used to split the input data and clustering is applied within the subsets to successively generate new tiers of the classification as shown in Figure 3-6. Although this method can be repeated as many times as it is desired, it is sensible to stop when sub-clusters begin to display no obvious differences from the parental clusters.

A bottom-up hierarchy involves clustering the data into K clusters representing the finest level of a classification, which are then merged to form a higher tier within the hierarchy; typically with Wards clustering methods.

An iterative allocation-reallocation method uses a different technique to compute cluster assignments. A K-means algorithm works by setting seeds, which are a random allocation within the vector space. The number of initial seeds (k) is equal to the pre-determined optimal number of clusters to be output. Several methods exist to aid the user in defining this number, though the process is usually iterative and involves extensive testing / user consultation. One such method involves the use of 'Clustergrams'; visualisations of the assignment and re-assignment of observations to clusters across a range of values for k. In this regard, it is similar to dendrograms, but can also be implemented on non-hierarchical data. The clustergram is constructed as follows; "for each cluster within each cluster analysis, compute the mean over all cluster variables and over all observations in that cluster" (Schonlau, 2004, p.5). For each cluster, the cluster mean versus the number of clusters is plotted, and consecutive clusters are joined by parallelograms. This visual method can assist in the selection of an optimum k value as it is possible to identify which clusters split to form new clusters and assess similarity or 'closeness' of newly formed clusters, as well as assess the number of observations within each cluster based on the width of the parallelogram (Schonlau, 2004).

Clustergrams are a relatively innovative method, but much improved in terms of ease of comparison when compared to existing methods such as elbow plots and the gap statistics; all of which are discussed in greater detail in Chapter 5 during the building of EPC K-means cluster typology.

Once the initial number of seeds has been set, the algorithm then begins to assign observations to each of the seed locations based on proximity, typically measured by Euclidean distance. This initial allocation represents the first iteration of the algorithm. The centroids of the newly formed clusters are then calculated and become the centres for the next iteration of assignments. The algorithm aims to minimise the total within cluster sum of squares (WCSS), which is the cumulative sum of all the squared Euclidean distances from observations to cluster centroids. Smaller WCSS values represent more homogeneous (or similar) clusters. The algorithm repeats for many iterations until convergence, when assignments no longer change and WCSS values have been minimised. Figure 3-7 describes how the algorithm works on a 2 dimensional dataset.

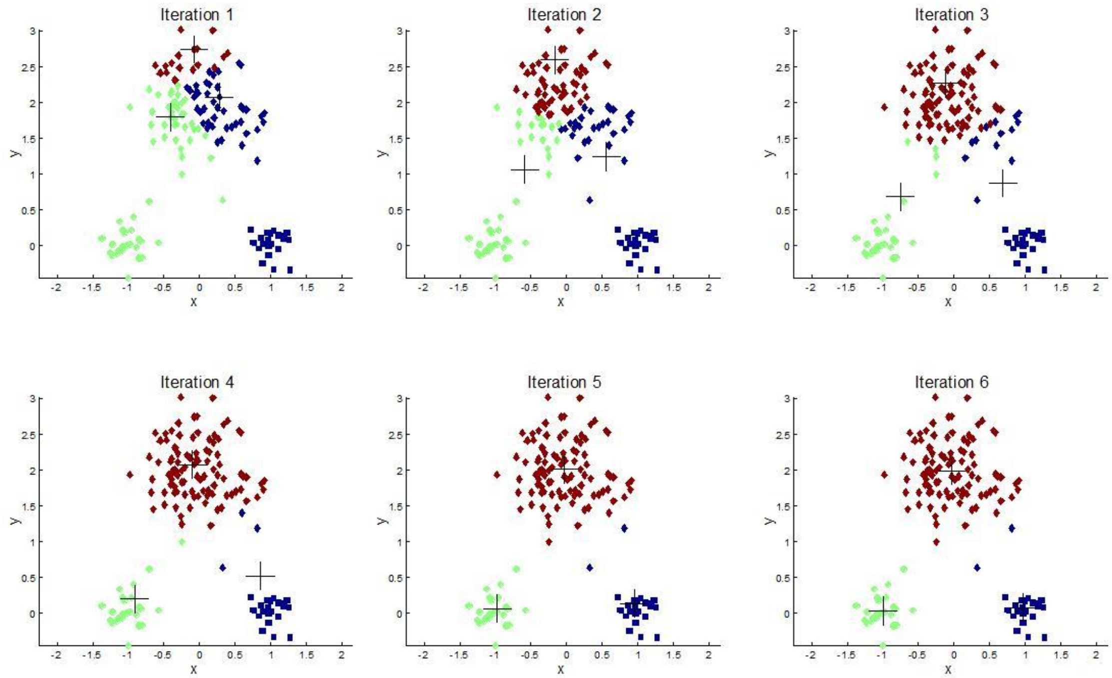


FIGURE 3-7 ITERATIVE ALLOCATION CLUSTERING (Practical Computing Applications, 2017)

3.7 Summary

The preliminary analysis of the smart meter dataset location attributes suggested that these data are biased towards certain areas of the country, and therefore segments of the population, primarily towards northern urban locations, with the West Midlands, North West and North of Wales over-represented. This variation in representativeness informs further investigation undertaken in the following chapter and the legacy of regional bias is fundamental to take into account if attempting to extrapolate the dynamics of smart meter users to that of the general populace.

Given the uniqueness and modernity of the smart meter consumption data and EPC data accessed for this study, it was considered essential to incorporate external data from ancillary sources that contextualise findings and provide the opportunity for developing richer profiles. These are introduced in section 3.4. Census data was considered the most relevant source of triangulation, given that it is the most complete source of population data available to researchers. The IUC, IMD, Building Age data, OAC, Urban Rural Classification and current fuel poverty statistics were all employed to provide supporting or contextual information to the insights which were generated from the smart meter and EPC data with a view to realising the value in the linkage of multiple data sources in order that a much broader understanding of the populace than one that is possible by traditional data means be generated. Some data quality issues are also acknowledged in the EPC data, which are investigated in situ in Chapter 5, relating to the quality of the data pre-processing and to the non-standardised elements which have introduced a level of uncertainty to the data.

Furthermore, it is important to acknowledge the practical implications and limitations of some of the methodological framework detailed in this chapter. A population based dasymetric reweighting methodology was chosen above GIS methods for the robustness of the resulting output and computational expense of implementation, but could be criticised for the fact that it makes assumptions about the spatial distribution of the population across both the postcode and census geographies.

In regards to the geodemographic framework described in this chapter, it is fair to say that elements of it are inherently subjective, such as the cost benefit of normalising skewed data, and by what method, but we can assert that variation in methods is acceptable and indeed common in the construction of a geodemographic classification based on the methods suitability to the data. It is important to balance the need to clean and transform data to optimise the resulting outputs with the fact that these methods will reduce the impact of any outliers and therefore potentially mask interesting differences and produce more homogenous clusters within the classification. In addition to this, whilst there are many, K-means is the most commonly used technique in geodemographics, and the approach adopted in the empirical chapters of this thesis therefore follows in the path of existing literature surrounding conventional geodemographics (Harris *et al.*, 2005)

Clustergrams are discussed here as an improved method of identifying the optimal number of initial seeds input into the K-means clustering algorithm. By removing some of the iterative testing and user consultation associated with other methods such as elbow plots, it is anticipated that utilising Clustergrams will increase the reliability of the pre-determined optimal number of clusters.

It is also important to caveat that aggregating the DEP and EPC data from an individual consumer or household view to Postcode Sector level gives rise to issues of ecological fallacy and the modifiable areal unit problem (MAUP) (Openshaw, 1984). For instance, analyses conducted on DEP data may be subject to zonation effects, given that census geographies and the DEP data are derived from different base populations and scale effects, arising from the fact that the DEP data is limited in its spatial granularity. Any outcomes, which are the majority in this case, generated from analysis on the aggregated DEP data will be subject to these limitations; justifiable by the necessity for data to be available for analysis of the data outside of the secure facility, to ensure non-disclosive presentation of results and to facilitate linkage to census data in order to contextualise the novel data in terms of general population characteristics.

4 The Geodemographics of Energy

4.1 Introduction

The aims of this chapter are twofold; firstly to describe the characteristics of the smart meter data provided by the ESRC Consumer Data Research Centre; and to map aggregate patterns of domestic energy consumption: specifically exploring issues of representativeness, and spatio-temporal signatures of aggregate residential consumption. Secondly to evaluate the socio-economic determinants of energy consumption and their relationship to deprivation and in particular, fuel poverty. To the author's knowledge this is the first dataset of its kind to be analysed at the half hourly cadence for the national extent, and the first energy study in the UK to analyse both gas and electricity consumption in tandem with demographic characteristics to understand contextual aspects of fuel poverty.

In existing energy literature, there is disparity across definitions of the term "energy", depending on the focus of the study. In many cases it is taken to mean specifically electricity but may also refer to both gas and electricity (Druckman and Jackson, 2008; Jones and Lomas, 2015; McLoughlin *et al.*, 2015; Viegas *et al.*, 2016). In this thesis, references to 'energy' are taken to mean both gas and electricity, and as such this chapter aims to provide a more comprehensive assessment of a household's overall consumption. This chapter begins to unpack the myriad indicators of fuel consumption, the barriers to access and improvements and the relative benefits that Smart technologies could provide if their full utility is realised.

4.2 The Domestic Energy Provider Dataset

The national dataset of smart meter readings were sourced through the ESRC Consumer Data Research Centre (CDRC) and relate to one of the UKs Big Six energy suppliers. The data contains details of around 1,080,000 gas and electricity domestic smart meters, providing meter readings at a half hourly cadence for the 2015/16 financial year, representing 43% of the 2.3 million smart meters installed at that stage. The spatial granularity is at Postcode Sector. Given the timing of the data collection, it is likely that all the meters from which readings were collected were the first generation SMET1 smart meters.

4.2.1 Representativeness in the Smart Meter Dataset

The dataset represents the early stages of smart meter rollout. There are biases associated with this, primarily that those with the oldest meters were the first properties to receive an upgrade. The second most prevalent source of bias arise from the fact that the first households to receive an installation were more likely to be at home during the campaign; skewing the customer representativeness towards the elderly, families and the unemployed (Chapter 3). Indeed, the DECC (2014a) found that smart meter installations into domestic properties had been as a result of the majority of consumers being contacted by energy suppliers rather than consumers proactively requesting one (84% vs 5%).

4.3 Spatial and Temporal Trends Within the Smart Meter Dataset

As discussed in Chapter 3, smart meters are installed at an address level and so it is possible to utilise the installation figures as a proxy for coverage. Ushakova *et al* (2018) split the dataset by energy type, but to adhere to the definition of energy discussed at the beginning of this chapter, the penetration of all smart meters regardless of their energy type are shown in Figure 4-1 to explore the overall spatiality prior to the data aggregation. It displays the penetration of smart meters as a percentage of total homes in each PCS (Postcode Sector). This corroborates previous studies and does suggest that there are some urban/rural variations in the ease of physically accessing housing stock, ease of installation into homes, or availability of the prerequisite smart meter infrastructure such as access to fixed line broadband, all of which become more challenging in rural areas (Sovacool, 2015). During the 2015-16 financial year when this data was collected, the DEP had no higher than 16% penetration anywhere in the UK. It is important to consider the size and spread of this particular DEPs customer base; it may be the case that there are smart meters provided by other DEPs in areas which look sparse and so this dataset should not be considered wholly representative of the entire population.

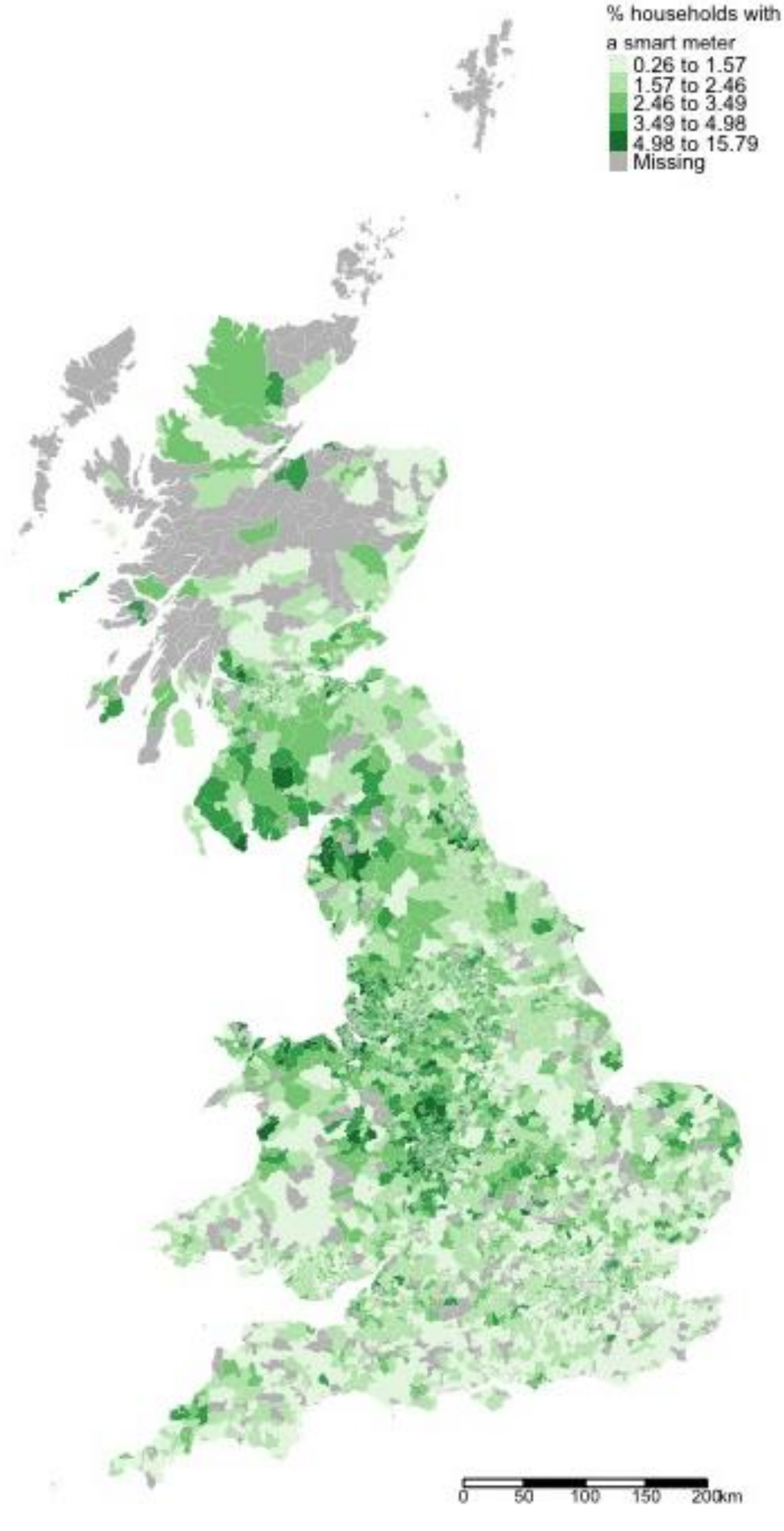


FIGURE 4-1 SMART METER PENETRATION RATES ACROSS GREAT BRITAIN

As previously mentioned, household smart meter installations can be used as a proxy for coverage and whilst Ushakova *et al* (2018) investigates the percentage of homes with smart meters across the country, this research builds on this by attempting to understand the extent to which this spatial distribution has occurred randomly. The Local Indicators of Spatial Autocorrelation (LISA) methodology attempts to identify and quantify local patterns of spatial association and the results can be interpreted as indicators of local 'hotspots' (Anselin, 1995). A Morans I statistic quantifies the extent of the spatial autocorrelation, or to what degree similar features cluster over space. In this instance neighbours were defined as contiguous polygons. The results of the LISA analysis are displayed in Figure 4-2, where clear hotspots (and cold spots) have been identified.

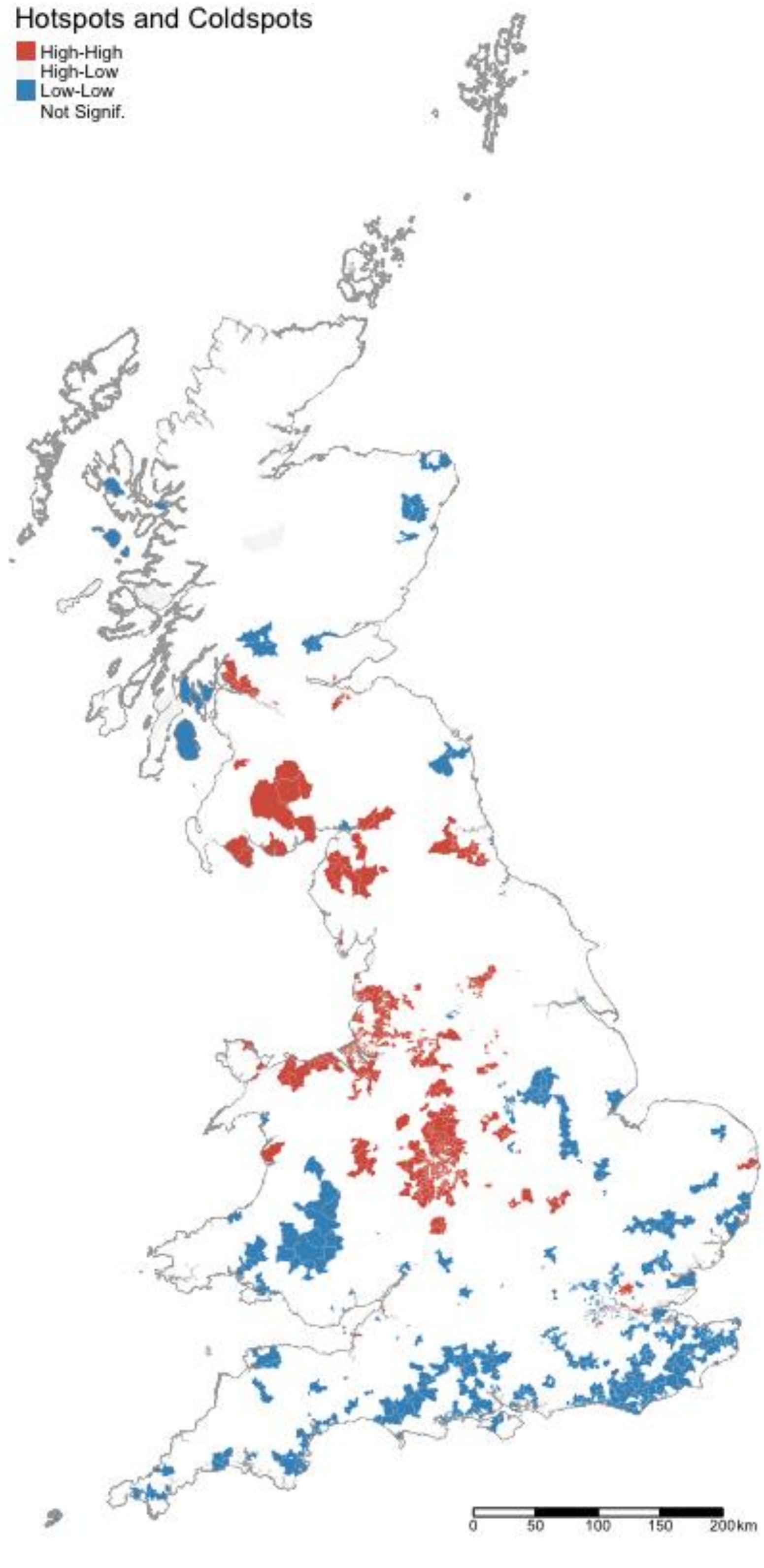


FIGURE 4-2 LISA ANALYSIS OF SMART METER ADOPTION AT POSTCODE SECTOR LEVEL

There is a clear north south divide, and whilst any reasoning for this would be purely speculative because of the lack of information regarding the DEPs rollout programme, it is true to say that in the North of the UK clusters of high penetration areas tend to be more prevalent, whereas in the South of the UK the exact opposite is true with the pattern of low/low areas particularly prominent along the south coast. A Global Morans I statistics had a positive value of 0.51, with a p value less than 0.05 suggesting that this spatial clustering has not occurred randomly. This clustering could be symptomatic of where the DEP has the strongest customer base or could suggest some level of targeting consumers in the North.

4.3.1 Data Cleaning and Description

Before a temporal analysis of the DEP smart meter dataset was undertaken, data cleaning steps were considered in order to ensure that the data was fit for purpose and did not contain potentially sensitive individual data records. Initially, data cleaning was undertaken to remove individual records with null values or lack of spatial attribution. Exact duplicates and rows with missing categorical data such as postcode attributes or meter type were removed as these could not be accurately imputed. .

visualizes this data cleaning process, which also acted as a data minimization method; the final aggregated dataset represents a 97% reduction in size, making the processing time of the proceeding analysis less arduous and computationally expensive (Jiawei *et al.*, 2012).

TABLE 4-1 THE DATA CLEANING AND MINIMISATION PROCESS

Step	Number of records removed	Percentage records removed	Total number remaining records
Prior to data cleaning			292,855,095
Removing >50% 'NA' consumption data	1,275	0.0004	292,853,820
Remove 'NA' categorical data	66	0.00002	292,853,754
Aggregation to Postcode Sector			6,141,494

Because of the nature of this innovative dataset, it is difficult to know whether or not a zero value for a meter reading is a true zero where no energy has been consumed or is a false reading. We already know that SMET1 meters were prone to technical faults causing missed readings and it is unlikely that a true zero would appear in the dataset due to 'standby energy consumption' – the small units that are consumed by appliances always left on such as fridge freezers or security systems (Wyatt, 2013). In cases where over half of the days meter readings were zero, the entire row was removed as it would have been difficult to accurately impute and risked introducing error into the overall results. The frequency of the missing records would also suggest a fault with either the meter itself or the signal strength required to feedback to the DEP. Where under half of the meter readings were zero, the value was imputed using the mean value of the postcode sector in order to provide the most complete version of the dataset for analysis prior to aggregation (Lavin and Klabjan, 2015). Cleaning and preparation steps were taken to ensure usability and accuracy in the later analysis; an essential step in order to avoid influencing the results with atypical values (Ramos and Vale, 2008).

Other instances of unrealistic energy consumption were also considered at this stage. For each household, unrealistic consumption was considered to be anything above 3 standard deviations and so were removed. A value over 3 ʃ from the mean should be considered an extreme outlier (Field *et al.*, 2012). Other values that were considered outliers to a lesser extreme were kept as this variability could be accounted for by different devices being present; for example a household with an electric car will use significantly more energy during overnight hours, but should not be removed from the dataset.

Finally, the data were manipulated to preserve the anonymity of the users and prevent re-identification as per the requirements imposed by the DEP and the CDRC. Having already understood the representativeness of the individualised dataset, it was aggregated to Postcode Sector Level and any sector containing less than 10 records was removed as per the secure data policy to avoid any individual being re-identified. By aggregating to Postcode Sector, the dataset was reduced to 6,141,494 records in total, which represents a single row per Postcode Sector for each day and meter type.

4.4 Consumption Trends

At a diurnal granularity, energy usage has two clear peaks as can be seen in Figure 4-3. There is a minor peak between 07:30 and 08:00 as people begin their daily tasks, before dropping as people leave their homes to go out to work and school. There is a slight increase as people who are at home during the day, such as those with caring responsibilities, shift workers or the unemployed undertake their lunchtime routines, before falling away again until around 17:00 when consumption increases steadily to the major peak at around 19:30, when typically most family members are at home and engaged in energy consuming activities such as cooking, cleaning and relaxing (television, tablets and games consoles) as well as homes requiring greater levels of lighting in the evening as darkness falls outside, which remain in constant use until people go to bed (Yohanis *et al.*, 2008).

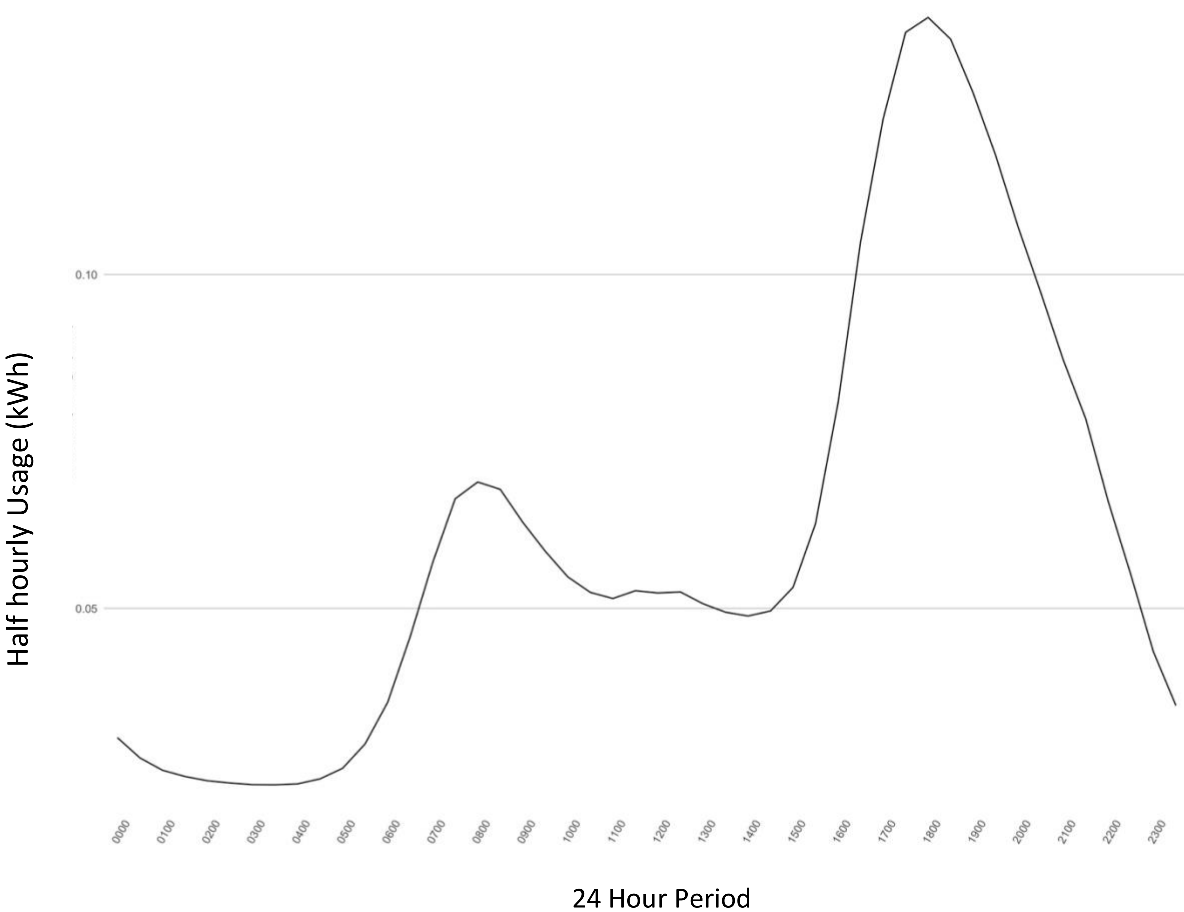


FIGURE 4-3 HOURLY RATES OF TOTAL ENERGY CONSUMPTION

After 21:00 domestic activities slow and family members go to bed, making an effort to turn off almost all of their devices, as this is the time when most people recognise that leaving devices powered up unnecessarily is wasteful. Overnight consumption represents the minimum load, and as previously mentioned is sometimes called 'standby usage' or 'base load', which is made up of two components; usage by appliances that require constant power such as the fridge freezer or security systems, and reduced usage from devices (usually entertainment systems) which consume power whilst in an unused state. More frequently, overnight consumption also occurs when mobile devices are plugged in to recharge overnight for use the next day and for convenience – increasingly entertainment systems operate wirelessly, with one central control box being left on overnight to enable for example, family members to all watch television or access the wi-fi in their bedrooms (Wyatt, 2013). This pattern is an aggregate consumption profile across a 7 day week, and so the levels of daytime usage include weekend days where families are much more likely to be at home engaging in leisure activities. Daily usage is disaggregated by day of the week to examine this in greater detail in the following section. When the daily usage is disaggregated for individual days of the week, the consumption patterns confirm these notions, Figure 4-4 shows the split by day.

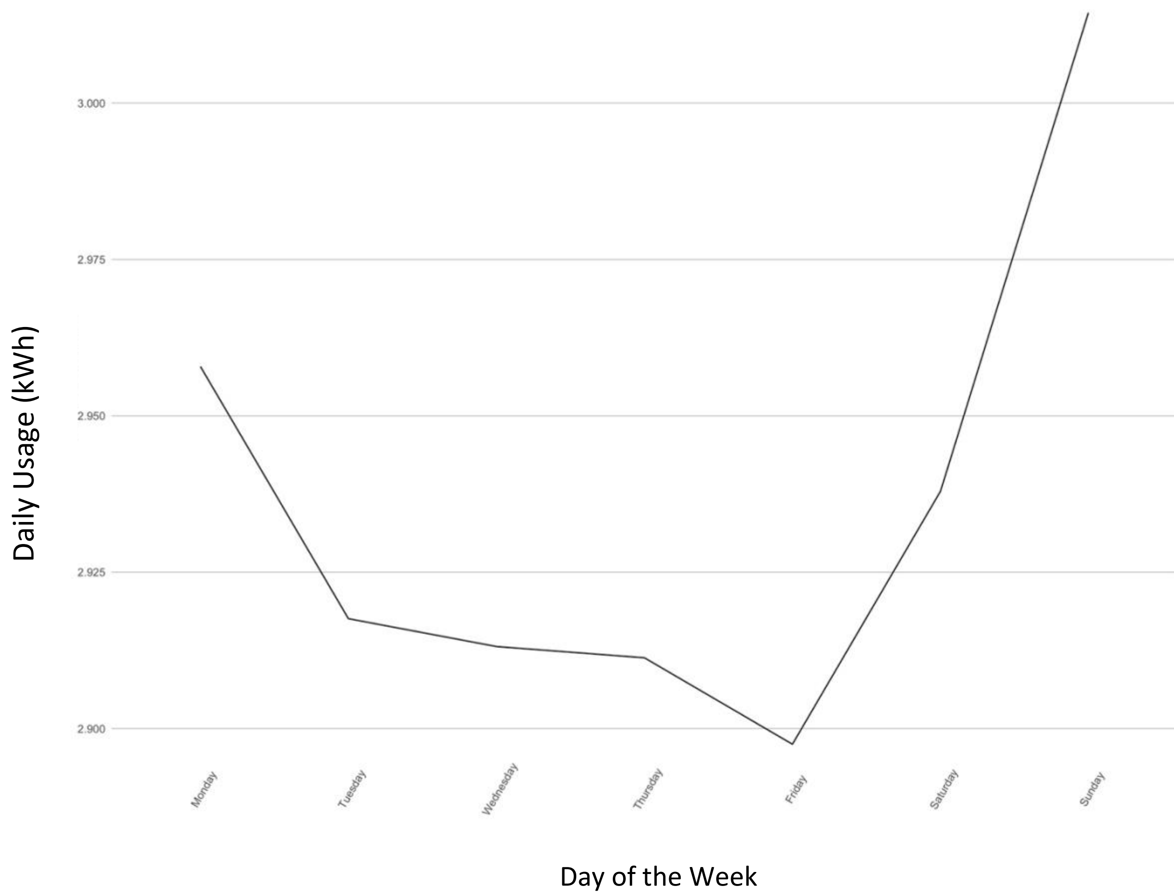


FIGURE 4-4 TOTAL DAILY CONSUMPTION

Figure 4-4 displays the total energy usage for days of the week, reflecting a typical weekly routine – lower usage during the working week whilst residents are out at work and school; and higher consumption at weekends when people tend to be at home engaging in leisure activities and chores. Friday's are the lowest day of the week, which could be the result of people both being at work all day and staying out of the home into the evening to socialise as the weekend begins and conversely Sunday's are proportionally higher, perhaps reflecting British traditions of a Sunday Roast – having all the family present and spending lots of time cooking. It might also indicate preparations for the week ahead – laundry, ironing, bathing etc.

When examining the disaggregated half hourly data by day of the week, these patterns are confirmed and show a clear difference in usage between the working week and the weekend. Figure 4-5 shows that the minor AM peak starts later at weekends; families have less pressure to leave the house to start work and school, and usage is higher throughout the day as households engage in leisure activities in the home, but evening peaks on Fridays and Saturdays are the lowest, suggesting that evening entertainment at weekends takes place away from the home – despite Friday's being relatively similar to the rest of the working week, it is this reduction in evening consumption that leads to it being the lowest day of usage overall. The timing of the major evening peaks are relatively similar as natural light fades consistently.

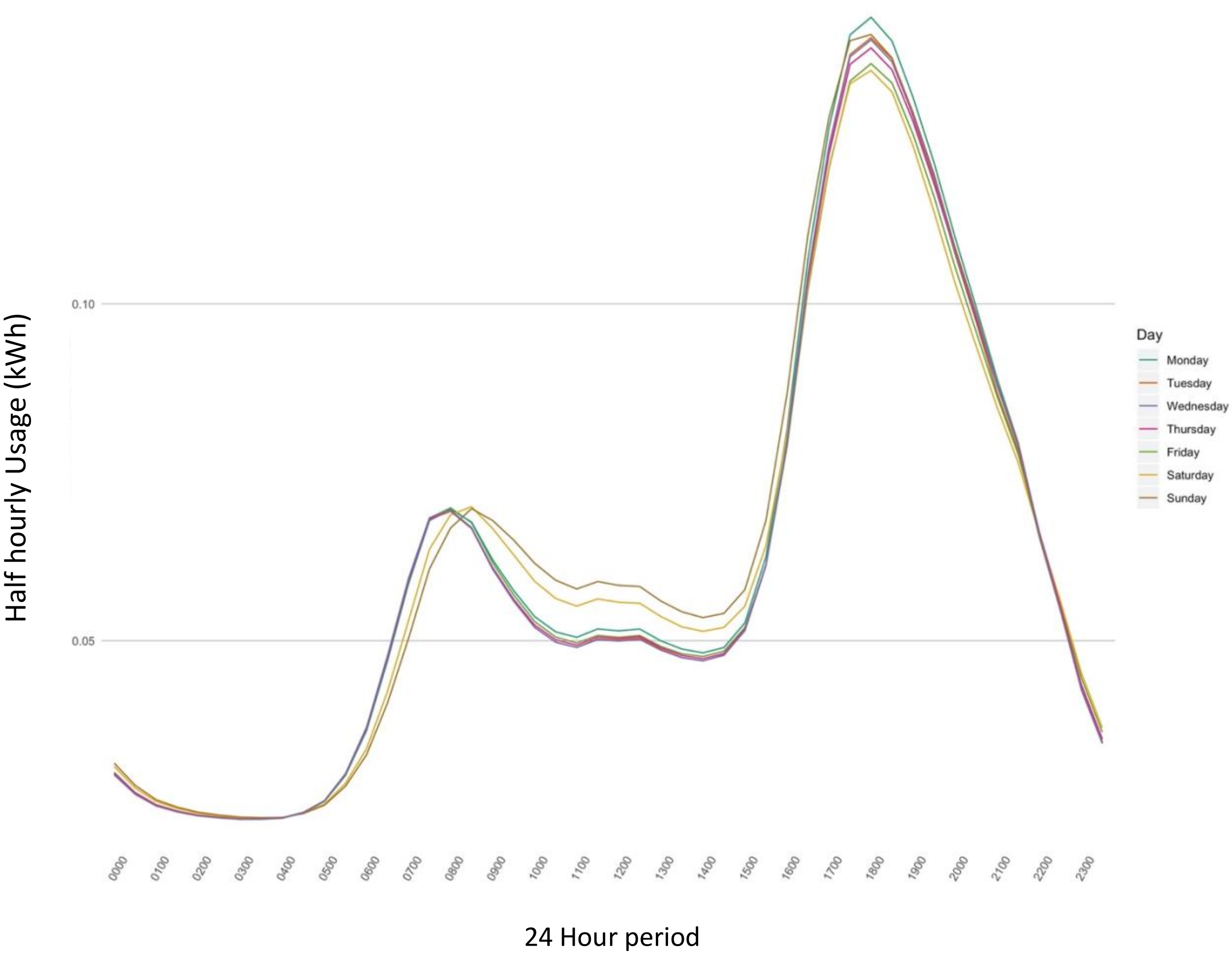


FIGURE 4-5 HALF HOURLY DISAGGREGATE USAGE BY DAY

Figure 4-6 shows that there are also seasonal patterns reflecting the UKs temperate climate; very high usage in winter when heating and lighting requirements are much greater due to colder weather and longer hours of darkness than in the summer. The lowest usage between June and August is reflective of people not only needing less heating and lighting but also taking advantage of other energy saving measures such as the ability to dry clothes and occasionally cook outside in warmer weather. People are also more likely to undertake leisure activities outside the house in good weather, resulting in lower consumption levels by devices and electronic equipment. It might also be indicative of patterns of extended low use as people take a summer holiday and leave their property unoccupied for several weeks at a time.

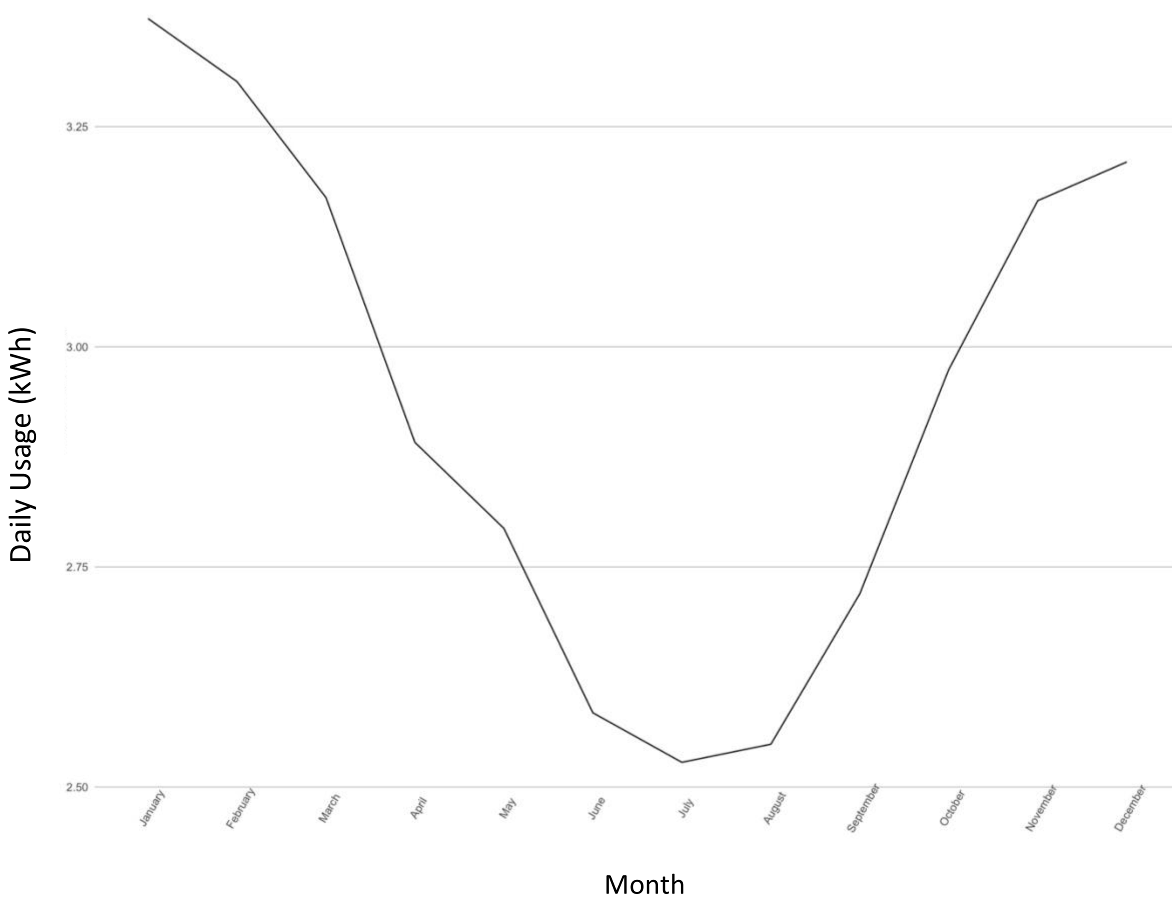


FIGURE 4-6 SEASONAL RATES OF TOTAL ENERGY CONSUMPTION

The half hourly disaggregate figures for each month reiterate these patterns and offer a more detailed insight; Figure 4-7 reflects the higher usage overall in winter months (December, January and February), which occurs due to sustained higher consumption throughout the day. The summer months of June July and August show the biggest variation from the overall trend with much less variation in maximum and minimum consumption – the evening peak is slightly earlier and greatly reduced, and the decline to the 'stand-by rate' is much more gradual as the requirement for heat and light remains lower throughout the day.

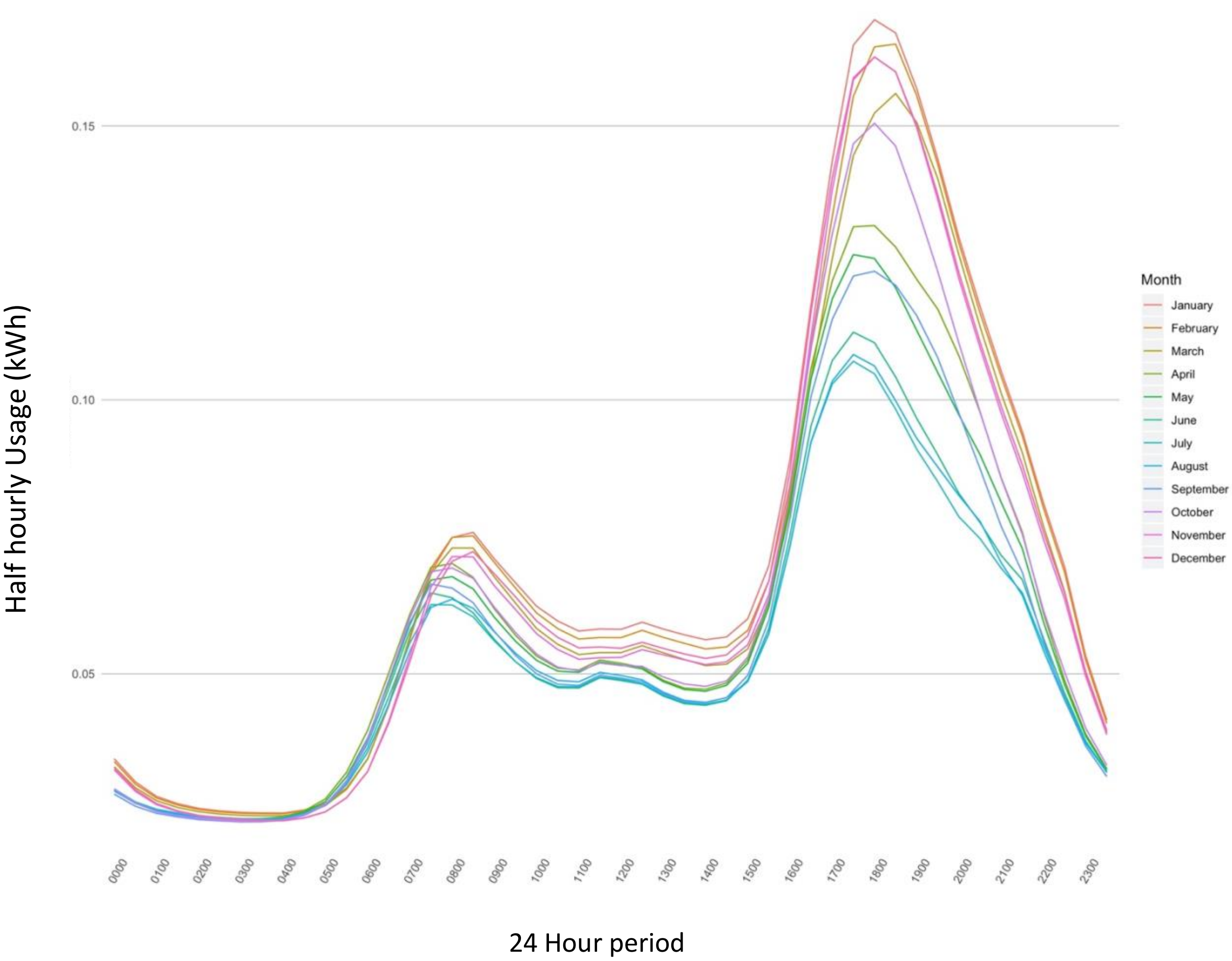


FIGURE 4-7 HALF HOURLY DISAGGREGATE CONSUMPTION BY MONTH

4.5 Characteristics of Smart Meter Users

For the study of populations, geodemographics represent the analysis of people by where they live (See Chapter 3) (Longley, 2017). They have utility in both commercial and academic settings and have many applications to aid the understanding of the relationship between population characteristics and consumer behaviours and provide contextual validation to the smart meter data. The following section takes the Output Area Classification (OAC) to quantify the relationship between population characteristics and their energy consumption and smart meter adoption rates. The most recent iteration of the OAC is generated from the 2011 census by the ONS; it comprises 8 Supergroups, 26 Groups and 76 Subgroups (See Chapter 3) (Gale *et al.*, 2016). Because of the need to reweight the OAC classification to Postcode Sector Level (See Section 3.5) the rates of smart meter adoption in individual areas lead to sparse results when combined with the 76 Subgroups and so the 26 Groups were utilised. Figure 4-8 shows the smart meter adoption rate within each OAC Group.

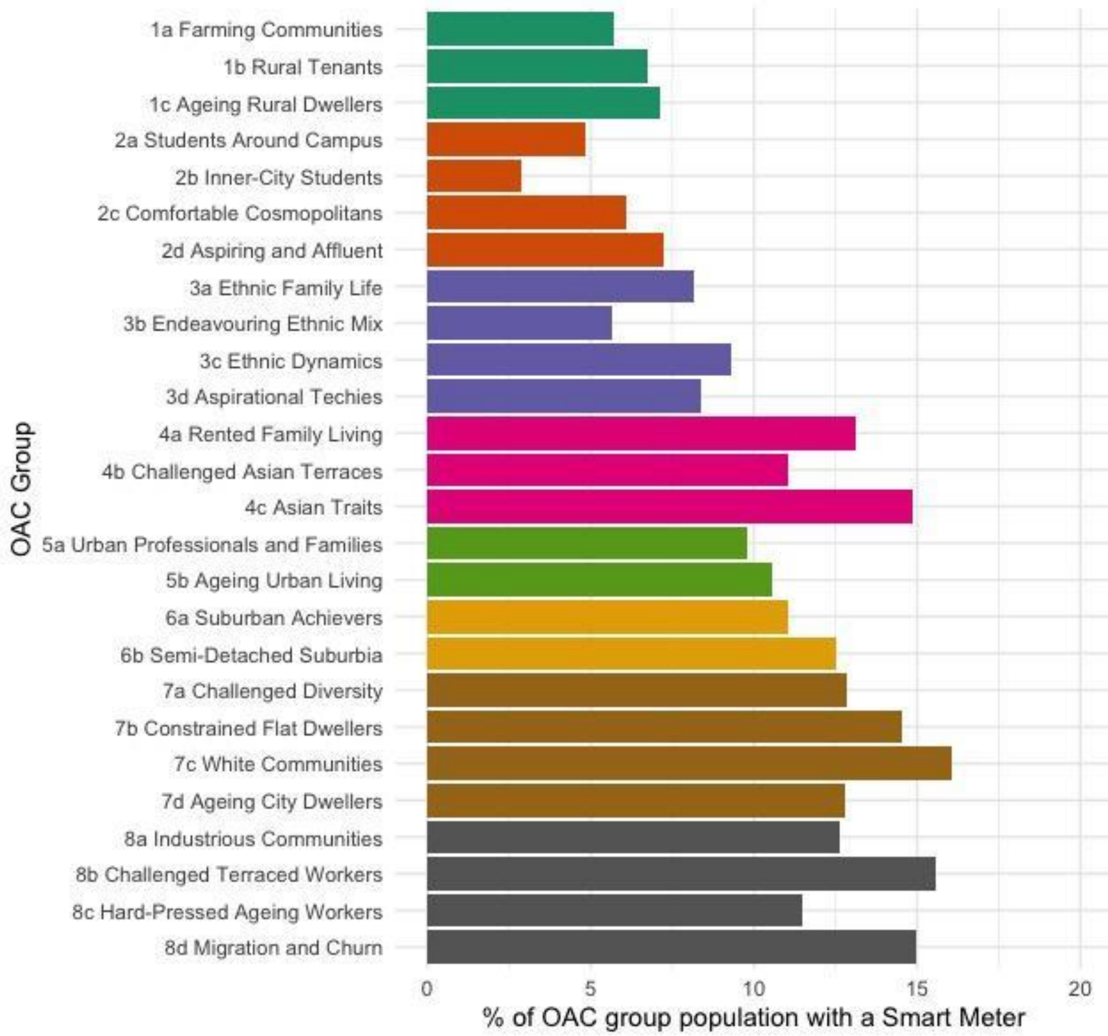


FIGURE 4-8 SMART METER ADOPTION BY OAC GROUP

The adoption of smart meters amongst these different groups reiterate bias within the data. There is an obvious urban rural split, with groups in Supergroup 1 (1a, 1b and 1c) who are all considered rural dwellers having some of the lowest proportions of smart meter users.

Whilst rurality appears to contribute to the low propensity, possibly as a consequence of physical access to the properties (Section 2.3.3) , these groups characteristics indicate that they are also likely to be employed in "agriculture, forestry and fishing industries" and therefore away from their properties during the day, meaning they are less likely to be exposed to doorstep marketing.

Furthermore, "an above average number of people live in communal establishments (most likely to be retirement homes)" indicating that occupants are unlikely to have an individual meter for which they are responsible. These patterns both reiterate the assumed bias noted by Ushakova *et al* (2018). Group 7C (White Communities) have the highest proportion of homes with a smart meter installed, and their profile suggests that they are "more likely to own their semi-detached and terraced properties" confirming the notion that having the autonomy to make structural changes to a household leads to a higher rate of successful installations than those that are limited by contractual obligations to a landlord.

Group 4C (Asian Traits) are particularly interesting, as their profile suggests they are likely to be "owner occupiers of detached and semi-detached homes" and work in "industries associated with information, communication and finance". This technological and financial awareness might be indicative of increased knowledgeability of modern technology and its positive implications for energy and monetary savings, hence their higher smart meter propensity rate relative to other members of their Supergroup.

Existing literature suggests that material deprivation and fuel poverty are closely linked and also that demographic characteristics have an influence over a household's ability to consume energy. The link between smart meters and fuel poverty is important as the UK government utilise this scheme to improve energy efficiencies, particularly for the most vulnerable. To better understand the link between material deprivation and a household's ability to consume energy, total energy consumption was also profiled by OAC Group, as described by Figure 4-9.

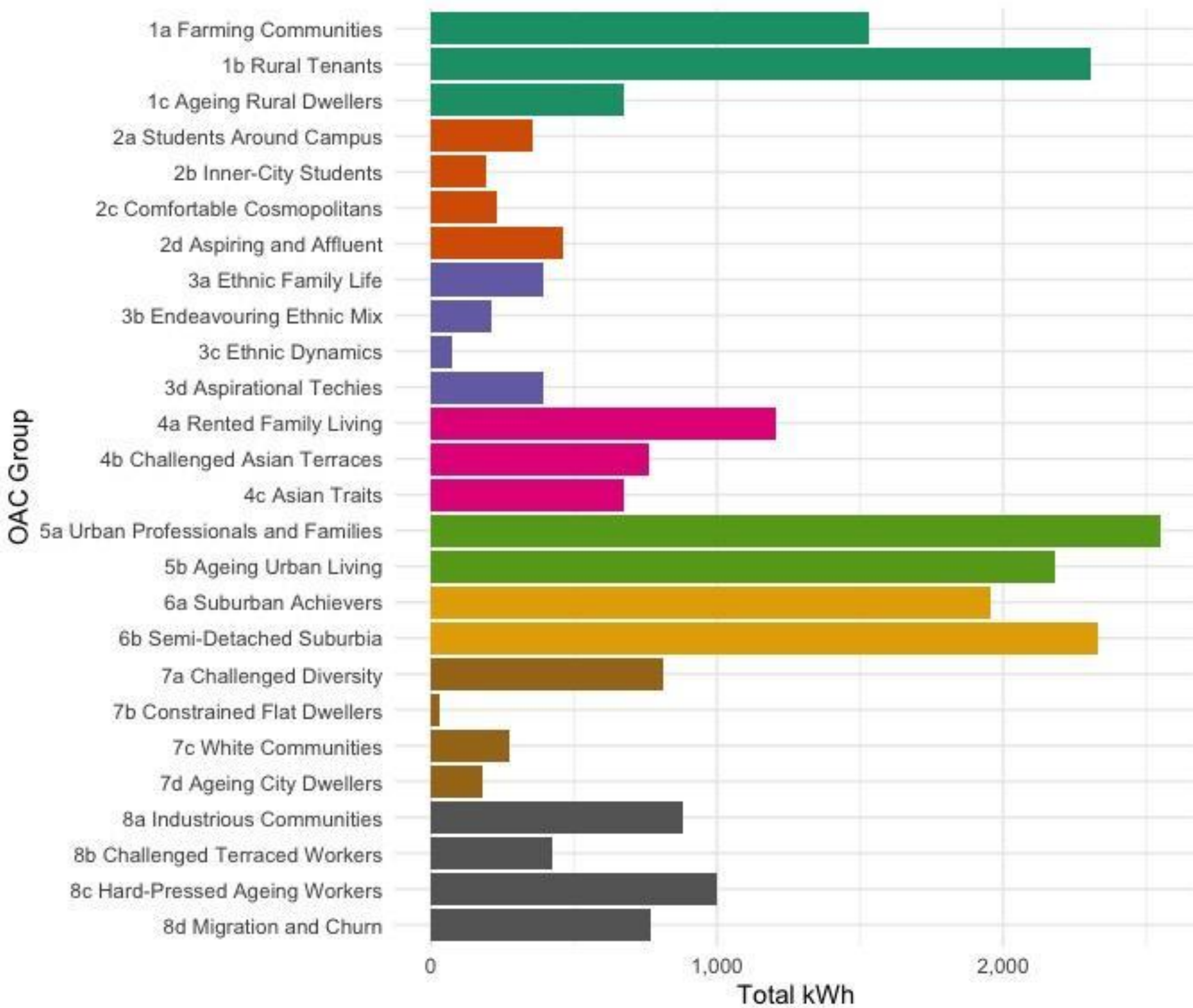


FIGURE 4-9 TOTAL ANNUAL CONSUMPTION BY OAC GROUP

The most interesting patterns occur at the highest and lowest usages. Group 6B 'Semi-detached Suburbia' use the most energy overall. Their pen portrait indicates that a high proportion of this group live in rented accommodation, corroborating the findings in existing literature that rented homes are the least likely to be energy efficient, thus consuming higher amounts of energy overall to achieve comfortable living conditions (See section 2.2.3). Furthermore, despite the fact that these homes are the ones where smart meters are installed, they are of no cost to the landlord and structurally, these homes are still some of the most inefficient housing stock in the UK and most likely to suffer from the spilt-incentive problem where landlords are unwilling to invest in further efficiency measures that require greater capital investment. (Section 2.2.5). By investigating the characteristics of the Subgroups within Group 6B, further high usage characteristics are revealed; Subgroup 6B1 ('Multi-ethnic Suburbia') are more likely to live in overcrowded conditions; more devices and appliances are going to be in use and more rooms will have a requirement for heat and light to meet the needs of more people. Subgroup 6B3 are also mostly aged 65-89, therefore retired and at home throughout the day, and may also have a greater requirement for heat in order to manage age related health complaints, which are often compounded or made worse by underheating (Section 2.2.6).

Those in Group 1B; 'Rural Tenants' are also worthy of note. Despite their relatively low propensity of smart meter ownership, those that do own one display relatively high usage. This is most likely due to the age structure of the group who are mostly middle aged and retired, with the associated energy requirements as discussed above. They also have an increased likelihood of living in rented accommodation than others in their Supergroup (1A and 1C display much lower usage). It might also be indicative of the fact that rural properties overall are more likely to be built with solid walls rather than cavity wall, which makes retrofitting energy efficiency measures more difficult (Roberts *et al.*, 2015).

Group 7B 'Constrained flat dwellers' have the lowest energy use. The pen portrait indicates several characteristics associated with material deprivation as a whole, such as living in socially rented accommodation and owning fewer cars. For this reason, it might be fair to assume that their very low consumption is a symptom of those households restricting their usage either to avoid, or because they already find themselves in fuel poverty. However, their usage is already low given their lower square footage, number of rooms and fewer people per household. Group 3C 'Ethnic Dynamics' use only slightly more energy, which can again be attributed to also living within accommodation where there is a lower square footage than average and having economic constraints through unemployment; but this may also explain the slight increase; some may use slightly more energy due to being more likely to occupy the home during the day.

4.6 Discussion and Conclusion

The aim of this analysis was firstly to address the issues of representativeness in the smart meter dataset and demonstrate the utility of smart meter energy consumption data for describing high -level aggregate consumption patterns. The results of the LISA analysis demonstrated that the bias towards locations in the north of England and Wales had not occurred randomly and is most likely present as an effect of the DEPs existing customer base and roll-out programme. It also indicated an urban/rural bias, perhaps indicative of the infancy of the roll-out programme at the time as well as physical access constraints at rural properties.

The results of the aggregate consumption patterns show clear patterns at various temporal granularities and suggest that these smart meter users consume energy in a way that is a fair representation of everyday life; with clear, sensible waking times and evening peaks, weekday and weekend variations and seasonal consumption patterns that reflect the warmer and lighter summer months and colder, darker winter months. The daily 'two peak' pattern is also reflected in other research which utilises smart meter consumption data to estimate diurnal patterns (Haben, Singleton and Grinrod, 2016). Despite this, when the smart meter consumption data is transformed using the equation to convert watts into kWh (Section 3.2), both the temporal consumption profiles and the total annual consumption by OAC group display values that would be considered below the national average for even the lowest users. These estimate 9,900 kWh of combined gas and electricity consumption annually (Section 2.3). This raises questions of data quality, and the requirement for additional validation of the smart meter data. Again, it is not possible to know if the DEP roll out programme deliberately targeted consumers with a lower than average rate of consumption. It is also a possibility that the multiplying factor used in the equation to transform half hourly data into kWh has been misinterpreted, and so would benefit from an increased understanding of the applicability of this equation in respect of data that is presented in watts per half hour in it's raw format.

Secondly this chapter aims to understand the intersection between smart meter adoption and OAC groups, generating insight into the demographic characteristics associated with adoption and consumption. The results showed interesting patterns particularly in rural areas, highlighting potential constraints to installation. Ageing rural tenants are the least likely to have a smart meter install which could be linked to their properties not being attached to mains gas and therefore no meter is required, or could be indicative of additional constraints in accessing smart metering technology in rural areas; more limited availability of engineers given the increased burden of travelling out to rural homes, technological constraints such as limited broadband speeds or older properties with thick walls making them unsuitable for the first generation of smart meter technology. In addition to this, the results also suggested that OAC groups with high prevalence of rented tenants were very unlikely to have smart meters installed, re-enforcing the idea that tenants have very little autonomy over their household's energy efficiency and contractual constraints imposed by landlords mean that tenants are unable to have smart meters installed without their permission.

This exploration into the geodemographic characteristics associated with varying levels of energy consumption provides utility in framing the link between energy consumption and contributing contextual factors which begin to build a narrative around the need for a multifaceted fuel poverty definition; and has also provided a level of external validation to the data. Those patterns exposed within the energy use are consistent with hypothesised usage that one might expect given the characteristics of those people and the places in which they live, as identified by the Output Area Classification. It has implications for proceeding analysis in informing the characteristics which are most prevalent in smart meter adoption and under consumption of energy.

5 Reconsidering Fuel Poverty Through The Energy User Classification

5.1 Introduction

The previous chapter investigated patterns of access to upgraded energy efficiency technologies such as smart meters, revealing geographical disparities. Profiling by the Output Area Classification (OAC) revealed that propensity was not evenly distributed across all groups, suggesting that access is likely to be influenced by a number of factors as well as a broadly suggesting inequity in the prioritisation of infrastructure upgrades, forming a basis to support a conceptual framework for an energy based classification. Attributes pertaining to age, population density and occupation as well as some more physical attributes such as accommodation type, building type and ownership have all been discussed previously as having links to overall consumption and would likely show utility in building a typology by introducing measures of demographic characteristics to supplement consumption data, and as such, create a broader view of energy efficiency and fuel poverty.

Based on these findings, this chapter integrates Energy Performance Certificate data (EPC) alongside demographic measures to challenge the robustness of the current fuel poverty definition and the extent to which a household's ability to consume energy is impacted by factors other than those purely monetary. Given that the Government targets for improving energy efficiency centre around improving their EPC rating (Band C by 2030', with interim milestones of 'Band E by 2020' and 'Band D by 2025' (Department of Energy and Climate Change, 2015)), there is logic to the inclusion of these data in analysis with a view that they provide different insight beyond those social determinants of energy use.

The proceeding chapter is structured as follows; the literature in Section 5.1 reiterates the current fuel poverty definition and addresses why this definition does not fully encompass the lived experience of fuel poverty and argues for a much broader, multifaceted definition. Section 5.2 introduces the data on which current fuel poverty statistics are based, as well as the EPC dataset and the energy efficiency of the OAC Groups in order to understand what high level demographic variation exists, before Section 5.3 introduces the methodological approach to building an Energy Consumption typology. The Chapter concludes with Section 5.4, which summarises the results of the Energy User Classification, with pen portraits for the four Supergroups.

5.2 Energy in Context

5.2.1 Current Fuel Poverty Statistics

The Government produced an openly available fuel poverty statistic for the year 2016 which was disseminated at the LSOA level (Section 3.4.8). Because the overarching aim of this chapter is to prove the utility of external datasets in quantifying fuel poverty beyond the current definition, providing directly comparable results for the current statistics required the reweighting of the current fuel poverty data. Figure 5-1 shows the spread of fuel poverty across England and Wales for all areas where data was available, by proportion of houses per PCS. Some spatially based characteristics enhance the likelihood of a household experiencing fuel poverty; the material and infrastructural characterisation of an area as well as aggregated attributes such as demographics vary between different household types and therefore also geographically (Robinson *et al.*, 2018b).

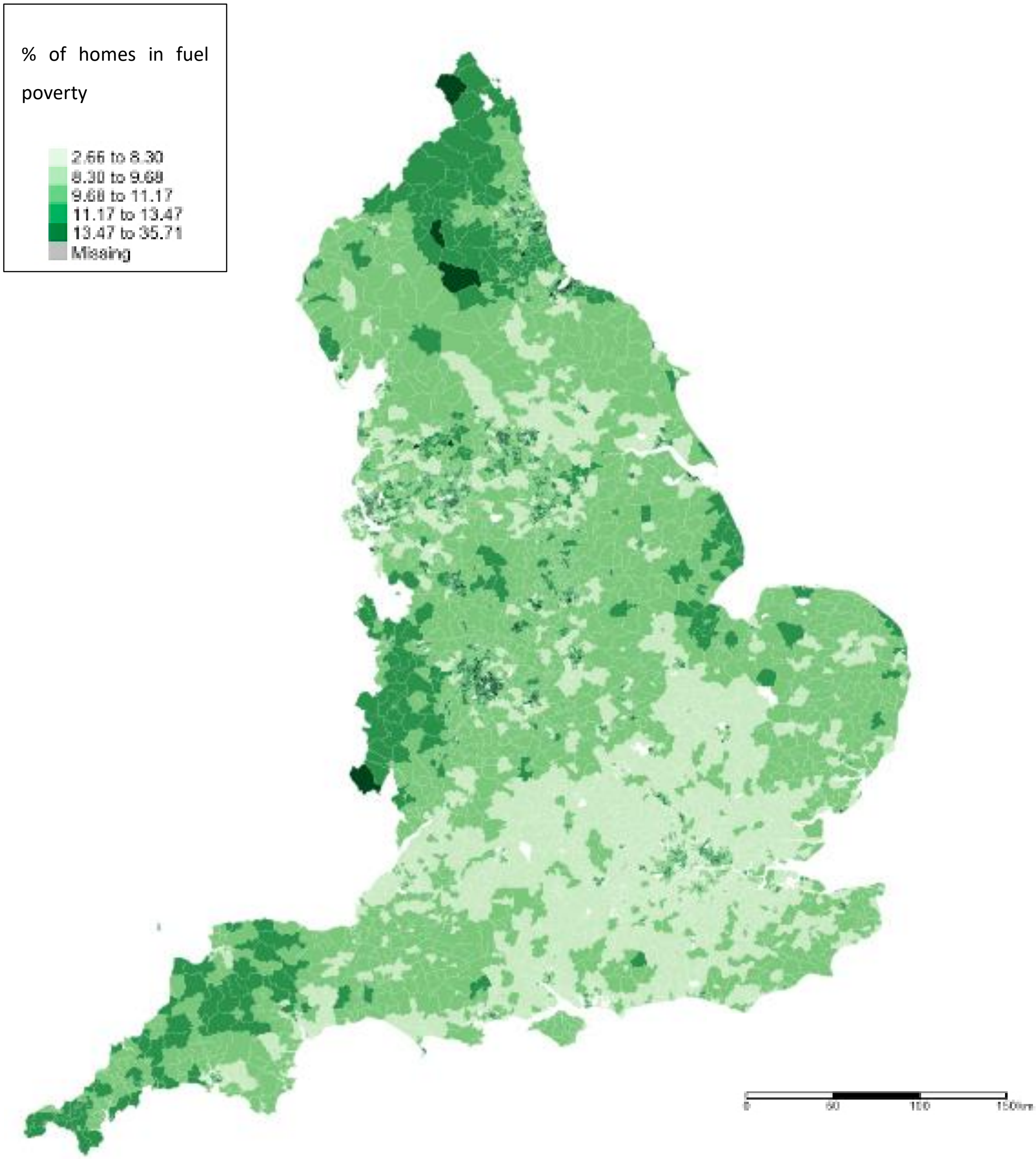


FIGURE 5-1 CURRENT DISTRIBUTION OF FUEL POVERTY IN ENGLAND

Figure 5-1 shows there is a clear spatial disparity between the North and South of England, where the North has a much higher percentage of fuel poor households than the South. There is also some evidence of a disparity between urban and rural areas, both of which are likely due in part to the fact that the LIHC indicator equivalises fuel costs and income. Equivalisation adjusts household income based upon different demands for resources, considering household size and composition, thus reflecting how larger households require more energy to heat and tend to have less disposable income than smaller households. This results in an under-representation of larger under-occupied households in fuel poverty. Under-occupancy is most common in owner-occupied homes, most prevalent in rural areas (Robinson *et al.*, 2018b). Also contributing to the relatively urban nature of fuel poverty is the fact that inner city areas are often disproportionately affected by inefficient housing in the private rented sector, where tenants lack housing rights and access to retrofitting schemes (Hope and Booth, 2014; Robinson *et al.*, 2018a). Data for Scotland and Wales is not available; these data represent experimental statistics collected from the *'English Housing Survey'* when they were published and were the best available.

5.2.2 Before and After Housing Costs

The Hills definition uses AHC measure to make fuel poverty a relative measure; the equivalisation of incomes is intended to make low and high income households more directly comparable. The AHC measure has benefits; it is much more representative; a household cannot spend their housing cost on fuel, but what the annualised statistics on which the reporting figures are based do not take into account is the fact that especially for those populations who are trapped in precarious tenancies, housing cost changes frequently and thus so does the disposable income available to spend on fuel. Under new Government policies such as Universal Credit, which disproportionately affects those on a low income, the amount of benefit a household is eligible for can vary month to month, making budgeting and forward planning difficult. Owner-occupied houses are less likely to suffer from this as mortgages provide some level of stability with regard to consistent payments. By investigating the change over time in BHC and AHC from the Small Area Income Statistics dataset (Section 0), it is possible to get an overview of the volatility in changing incomes between 2012 and 2016; the analysis below shows annual changes in the percentage of income accounted for by housing costs. Figure 5-2 shows that for the lowest income quintile, the gap created by housing costs widens over time, whereas higher earning groups have seen their housing cost relative to their income remain steady.

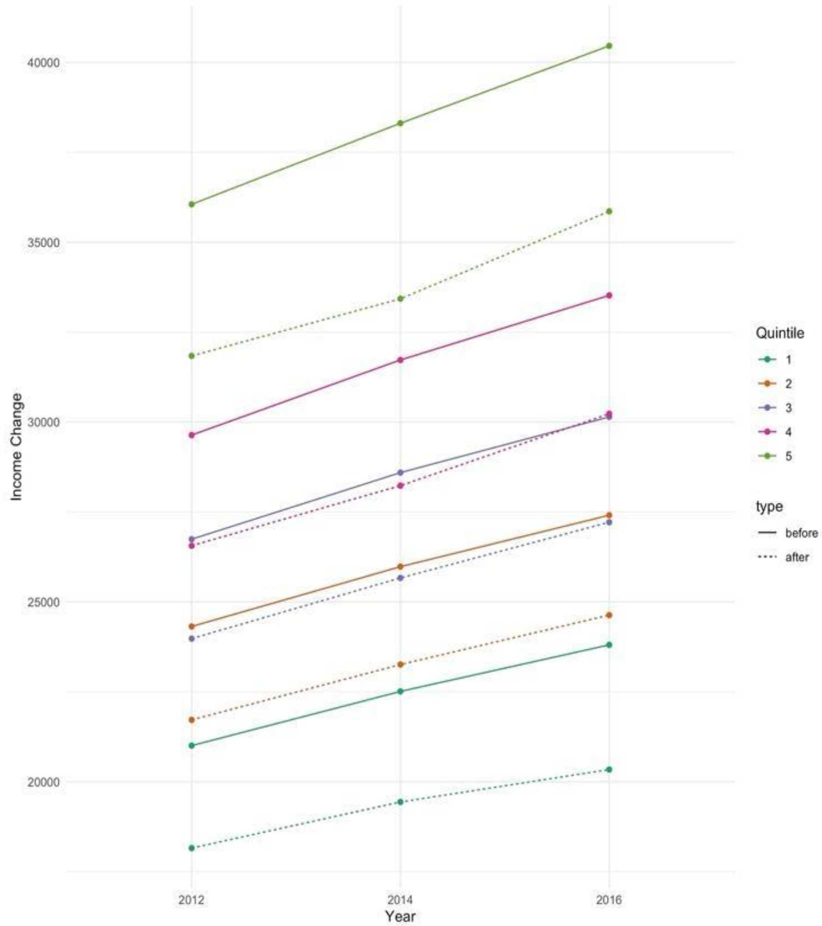


FIGURE 5-2 POSTCODE SECTOR WEIGHTED INCOME QUINTILES SHOWING CHANGES IN INCOME BEFORE AND AFTER HOUSING COSTS 2012 - 2016

The percentage change in AHC is shown in Figure 5-3 and reiterates that the lowest earners have seen the largest percentage housing cost increase; in 2012 14.25% of income was accounted for by housing cost which had risen to 15.75% by 2016 while salaries in this group only increased by 11% - hence the widening gap. This relative decrease in disposable income may lead to those households finding themselves in either short term fuel poverty while their finances recover, or perpetually unable to meet their energy costs once their housing costs have risen. They may choose to maintain inadequate heating and lighting in their homes, or forgo other necessities such as food or transport to provide thermal comfort and in order to minimise the shortfall (Middlemiss and Gillard, 2015). Meanwhile, the higher earning groups have seen this gap close as their salaries increase above the increase in housing cost.

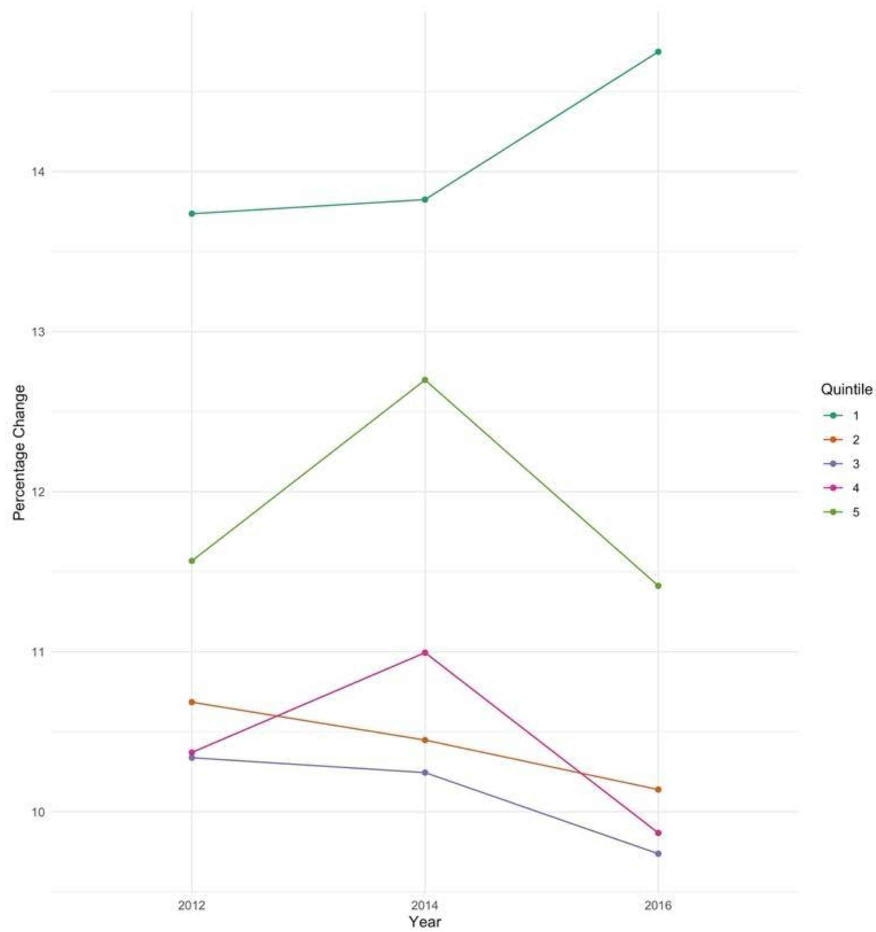


FIGURE 5-3 POSTCODE SECTOR WEIGHTED INCOME QUINTILES SHOWING PERCENTAGE CHANGE IN INCOME AFTER HOUSING COST 2012 - 2016

Other groups have seen their housing cost relative to their income reduce or remain steady, meaning they are less likely to find themselves at risk of fuel poverty. They are also more likely to have the ability to balance any increase as part of their usual outgoings given that they have seen a relative increase in disposable income, and they may have the ability to save some of their income to deal with unexpected expenses to avoid becoming fuel poor. Their rise in disposable income may also enable them to afford to make upgrades to their houses and appliances to decrease their energy costs further.

5.2.3 Energy Performance Certificates

The practicalities of Energy Performance Certificates were discussed in Chapter 3 (Section 3.3); EPCs are a mandatory certificate detailing a building's energy efficiency, predicted running costs and scope for improvement. When they were first introduced, they were intended to make homes more comparable to prospective buyers and encourage sellers to make energy efficiency improvements to increase the attractiveness of a home (Energy Saving Trust, 2020). The utility of the certificates in this situation has been debated (Section 3.3.2), but because they are now compulsory, the coverage across England and Wales is very high and covers domestic purchases and rentals since 2008. The data details a multitude of structural characteristics which provide the basis for the Energy User Classification. Given the Government's energy efficiency focussed targets which are based on EPC ratings, with regard to both fuel poverty and carbon emissions, dissecting the household characteristics which lead to properties being categorised as inefficient will lead to a broader view on energy efficiency and thus the facets of fuel poverty above and beyond a household's income.

Chapter 3 discussed the bias introduced into the dataset through collection techniques and discrepancies in the quality of assessors. The following section looks at the high level distributions within the data to understand if there are other factors which may bias the dataset. Table 5-1 overleaf provides descriptive statistics for the headline variables.

TABLE 5-1 EPC DESCRIPTIVE STATISTICS

Variable	Category	Count	Percentage
Current Energy Rating	A	19,838	0.13
	B	1,393,618	8.91
	C	4,155,004	26.59
	D	6,136,780	39.28
	E	2,874,748	18.4
	F	803,690	5.14
	G	239,614	1.53
	Invalid	244	0.002
Built Form	Detached	3,607,949	23.09
	Semi Detached	4,513,831	28.89
	Terrace	6,680,032	42.76
	Other	821,724	5.26

When compared to the housing stock data from the English housing survey (Table 5-2), terraced houses are overrepresented in the EPC dataset (only 28.4% of the housing stock is terraced, compared to 42% of EPC certificates). This is likely to be because of the affordability of terraced houses and the increase in the share of the UK housing stock which is privately rented; they appeal to first time buyers and younger residents looking to get on the property ladder, as well as to private landlords for rental properties, as so are more likely than larger family homes to have been bought or rented since the introduction of EPCs – on average owner occupiers stay in their homes for 17.8 years and so many of these properties will never have required an EPC (MHCLG, 2019).

TABLE 5-2 EXTRACT FROM THE ENGLISH HOUSING SURVEY (MHCLG, 2019)

Dwelling Type	Owner occupied	Private rented	Local authority	Housing association	All dwellings
small terraced house	6.8	16.9	10.4	12.1	**9.7**
medium/large terraced house	19.2	19.3	14.2	17.3	**18.7**
semi-detached house	30.8	15.9	16.9	15.9	**25.3**
detached house	24.7	6.1	0.2	0.8	**16.7**
bungalow	10.0	4.4	11.1	9.3	**8.9**
converted flat	1.7	11.1	2.3	4.1	**3.9**
purpose built flat, low rise	6.1	23.5	38.2	37.7	**15.1**
purpose built flat, high rise	0.7	2.8	6.7	2.8	**1.7**

The other discrepancy between the EPC dataset and the results of the English Housing Survey (EHS) is the increased propensity of very energy efficient buildings being included in the EPC data, especially when property types such as purpose built flats and apartment blocks are considered; Figure 5-4 details the cross tabulation between the building type and assigned energy rating, and is reasonably comparable to the EHS.

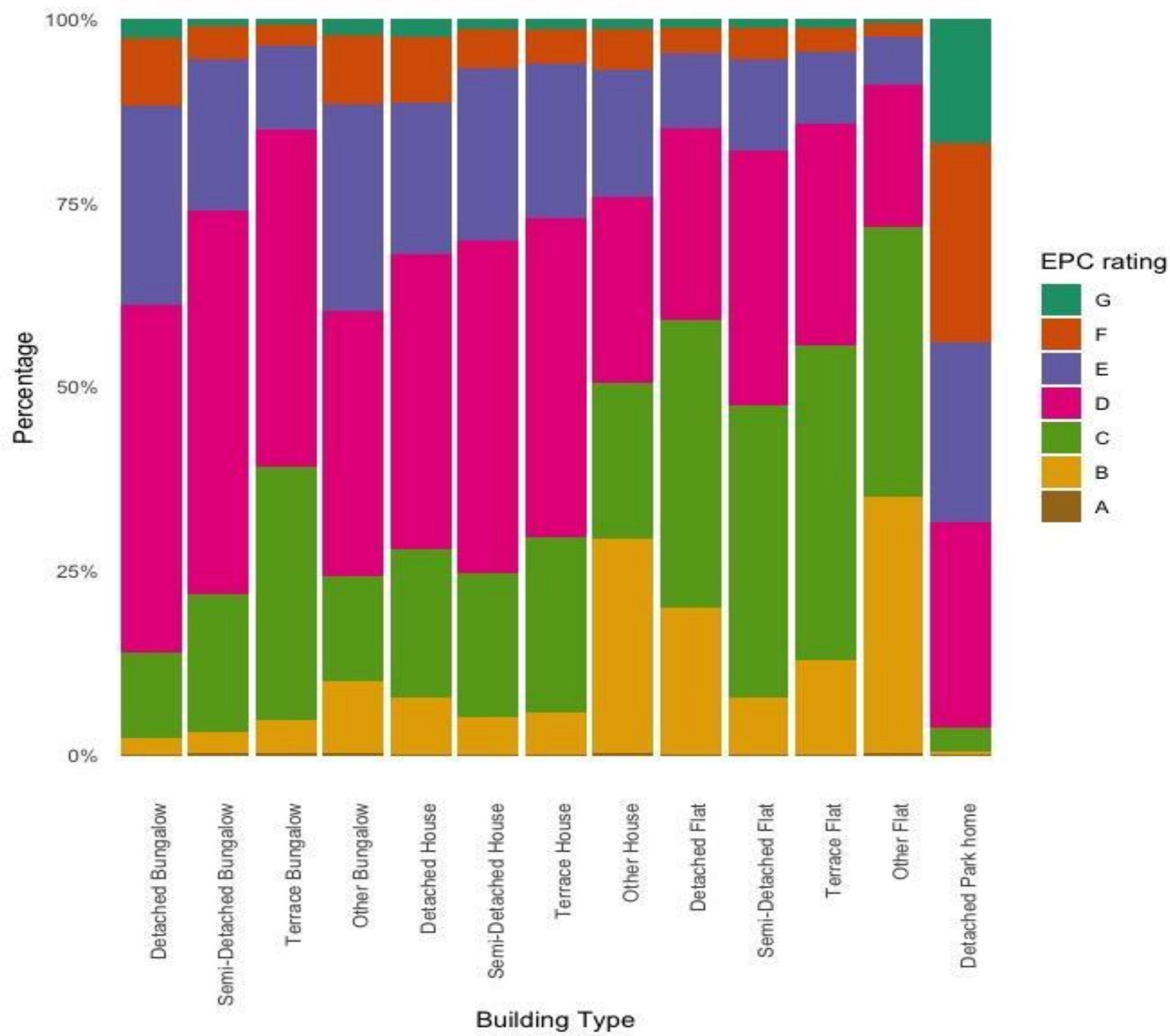


FIGURE 5-4 EPC RATING BY BUILDING TYPE

However, the EHS suggests that half of all properties are rated Band D – the EPC dataset contains all new build properties since the introduction of zero-carbon legislation (Section 2.1.4), reflected in the higher percentages of energy efficient buildings rated B and C, especially in the "Other Flat" category, which includes all purpose built apartment blocks. Looking at Table 5-2 and Figure 5-4 in conjunction helps to validate the relatively high efficiency levels of the "Terrace Bungalow" in the EPC data, which in the English Housing Survey may be classified as a 'bungalow' or a 'purpose built flat, low rise'; both of which are predominantly social rented properties, which is the highest performing tenure type with between 20% and 29% of housing association and local authority tenants being rated A-C (Hope and Booth). The distribution of Park Home efficiencies is affected by their small sample size – (2974/15,623,536).

5.2.4 Output Area Classification

To investigate the aggregated socio-spatial structure of energy efficiency, the Output Area Classification (see Section 3.4.6) was appended to the EPC data and the mean energy efficiency score per OAC group was calculated. The Energy Efficiency Score is a linear scale between 1-100 and is also referred to as the SAP (Standard Assessment Procedure) score, with the rating relating to EPC Bands – Table 5-3 details the splits.

TABLE 5-3 EPC RATING BANDS AND SAP SCORE REFERENCE (EDF ENERGY, 2020)

Band	SAP Points
A	92-100 (Most efficient)
B	81-91
C	69-80
D	55-68
E	39-54
F	21-38
G	1-20 (Least efficient)

Despite the relatively small range in energy efficiency ratings between the highest (71.04) and lowest group (51.46), there are clear disparities between the groups. Figure 5-5 shows the average SAP score for each OAC group.

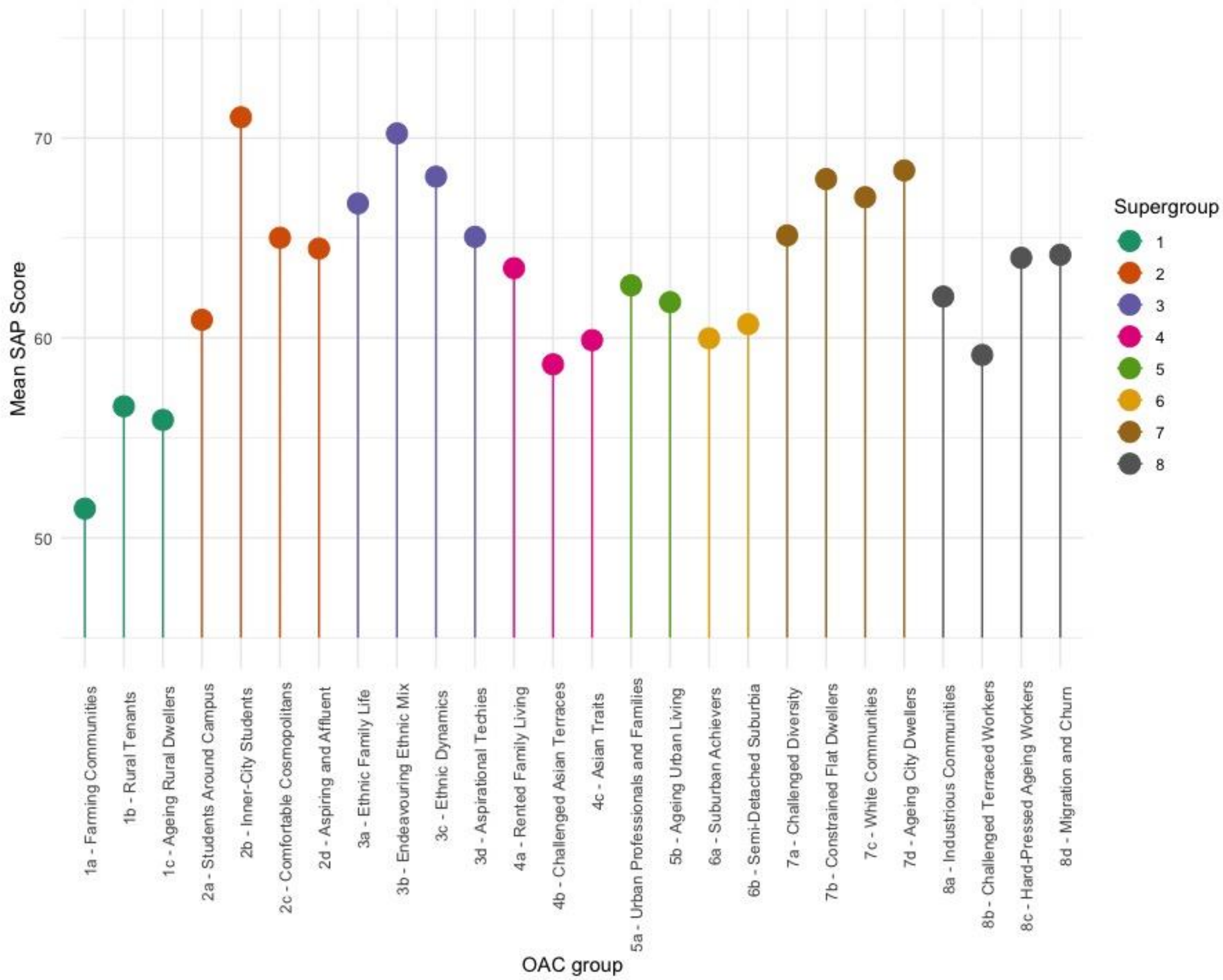


FIGURE 5-5 AVERAGE SAP SCORE BY OAC GROUP

The three groups within Supergroup 1; 1a - Farming Communities, 1b - Rural Tenants and 1c - Ageing Rural Dwellers pertain to the lowest energy efficiency ratings between 51.46 and 56.58. We know from the EPC dataset that these groups have the highest propensity to use non-standard fuel types such as wood, coal or oil as their main fuel source due to being disconnected from the mains gas network; Supergroup 1 has an average of 49.3% not connected, compared to the average in the EHS at 14% (MHCLG, 2019). The highest rated group are 2b - Inner City Students with a score of 71.04, likely due to them living in very modern and efficient, newly built halls of residence style accommodation (94% of Inner City Students lived in flats in the EPC dataset). See Chapter 2 (Section 2.1.4) for an overview of the regulations introduced to improve the quality and efficiency of new buildings. The groups characterised by living in terraced accommodation such as 4b - Challenged Asian Terraces and 8b - Challenged Terrace Workers typically score lower than others after the rural dwellers, likely due to the poor quality housing stock symptomatic of the UK, but also because of, as previously mentioned their inability to seek out and afford energy improvements.

5.3 Building a Classification of Energy Consumption

As introduced at the beginning of this chapter and in the literature (Chapter 2, Section 2.2.8), attributes pertaining to age; population density and occupation as well as physical attributes such as accommodation type, building type and tenure would likely assist in building a typology by introducing measures of demographic characteristics to supplement consumption data, and as such, create a broader view of energy efficiency and fuel poverty. The drivers of fuel poverty are highly multidimensional and often interlinked and can be summarised by three taxonomical elements :

- Demographic and contextual attributes
- Structural and physical fixtures and fittings of households
- Access to technological upgrades

The application of the fuel poverty definition to low income households only is not rational, as households above the general poverty line may be in fuel poverty. Energy poverty that is currently hidden under the current definition could be revealed through the inclusion of demographic and objective indicators such as those that fall into the categories above. It is important to include these supporting indicators and the exploratory descriptive work that has already been undertaken highlights the multidimensionality of them. The indicators are often interrelated and non-linear; in the previous chapter, patterns of access to upgraded energy efficiency technologies such as smart meters were investigated and a Morans I test revealed geographical disparities (Section 4.2.1). Profiling both smart meters and energy efficiency by the OAC (Sections 4.4 and 4.5 respectively) revealed that neither are evenly distributed across all groups. As an example, those in rural areas are particularly affected by both lack of access and poor energy efficiency, suggesting that both consumption and efficiency are influenced by multiple demographic, geographic and structural factors, as well as broadly suggesting inequity in the prioritisation of infrastructure upgrades, manifested in a high proportion both lacking access to mains gas and smart metering technology, thus forming a basis to support a conceptual framework for an energy based classification.

5.3.1 Selecting Measures

As a first step, consideration of previous findings and existing literature was required to identify those variables that would form useful inputs into the classification. Based on this 176 preliminary attributes were selected over the three taxonomical elements, derived from the EPC data and the 2011 Census. While both these datasets contain a wide variety of possible candidate variables, a large number are highly correlated or homogenous across space and so deemed less effective in classification building; for instance, any variation in sex is considered to be of lower importance since overall the ratio in small areas is the same. Furthermore, some variables in the EPC dataset were simply unsuitable and did not address the needs of the end-user such as those linked to potential efficiency and consumption. The proceeding subsections in Section 5.3 detail the steps taken to eliminate and transform variables in order that the input variables for the classification were suitable.

Due to the sparse nature of some variables at the individual certificate level and to allow for consistent comparison with results throughout this thesis, all measures from the EPC dataset have been aggregated to PCS level. Before evaluating the candidate variables, the dataset was checked for missing variables and as a result, 17 variables were removed due to being over 80% missing[8]. These pertained to measures providing very high level descriptors affecting the overall energy rating of very few households, for example “9 or more rooms”.

Contextual and demographic indicators were obtained from the census and broadly represent attributes that are known to correspond with levels of energy consumption and fuel poverty such as age, income and accommodation type. As before, the reweighting of census variables to the postcode

[8] Variable over 80% missing: current energy consumption, current lighting cost, current heating cost, current hot water cost, count of flat storeys, count of extensions, count of number of habitable rooms, headcount, 1 room, 2 rooms, 3 rooms, 4 rooms, 5 rooms, 6 rooms, 7 rooms, 8 rooms and 9 or more rooms.

sector level utilised the ONSPD and in the case of census variables, it was found that incompleteness was due to the chosen variables being associated only with Scottish census responses, meaning our study area was not covered. Table 5-4 details the standard census key statistic tables that were replicated using the download service from the UK Data Service.

TABLE 5-4 CENSUS KEY STATISTIC TABLES

Census Table Name	This table provides information on:
KS103EW	Marital and Civil Partnership Status
KS401EW	Dwelling, Household Space and Accommodation Types
KS402EW	Tenure
KS403EW	Rooms, Bedrooms and Central Heating
KS601EW	Economic Activity
KS611EW	NS-SeC

The dataset as a whole totalled 159 variables for approximately 7500 postcode sectors. It was apparent that a number of these attributes displayed skewed distributions which required further data mining steps prior to clustering, discussed in the subsequent section of this chapter.

5.3.2 Assessing Skew

Given the large number of input variables (159) a test of skewness was applied in favour of a visual inspection of a histogram for each variable to reduce the likelihood of interpretative error and to provide a quantitative measure for comparative purposes despite the large sample size. Skewness is a measure of the asymmetry of a frequency distribution. When the frequency scores are clustered at the lower end of the distribution and the tail leads towards higher or more positive scores, the data is considered right skewed and the value of the skew is positive . Conversely when the frequent scores cluster at the higher end of the distribution and the tail points towards lower values, the data is left skewed and the value of the skew is negative. If the data is normally distributed, the left and right tails are balanced and the value of the skew is zero (Doane and Seward, 2011; Field *et al.*, 2012). Ideally, all variables would display a normal distribution to ensure the optimal performance of some clustering algorithms such as K-Means, which is designed to find spherical clusters, however in practicality, this is very often not the case and as such, the skewness of the data should be understood to inform the extent of the data normalisation process.

In determining the extremity of the skew, upper and lower limits are applied as rule of thumb, and the conservativeness of the boundaries is largely subjective. In this instance it was decided that absolute values within ±2 are considered relatively normal. More stringent applications may lower this to ±1, and more conservative may increase it to ±3 (Lomax and Hahs-Vaughn, 2013).

Table 5-5 summarises the results of the skewness test, giving the number of variables which fall into each of these categories.

TABLE 5-5 SKEW DISTRIBUTION SUMMARY

Skew	**Frequency**	**Percentage**
Highly Negative	0	0
Moderate Negative	4	2.5
Approximately Symmetric	63	39.8
Moderate Positive	40	25.3
Highly Positive	51	32.2

The majority of measures that were assessed were approximately symmetrical in their distribution, but a large proportion of measures were also highly skewed in their distribution, and therefore likely to impact on cluster assignments, but it is argued that measures exhibiting skew would either be normalised using power transformations to reduce skewness or used regardless of skew, as the outliers within these measures may assist in producing distinct clusters (Singleton and Spielman, 2014). For these reasons, no variables were eliminated based on their skewness - especially as some of the measures would be expected to display skewed distributions given the domain. For example, the distribution of access to mains supply gas; indicators of infrastructure performance and prevalence would be expected to be skewed away from rural areas, as these contain smaller populations and generally have the poorest access to networked infrastructure.

5.3.3 Data Evaluation

As per the framework for designing a geodemographic classification introduced in Chapter 3 (Section 3.6), a correlation matrix was also produced in addition to an assessment of skewness. It is generally discouraged to include highly correlated measures as it can result in duplicate information where multiple measures adequately capture the same relationship (Harris *et al.*, 2005). This correlation can be addressed in one of two ways; by omitting multiple highly correlated measures to leave a single variable that is correlated with the largest number of other measures in order to ensure robustness, or alternatively, all measures can be included with or without applying weights. Weighting can be problematic as the process of selecting weights for individual measures can be argued to be subjective (Harris *et al.*, 2005).

It is possible to summarise the most notable correlations observed between sets of input measures. In general:

- Measures relating to physical properties of and within homes were strongly correlated to one another. For example, Gas Central Heating 'TRUE' was significantly related to Mains Gas Flag 'FALSE', $r = -.93$, $p < 0.05$
- Measures relating to current energy ratings were also highly correlated with physical properties of buildings. For example, there was a significant relationship between EPC Band F and Solid Fuel Central Heating, $r = .69$, $p < 0.05$

Based on the examination of the correlation matrix, six variables related to two measures were removed. These pertained to information related to Top Storey Flat; TRUE, FALSE and NA and

Solar Water Heating; TRUE FALSE and NA. The Top Storey Flat variable applies to very few records overall, and the properties of a household it describes are already recorded by a more descriptive variable which classifies all flats based on their floor level. Table 5-6 shows the significantly positively correlated pairs, with values close to 1 (FL_* standing for Floor Level, FTS_* standing for Flat Top Story).

Table 5-6 Highly Correlated Variables

Variable 1	**Variable 2**	**r**	**p**
FL_ground	FTS_N	0.53913201	< 0.05
FL_middle_floor	FTS_N	0.81021932	< 0.05
FL_unknown	FTS_NA	0.93519354	< 0.05
FL_top_floor	FTS_Y	0.95227015	< 0.05
SWHF_NA	PHOTO_NA	0.54893739	< 0.05
SWHF_N	PHOTO_FALSE	0.55329887	< 0.05
SWHF_Y	PHOTO_TRUE	0.58607444	< 0.05

Likewise, Solar Water Heating Flags (SWHFs) are covered by a variable describing the presence of photovoltaic panels - you cannot have solar water heating if photovoltaic panels aren't in place and again, SWHFs apply to so few individual level records it is fair to say that their removal is unlikely to make any significant difference to the final clusters.

Broadly speaking, other correlated variables were not unexpected and have been observed in previous literature, such as the positive relationship between demographic characteristics and tenure type (e.g. Private rented tenure and Age 25 – 29, $r = .76$, $p < 0.05$). As such, no other measures were removed as a result of this step. This decision was made on the basis that removing correlated variables based on the analysis of global statistical relationships could potentially mask local variation and lead to the smoothing of important non-linear patterns at a more granular level (Singleton and Spielman, 2014).

Table 5-7 details the final measures which were included in the dataset used to build the classification; each measure is broken down into factorised variables, detailed in the full variable table in Appendix 9.4.

TABLE 5-7 FINAL MEASURES FOR CLASSIFICATION BUILDING

Domain	Geo-locator	Energy Information	Physical Attributes	Fixtures and Fittings	Demographic
Sub-domain	Postcode sector	Current energy efficiency	Property type	Glazing	Economic Activity
		Current environmental impact	Built Form	Hot Water	NSSEC
		Current energy consumption	Mains Gas Flag	Secondary Heating	Marital Status
		Current costs	Floor level (flats only)	Central Heating	Tenure
		Number of storeys	Top Storey Flat (flats only)		Age group
		Habitable rooms	Extensions		
		Energy Certificate rating	Wind turbines		
		Transaction type	Solar Water Heating		
		Energy tariff	Photovoltaics		
		Mains fuel source	Accommodation type		
			Number of Rooms		

5.3.4 Transformation and Normalisation

The next stage was consideration of transformation and normalisation processes. Several variable normalisation methods are frequently referred to in geodemographic literature; including Box-Cox, log10 and cube-root (see Section 3.6). Given what is known about the skewness of the variables in this dataset, a thorough evaluation of the different transformation methods is imperative; highly skewed data can lead to poor cluster assignments, especially when used as inputs into clustering algorithms that are optimised to find spherical groupings such as K-means (Gale *et al.*, 2016). Box-Cox and log10 methods require values to be positive and greater than one; as some variables here had values that fell below one, a constant of 100 was applied to ensure that transformations could take effect.

Log10 transformations, compresses the upper tail and stretches out the lower tail, making the transformed data appear more normal, but apply a globally standard method of normalisation across a dataset, leading to compressed differences between large values and increasing differences between small values to artificially reduce variance. The Box-Cox method computes an appropriate exponent (lambda, λ) to transform a variable (Y) and normalise its distribution (Equation 3 details the Box-Cox equation). Multiple λ values between -5 and 5 are tested and the one that results in the most normal distribution is used for the power transformation. This means that the extent to which a variable is transformed is dependent on its level of skew.

$$\boldsymbol{Yi(\lambda) = \{Yi^{\lambda} - 1/\lambda\ (\lambda \neq 0)\ log(y)\ (\lambda = 0)}$$

EQUATION 3 BOX COX TRANSFORMATION

Finally the cube-root (x to x^(1/3)), has a strong effect on distribution shape and is most commonly applied to right tailed data. Whilst it does not have as substantial an effect on the distribution as the log transformation, it has utility in that it can be applied to zero and negative values without the need to include a constant. The three methods of skew reduction were compared, and Table 5-8 details the results for a subset of the ten most highly skewed variables.

TABLE 5-8 NORMALISATION METHOD RESULTS FOR TOP 10 MOST HIGHLY SKEWED VARIABLES

Variable	Skew (Raw)	log10	Cube-root	Box-Cox
CER_A	79.12	77.95	1.21	0.07
CER_INVALID!	94.34	94.34	30.53	7.01
HWD_gas_other	91.15	90.84	0.77	-0.14
HWD_heat_pump	66.95	66.90	9.59	7.31
HWD_none	66.42	65.30	0.11	-0.91
HWD_oil	65.40	64.00	2.96	1.55
MF_Community_scheme	75.98	75.11	27.71	15.42
SHD_hot_water_only	79.59	79.58	17.51	14.12
SHD_NA	91.39	91.12	2.34	1.06
WTC_TRUE	92.30	92.10	0.95	-0.10

It is evident that the Box-Cox method significantly improves the overall symmetricity of the variables. Whilst some are still highly skewed variables they are much closer to zero than prior to transformation and others are now more moderately skewed. Given the level of improvement it has had on the skewness of the variables the Box-Cox standardisation was applied to the dataset for the subsequent analysis.

5.3.5 Standardisation

Before clustering, the variables were standardised using z-scores to create a common scale. This was applied to both the transformed and the naturally distributed datasets. As discussed in Section 3.6 z-scores are the most common method for data standardisation and scores are calculated by subtracting the population mean from an individual raw score then dividing the difference by the population standard deviation. This results in a set of scores that are positive if they fall above the means and negative if they fall below, meaning that all standardised variables have an adjusted population mean of 0.

The other well-known method for rescaling data is a range transformation, which rescales the values into a range of 0 – 1 and is most useful in cases where all parameters need to have the same positive scale. It's major drawback, particularly in respect of clustering is that information about outliers is lost and could result in clusters without particularly distinct characteristics (Alexiou, 2016) (Section 3.6). In order to compare the utility of the two methods, the naturally distributed and transformed datasets were replicated and standardised using a range transformation; the results are discussed in the following section.

5.3.5.1 Testing the Impact of Standardisation

The first stage of the clustering process was to cluster both the transformed and naturally distributed inputs to observe the effects of the transformation, standardisation and normalisation processes on cluster assignments. At this validation stage a detailed classification is unnecessary, so the total Within Cluster Sum of Squares (WCSS) and between cluster sum of squares were calculated for k = 2:10 (k is equal to the number of unique seeds, see Section 3.6) in order to ascertain which combinations are likely to produce the most heterogeneous clusters.

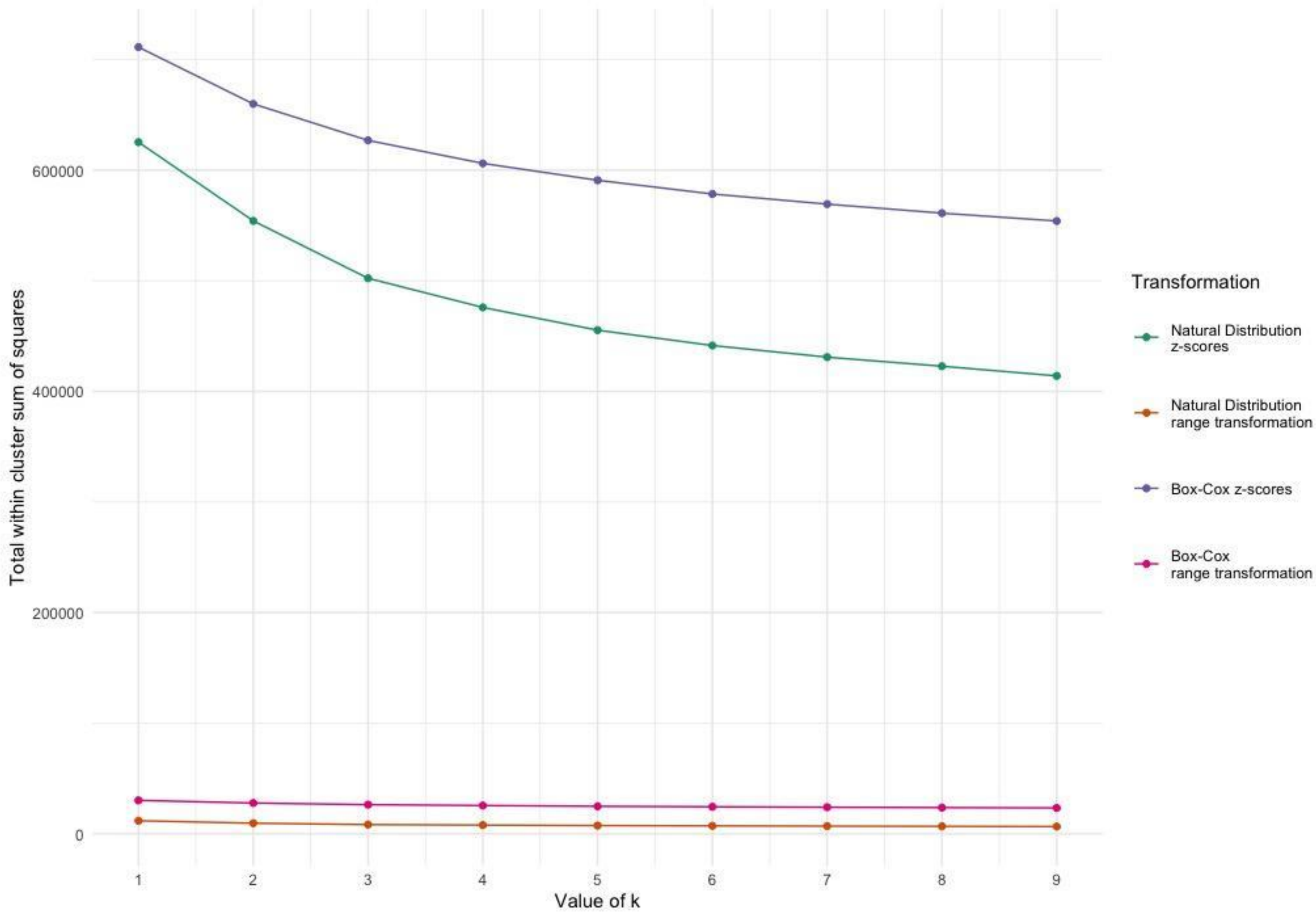


FIGURE 5-6 TRANSFORMATION EFFECTS ON WCSS FOR K = 2:10

It is clear to see from Figure 5-6 that the range transformed data for both the natural and Box-Cox datasets produce clusters with very low Total Within Cluster Sum of Squares (WCSS). Ordinarily it would be desirable to see a low WCSS score as a measure of 'goodness' of the clusters, however, if the WCSS is low for all values of k with no discernible "elbow" point where the increase in clusters no longer results in an improvement in WCSS this may suggest that cases within each cluster are too similar and would produce clusters with hard to distinguish heterogenous characteristics. This result is likely due to the nature of range transformations bounding the data and losing distinctive characteristics that appear as outliers in the data.

Figure 5-6 also shows how both the naturally distributed data and Box-Cox transformed data performed when z-scores are applied. Both show decreasing WCSS scores as the value of k increases, but the non-transformed data produced consistently lower scores, suggesting that it is the most likely to produce easily interpretable clusters with heterogenous characteristics. As a result of this evaluation, the range transformed data was discounted for further use.

To investigate the heterogeneity of the two z-score standardised datasets, the cluster assignments from an arbitrary run of k = 5 clusters was outputted and investigated. The initial output summary tables as shown in Table 5-9, and Table 5-10 and Table 5-11 overleaf, revealed differences in terms of cluster sizes, aggregate characteristics, and ease of interpretability. These summaries are a subjective interpretation of the outputted cluster centroids and have been condensed such that they can be presented overleaf. The full table used for the evaluation is included in Appendix 9.5.

TABLE 5-9 CLUSTER SIZES

Cluster	n (Natural)	n (Box-Cox)
1	880	1361
2	1121	1208
3	2201	2417
4	2616	922
5	843	1753

The naturally distributed data produced the most interpretable assignments, with more homogenous clusters formed. Based on the representation it is apparent that in the naturally distributed cluster assignments display the following characteristics:

- Cluster 1 is largely students and young professionals renting modern, efficient flats from private landlords.

- Cluster 2 represents mostly young, unemployed families in social rented properties of middling efficiency.
- Cluster 3 are middle aged or pensioners, living in homes of middling efficiency.
- Cluster 4 is a mix of ages representing both families with children and retirees, living in mortgaged homes with middling efficiency.
- Cluster 5 is middle aged and retired households, who live in larger detached properties which they own and are inefficient to heat.

In both cases, the tables highlight the variables where all the options were fairly equally represented in the clusters; Table 5-11 shows that this occurred much more frequently in the Box-Cox transformed dataset.

TABLE 5-10 NATURAL DISTRIBUTION CLUSTER CHARACTERISTICS

Cluster	Certificate A-G	Property Type	Built Form	Energy Tariff	Mains Gas	Glazing	Main Fuel Type	Economic Activity	Marital Status	Tenure	Age
1	B/C	Flats	Terrace/ Other	Unknown	**Mixed**	**Mixed**	Electric	Students	Single/Civil	Private Rent	18 - 44
2	C/D	Houses	Terrace	Single	TRUE	Double	Gas	Unemployed	Single/ Separated	Social Rent	0 - 44
3	D/E	Houses/ Bungalows	Detached/ Semi	**Mixed**	TRUE	Double	Gas	Part-time/ Retired	Married	Mortgage/ Owned	45 - 90+
4	C/D	Houses	Semi	**Mixed**	TRUE	Double	Gas	Full/Part-time	**Mixed**	Mortgage	**Mixed**
5	E/F/G	Houses/ Bungalows	Detached	Dual Fuel	FALSE	**Mixed**	Non-Standard	Self Employed/ Retired	Married	Owned	45 - 74

TABLE 5-11 BOX COX DISTRIBUTION CLUSTER CHARACTERISTICS

Clusters	Certificate A-G	Property Type	Built Form	Energy Tariff	Mains Gas	Glazing	Main Fuel Type	Economic Activity	Marital Status	Tenure	Age
1	**Mixed**	Bungalows	Detached	Dual Fuel/ Off Peak	FALSE	Triple/ Secondary	Non-Standard	Self Employed/ Retired	Married	Owned	45 - 74
2	**Mixed**	**Mixed**	Terrace	Single	TRUE	Double	Gas	Unemployed/ Carer	Single/ Separated	Social Rent	0 - 9/ 20 - 29
3	**Mixed**	**Mixed**	**Mixed**	**Mixed**	TRUE	**Mixed**	Gas	**Mixed**	**Mixed**	**Mixed**	**Mixed**
4	C	Flats	Terrace	Unknown	**Mixed**	Single/ Secondary	**Mixed**	Students	Single/Civil	Private Rent	20 - 29
5	**Mixed**	Bungalows	Detached/ Semi	**Mixed**	TRUE	Secondary	**Mixed**	Self Employed/ Retired	Married	Mortgage/ Owned	45 - 90 +

5.3.6 Construction and Hierarchical Design

Clustergrams were utilised to select the optimum number of clusters (Section 3.6) Plotting the distribution and redistribution of cluster centroids between outputs for a range of potential k values aids the interpretation of an optimum value of k, by being less subjective and error prone than other methods such as elbow plots and gap statistics. As seen in Figure 5-7, the Clustergram tested values of k from 2 through 10. For each iteration the method works by multiplying the cluster centres by the first loading of the principal components of the original data, thus offering a weighted mean of each cluster's centre dimensions, as indicated by the red point.

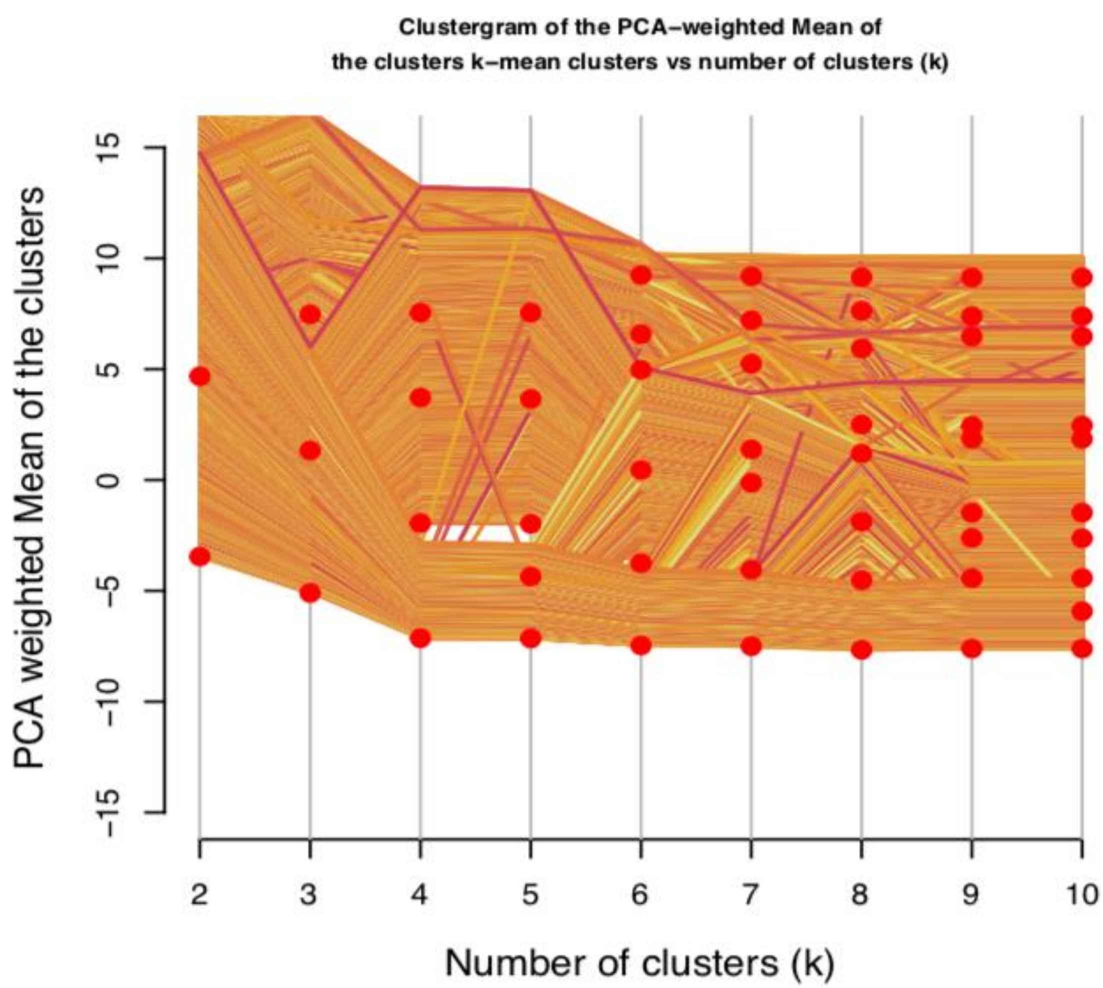


FIGURE 5-7 CLUSTERGRAM ITERATIONS FOR K = 2:10

Figure 5-7 visualises the resulting clustergram for a K-Means clustering algorithm. From this, it is apparent that when k = 2, the clusters are well spaced, as indicated by the red points, which visualise the cluster centre. The spacing suggests that the two clusters are sufficiently homogeneous in terms of their characteristics that they are easily distinguishable from one another. As the number of k is increased to 3, it is possible to track the reassignment of observations. In this case, a number of observations from each cluster are reassigned to form a central cluster or reassigned to each other. The same principle applies with k = 4 and 5, the clusters remain well spaced, but as the value of k increases, cluster centres become much closer together, and some overlap appears in the highest values, which may impact the interpretability of each cluster's characteristics. This suggests the optimum value for k has been exceeded.

Following this evaluation, a K-Means clustering algorithm was undertaken for k values of 4 and 5, as they presented the optimal values of k in the clustergram. Before a full examination of the clusters was undertaken, the principle components for each observation in the dataset were visualised to compare, as can be seen in Figure 5-8 and Figure 5-9.

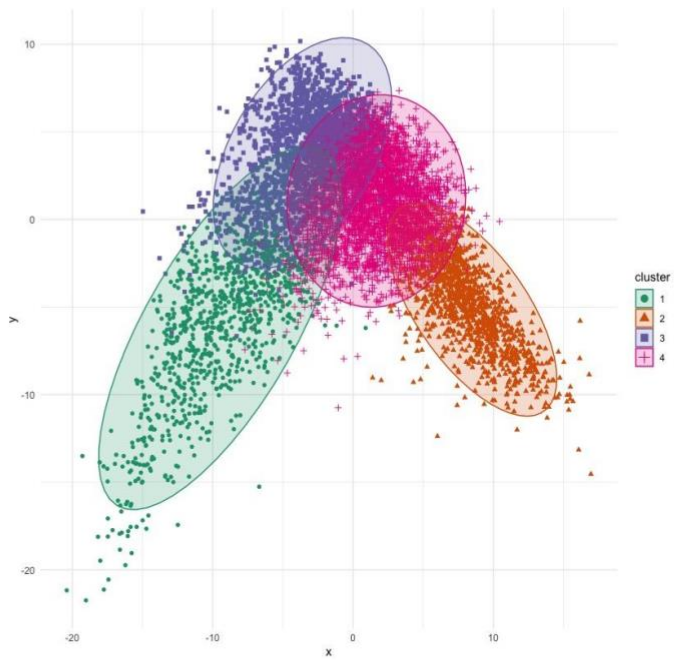


FIGURE 5-8 PRINCIPLE COMPONENT ANALYSIS FOR K = 4

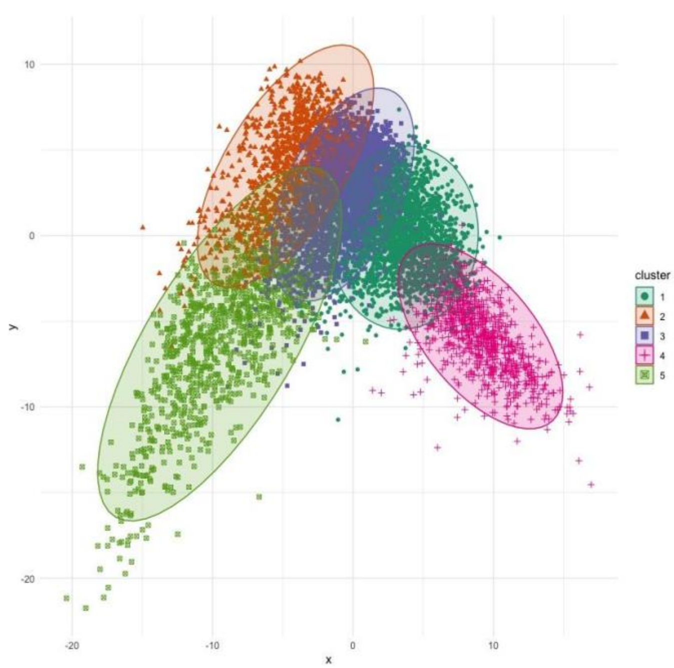


FIGURE 5-9 PRINCIPAL COMPONENT ANALYSIS FOR K= 5

Figure 5-9 shows that at k = 5 the clusters have a significant amount of overlap, with cluster 3 being almost completely covered by clusters 1, 2 and 5, with very few observations displaying distinctive characteristics different to any other cluster. This is likely to make uncovering the heterogenous characteristics of each cluster difficult. Whilst there is still some overlap when k = 4 (Figure 5-8), there are also still plenty of observations for each cluster which do not overlap, suggesting that of the two, this cluster assignment appears the most likely to provide easily interpretable and homogeneous clusters. Each of the four initial clusters were then separated and re-clustered in an attempt to build a second tier within the classification to give a more granular final typology. However, on investigation it was apparent that most clusters were unable to support results significantly different from the parent cluster and so the second tier was not investigated further. Following this final categorisation, the resulting classification was a single tier typology containing four clusters. For consistency with similar literature on geodemographics and to futureproof this classification, these top level clusters are referred to as 'Supergroups' going forward. The next stage was to study the characteristics of each and translate this information into a set of descriptive summaries.

5.4 Results

The process of summarising the Supergroup characteristics was achieved using a number of methods. A summary table of the cluster centroids was produced and visually inspected; interpretation of the results was aided by conditional formatting and can be reviewed in appendix 9.5. From this table, key characteristics of each Supergroup were extracted and recorded, providing the basis for the resulting textual summaries ('Pen Portraits'). Secondly the clusters were mapped to reveal their geographic distributions and the areas categorised using the urban rural classification to quantify this visual inspection. Through this combination of information, names and pen portraits were created for each of the four Supergroups.

5.4.1 Cold and Costly

Age	0 - 44
Marital Status	Single or Separated
Economic Activity	Unemployed / Carer
NS-SeC	5 and up
Tenure	Social Rent
Property Type	House
Accommodation	Terrace
Certificate	C/D
Energy Tariff	Single
Mains Gas	True
Glazing	Double
Hot Water	Mains/Gas Boiler
Main Fuel Type	Gas

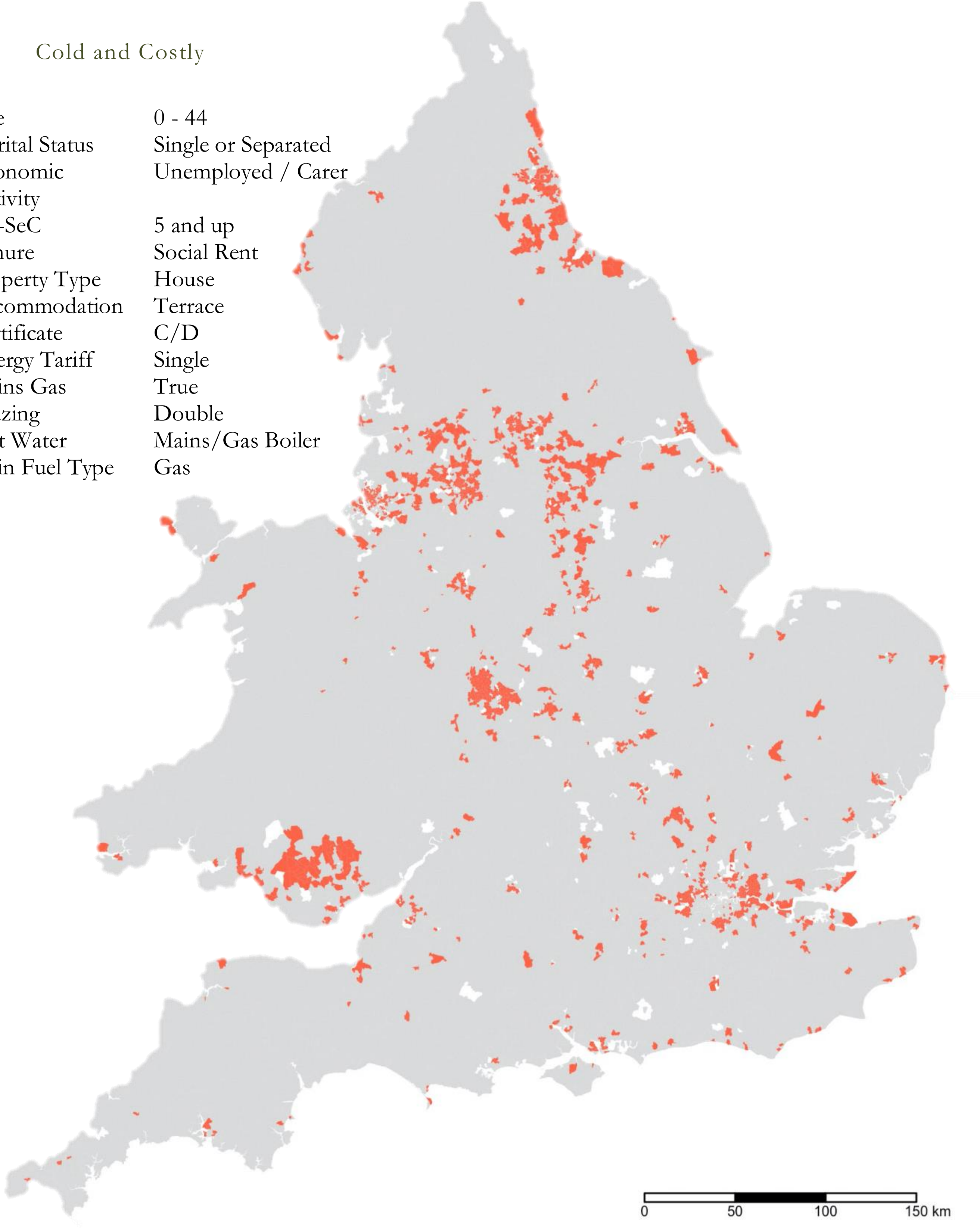


FIGURE 5-10 COLD AND COSTLY KEY CHARACTERISTICS AND GEOGRAPHY

Figure 5-10 shows the national distribution and key characteristics of the first Supergroup in the classification. It is apparent that this Supergroup is concentrated around major urban and suburban areas and towns, which typically attract younger populations and families.

This Supergroup is characterised by an increased likelihood of constituent homes being underheated or costly to heat; a high proportion use the more expensive single energy tariffs which are associated with pre-payment meters and despite the fact that there is some evidence of minor structural improvements being made, such as high levels of double glazing and EPCs generated from upgrade projects, the majority of homes are still more likely to be Band C or below. Members of this Supergroup are most likely to be housed in socially rented accommodation and therefore suffer from the tenant/landlord dichotomy, giving them little to no autonomy over the cost of their consumption. They are typically terraced or semi-detached houses, occupied by families with children, as indicated by the age range. The adults in this Supergroup are most likely to be long term unemployed, disabled or working in routine and semi-routine occupations. The above factors all combine to make this group the most likely to struggle to heat their homes consistently without becoming fuel poor, especially during colder months when they have been unable to build up any credit with their energy supplier to cover the increased usage. Comprising of 1921 postcode sectors (25%) and 25% of the population, it is the 2nd biggest cluster. On the basis of these features, the Supergroup name "Cold and Costly" was ascribed.

5.4.2 Off Grid Owners

Age	45 +
Marital Status	Married
Economic Activity	Part Time / Self Employed/Retired
NS-Sec	4
Tenure	Owned outright
Property Type	House/Bungalow
Accommodation	Detached
Certificate	E:G
Energy Tariff	Dual
Mains Gas	False
Glazing	Secondary/Triple
Hot Water	Immersion
Main Fuel Type	Non-standard

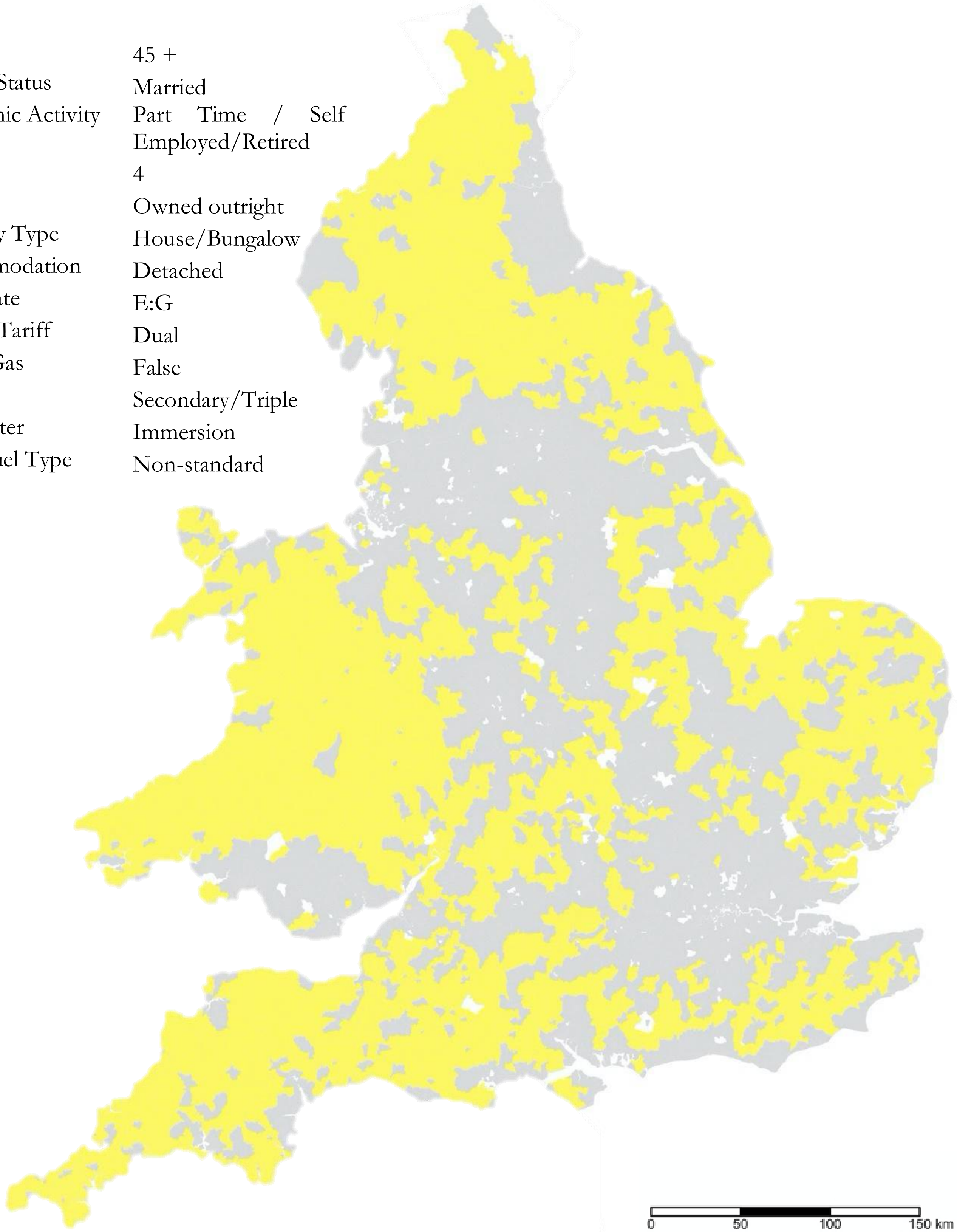


FIGURE 5-11 OFF GRID OWNERS KEY CHARACTERISTICS AND GEOGRAPHY

Figure 5-11 shows the national distribution and key characteristics of the second Supergroup within the classification. Unlike the first, it is clear that this Supergroup is widespread nationally, covering most deep rural and rural fringe areas. This Supergroup does not cluster around urban areas and towns; those areas generally associated with younger populations.

This Supergroup is characterised by aspects of demography and location, as well as some variations in energy efficiency characteristics. They are most likely to be aged over 45 or elderly, and married home-owners. They typically live in detached houses with a large floor area, and are the most likely to be self-employed, or semi/fully retired. There is evidence of them exercising autonomy over their energy consumption and its costs by making long term investments in their properties. The data shows that they have the highest propensity of undertaking energy upgrade assessments as well as undertaking general and efficiency based home improvements such as constructing extensions, installing solar panels and installing double or triple glazing. This leads to this Supergroup having the highest proportion of the four clusters with A rated properties. However, rurality clearly limits choice and opportunity to engage in more efficient energy consumption - those living in very remote areas rely on inefficient and expensive non-standard fuel sources such as oil or wood and are less likely to be connected to the mains gas network. These large, rural homes are likely to be inefficient and under occupied and as such, costly to heat to a comfortable level. It is these who may find themselves with increased costs during colder months; those on standard energy meters are more likely to have some credit with their fuel supplier as they have dual fuel tariffs. The "Off Grid Owners" Supergroup is made up of 1066 postcode sectors (14%) and accounts for 10% of the population.

5.4.3 Efficient City Living

Age	18-44
Marital Status	Single or Cohabiting
Economic Activity	Student
NS-Sec	1, 2 or 8
Tenure	Rented (Social and Private)
Property Type	Flats/Terrace Houses
Accommodation	High proportion shared accommodation
Certificate	B/C
Energy Tariff	Unknown
Mains Gas	Mixed
Glazing	Mixed
Hot Water	Immersion
Main Fuel Type	Electric

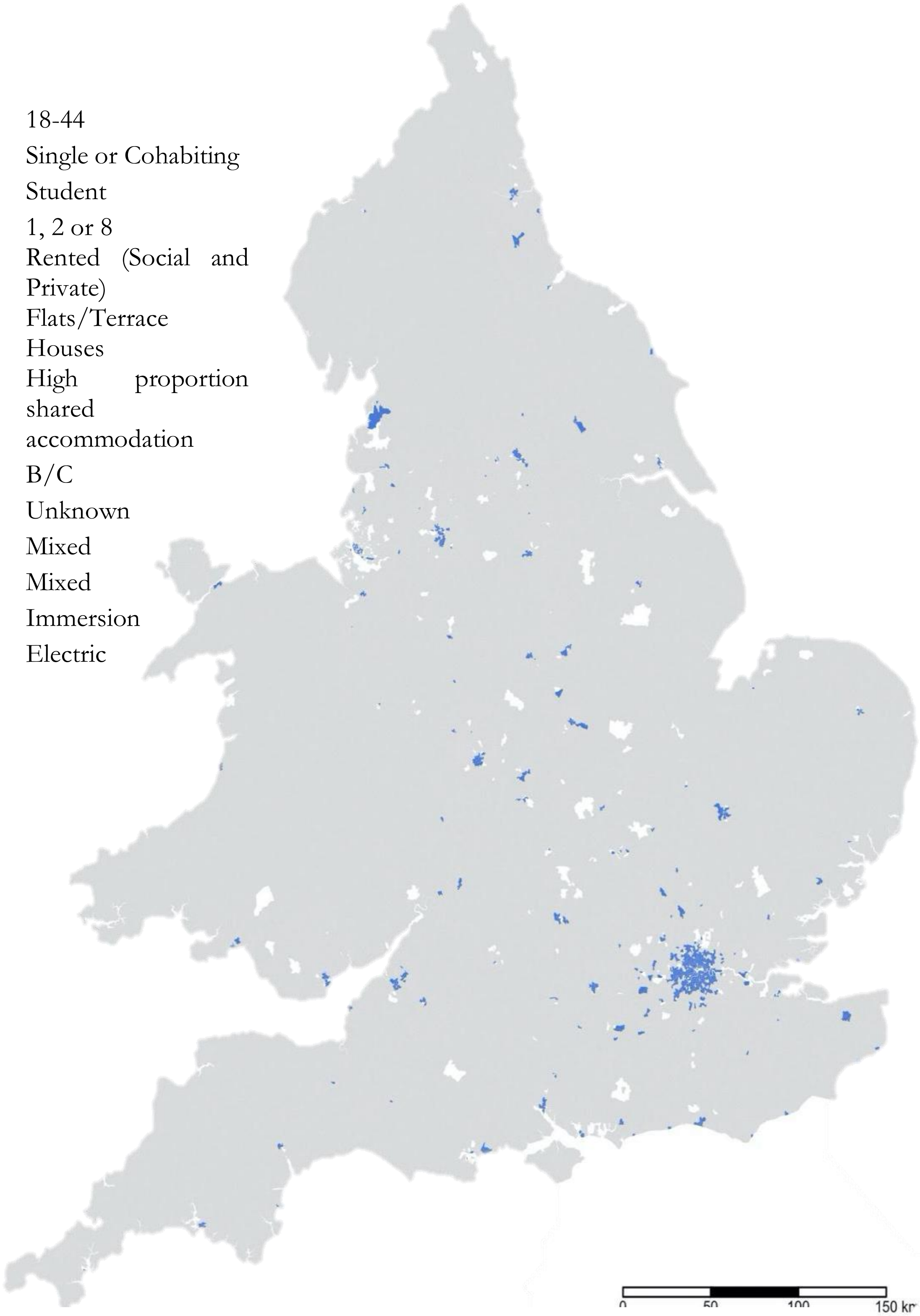


FIGURE 5-12 EFFICIENT CITY LIVING KEY CHARACTERISTICS AND GEOGRAPHY

Figure 5-12 shows the national distribution and key characteristics of the third Supergroup in the classification. This Supergroup has a tendency to cluster around major urban areas; predominantly London, but cities such as Manchester, Liverpool, Birmingham and Bristol are also highlighted. This Supergroup comprises populations who are young, students or working in higher managerial and professional occupations, and are living in major cities across the country. They typically live in privately rented flats or houses, which are newly built or purposely converted with updated, efficient fixtures and fittings such as triple glazing and modern boilers. This coupled with the fact that they typically have a lower square footage to heat and light means this Supergroup are the least likely to find themselves with high energy bills they are unable to alter. Their relatively high income allows them to absorb shocks to their bills and so even in the cases where expensive fittings such as immersion heaters are found, this group are the least likely to find themselves in fuel poverty. Some of this cluster are students living in purpose built halls of residence, whose all-inclusive living arrangement means that whilst they have no autonomy over their energy efficiency, they also do not need to consider energy bills as an extra cost and so, will not find themselves in fuel poverty. They are the smallest group, made up of 912 postcode sectors (12%), and are 10% of the population. This group was labelled as "Efficient City Living".

5.4.4 Typical Tariff

Age	45+
Marital Status	Married
Economic Activity	Employed
NS-Sec	1,2 & 3
Tenure	Mortgaged/Owned
Property Type	House/Bungalow
Accommodation	Detached/Semi
Certificate	D/E
Energy Tariff	Single
Mains Gas	True
Glazing	Double
Hot Water	Mains/Boiler
Main Fuel Type	Gas

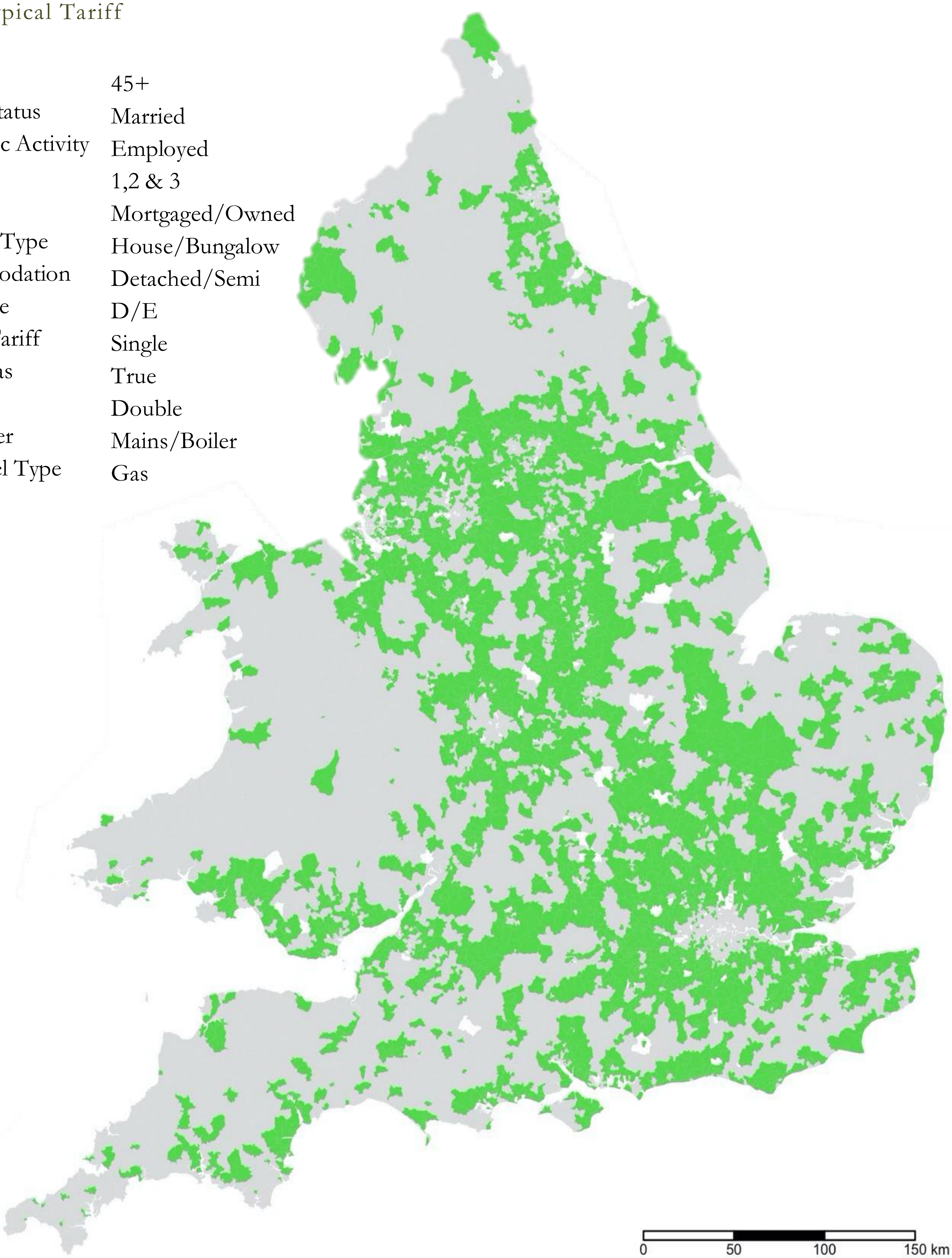


FIGURE 5-13 TYPICAL TARIFF KEY CHARACTERISTICS AND GEOGRAPHY

Figure 5-13 shows the national distribution and key characteristics of the fourth Supergroup in the classification. This Supergroup is distributed between towns and urban areas, but rural areas are also present. This Supergroup displays characteristics that are closest to the overall average. Areas are characterised by mixed energy efficiency and average floor area, and display a variety of fixtures, fittings and physical property attributes. Homes are typically semi-detached and are mostly mortgaged. There is a higher proportion of elderly people but an overall mix of ages and family types. Members of this Supergroup who are of working age are typically in middle or lower supervisory jobs. There are fewer shared houses and private rentals than other clusters. It is the largest cluster, accounting for 49% of postcode sectors (3762) and 52% of the population. This Supergroup was ascribed the name "Typical Tariff".

5.5 Conclusions

This chapter has shown that both the 10 percent and Low Income High Cost definitions of fuel poverty are lacking in terms of the lived experience of fuel poverty and it has re-evaluated the ways this multifaceted issue should be considered outside of its technical and structural policy framework. By integrating energy performance characteristics and demographic indicators, the EUC underpins the utility of consumer data as an asset to social science research endeavours and in this case in particular, the identification of at risk populations.

The results of the longitudinal small area income analysis showed the greatest levels of instability for the lowest income earners regarding their housing costs, which has numerous repercussions; it is harder for them firstly to plan for these changes and increases, but also harder for them to recover from financial shocks as they are the least likely to be able to save any of their disposable income in order to absorb them. This has implications for these households and leads to debt and restricts energy payment options which are available to them, often leaving only the most expensive pre-payment methods. As a result of the pre-payment tariffs, they are more likely to find themselves in short term fuel poverty in the winter as their costs change and may find themselves in perpetual fuel poverty if the gap between before and after housing cost continues to grow. All other income quintiles have seen the gap between before and after housing cost begin to close, giving them greater stability and more disposable income relative to their housing cost. When the cluster income spread is considered, it is clear that the lowest earners are again the most vulnerable to instability, although cluster 3 as the highest earners spend a significantly higher percentage of their income on housing costs.

The variations in energy efficiency rating between the OAC groups suggested that energy efficiency cannot be solely linked to the structural and physical aspects of a household. There were clear differences in the energy ratings of varying demographics and whilst building type and structural properties affecting energy efficiency were present, other factors such as rurality, ethnicity, employment and age were also at play.

By combining the EPC, census and income data, the resulting EUC validates the multidimensionality of fuel poverty by detailing the demographic characteristics present in areas which are overlooked by the current fuel poverty definition. It also shows that in each cluster of the EUC, there are areas which are currently considered fuel poor, again suggesting that the definition does not encompass all facets of fuel poverty. It clearly highlights one cluster which are the most at risk of fuel poverty under a multidimensional lense; the 'Cold and Costly' supergroup exhibit a large number of demographic, environmental and consumption characteristics that could be considered systemic factors of fuel poverty. Furthermore, it highlights the importance of both successful data linkage from multiple data sources in enabling a more detailed representation of the populace than has previously been possible from traditional data sources as well as highlighting the need for tools such as itself for enabling cohesive data partnerships between public and private entities for the effective enactment of fuel policy directives.

Overall it is true to say that energy efficiency and income do both have an role in a household's ability to consume energy, but they are not the only factors and under the current political framing of fuel poverty as a monetary problem within a disjointed policy framework which only aims to alleviate rather than eliminate fuel poverty, there are many other characteristics that are overlooked and left unresolved. The EUC provides utility in defining this multidimensionality and has implications for the policies and solutions of targeted fuel poverty alleviation. The chapter succeeding this endeavours to validate this classification through internal and external measures.

6 Evaluating the Energy User Classification's utility: Suggesting areas of improvement for the DEP to achieve the greatest social good.

6.1 Introduction

Previous chapters in this thesis have focused on the key characteristics and drivers of fuel poverty (See Chapter 2), their geography and how smart meter adoption rates of areas is differentiated by socio-spatial measures (Chapter 4). In addition, overall and temporal energy usage characteristics captured through smart meter technology were examined in Chapter 4. The result of the analyses in both Chapters 4 and 5 have suggested that there are geographic and socio-economic disparities that when collated, engender a complex geography of fuel poverty that far surpasses current definitions, whilst also indicating that the geographies of smart meters and access to and engagement with the technologies is worthy of further investigation. The Energy User Classification (EUC) in Chapter 5 utilises the EPC data and small area statistics to successfully generate a classification which shows that fuel poverty is associated to particular demographic characteristics which are not necessarily expenditure based, such as accommodation type, tenure and family life stage. It is clear from these results that the current definitions overlook the behavioural lived experience of fuel poverty.

Whilst the research presented in each of the empirical studies has made use of relevant and innovative sources of data, the outcomes currently remain independent of one another. To provide depth to these insights, there is an opportunity to utilize the measures and insights generated thus far within a local case study. This both provides additional validation, but also useful insight on the case study area; and to make recommendations of greatest social impact for Domestic Energy Providers in fulfilling their smart meter installation obligations. Chapters 2 and 4 both touched upon the multitude of constraints faced by both suppliers and customers during the smart meter roll-out process and this chapter also aims to address these in a constructive way and offer practical recommendations for overcoming those which are linked to the demographic characteristics of users.

The rest of the chapter is structured as follows; the Energy User Classification (EUC) is dissected; internal validation of the clusters interrogates the fit statistics, particularly in regard to those areas where the fit is poorest. External validation and correspondence to external indicators of deprivation and fuel poverty proves its utility in outlining the more nuanced definition of fuel poverty, which takes a much more multifaceted approach than the current definition, showing that demographic characteristics of fuel poverty are present at all income levels. Finally a practical validation of the cluster analysis in a targeted application of Wolverhampton is done to investigate smart meter adoption rates (SMAR) in the EUCs most "at risk" Supergroup with regards to fuel poverty, to discuss the environmental constraints faced by those households and provides recommendation to the Domestic Energy Provider with regards to overcoming them in order to have a positive social impact through their smart meter rollout schemes.

6.2 Cluster Fit and Outliers

In order to validate the Supergroups (clusters) in the Energy User Classification, a fit statistic was calculated to reveal how well each Postcode Sector is represented by its assigned cluster. When referring to clustering algorithms, similarity and distance are analogous and fundamental concepts and many of the measures of similarity used in cluster validation are comparable with Euclidian distance - that is - the greater the distance the more dissimilar the observations are. However, it is important to consider that the similarity amongst observations is more complicated in highly dimensional datasets with many variables and observations might be similar in some characteristics, but dissimilar in others, invariably skewing the fit statistics (Brunsdon and Singleton, 2015).

Within this context it was important to explore how well the cluster assignments fit the underlying distribution of calculated measures for each area; and where outliers were shown, what useful insights did these reveal and could they be explained through a more detailed investigation of the area characteristics. As such, the distances from the cluster mean were calculated for each Postcode Sector (PCS) and interpreted. As discussed above, it was noted that some variables were having a disproportionate effect on the overall fit of some PCSs. Figure 6-1 displays the spread of the distances for each of the four clusters, and suggests that particularly in cluster 2, 3 and 4, some PCSs displayed very high distances from the cluster mean.

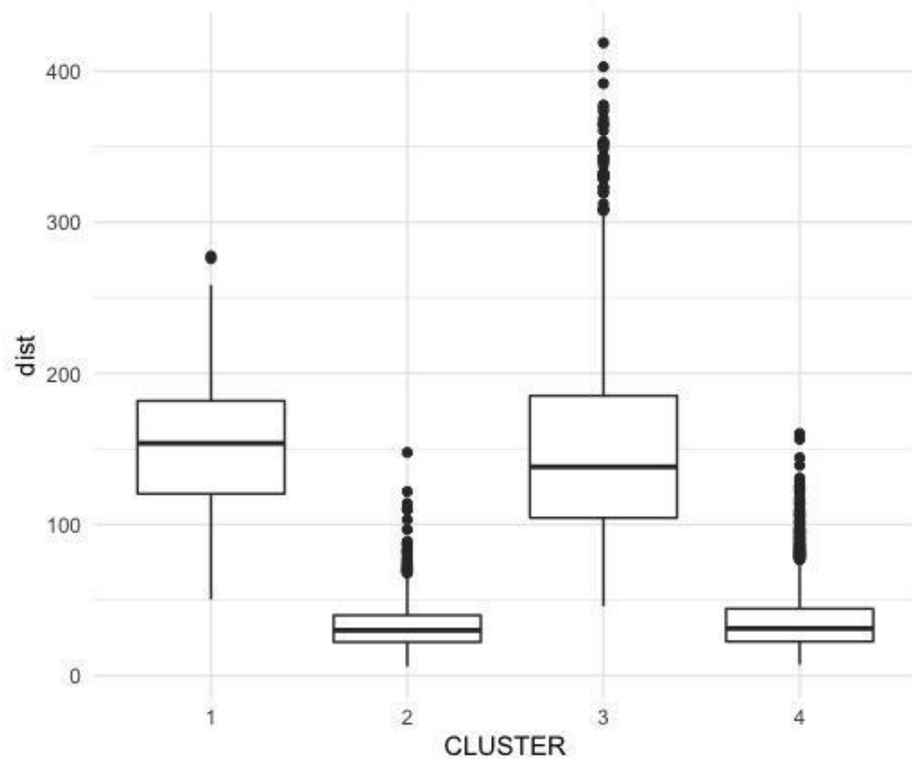


FIGURE 6-1 EACH PCSS DISTANCE FROM THE CLUSTER MEAN BY CLUSTER ASSIGNMENT

Several ways of ensuring this disproportionate effect did not affect the overall fit statistic were trialled, firstly by taking the median of the PCS variable scores before squaring them and also by removing extreme outliers that fell further than three standard deviations from the mean. By taking the median of squares instead of the sum of squares, the measure of central tendency was less affected by the extreme outliers. Whilst it did improve the scores overall, it was decided that the higher numbers caused by one or two variables still unrealistically skewed the data. By treating each PCS for outlying variables which fell above 3 standard deviations a much more realistic view of the overall fit of the classification was achieved, so areas were not being affected by a few variables with especially poor fit. It is fair to say however, that some naturally occurring variation may have been lost when compensating for the highest values and it is possible that these outliers provide interesting insight into the reasoning. Figure 6-2 to Figure 6-4 visualise each iteration.

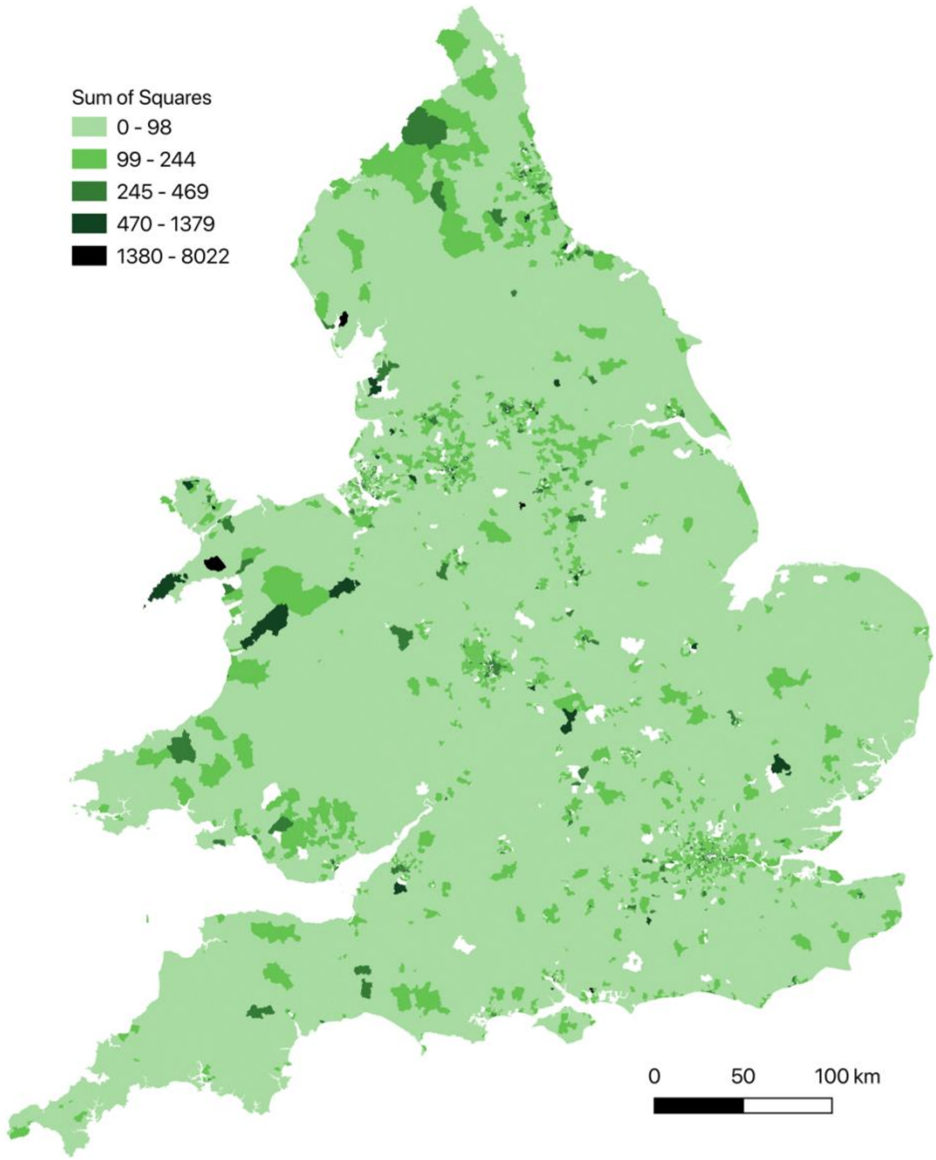


FIGURE 6-2 SUM OF SQUARES OUTLIER TREATMENT EFFECT ON OVERALL CLUSTER FIT

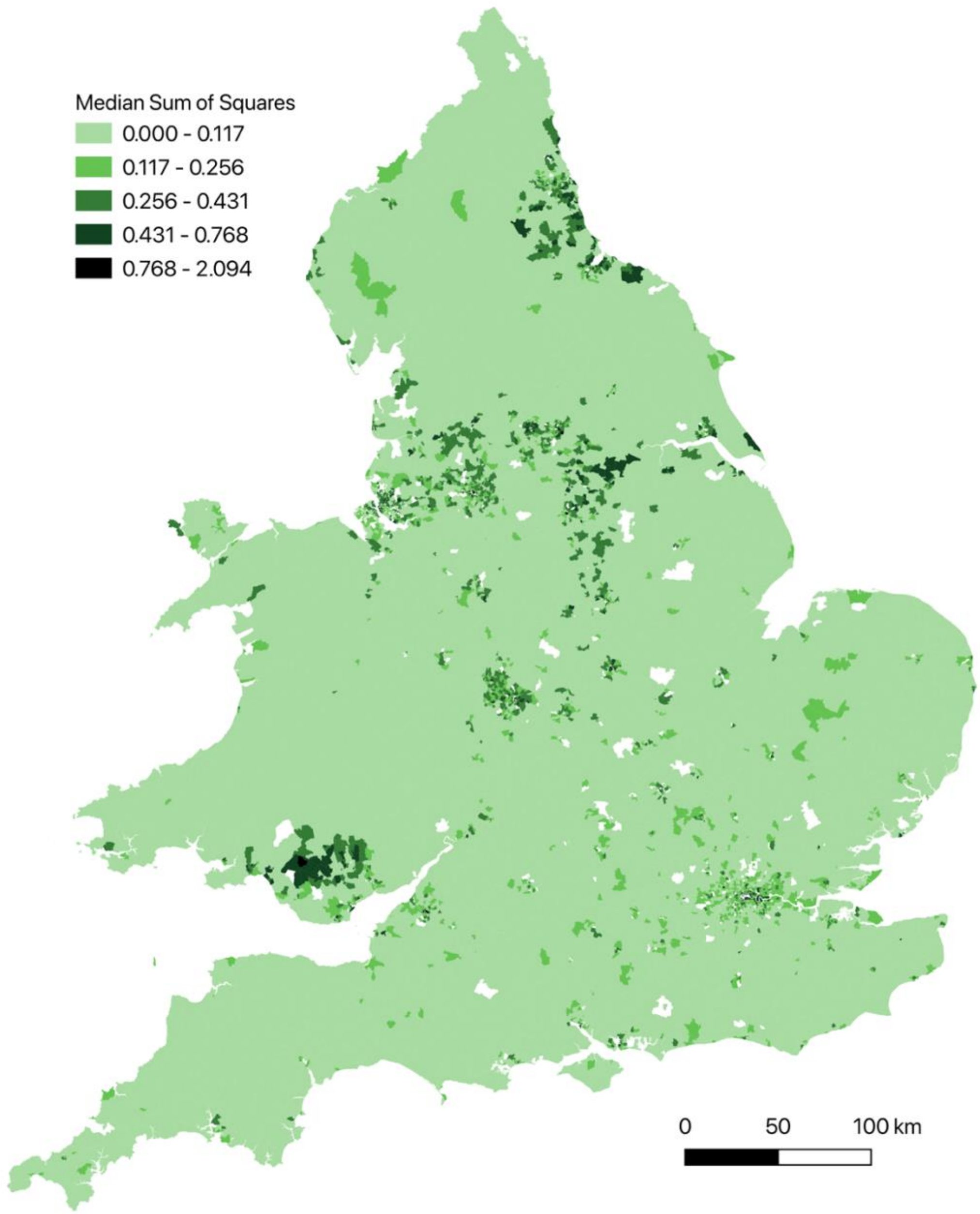


FIGURE 6-3 MEDIAN SUM OF SQUARES OUTLIER TREATMENT EFFECT ON OVERALL CLUSTER FIT

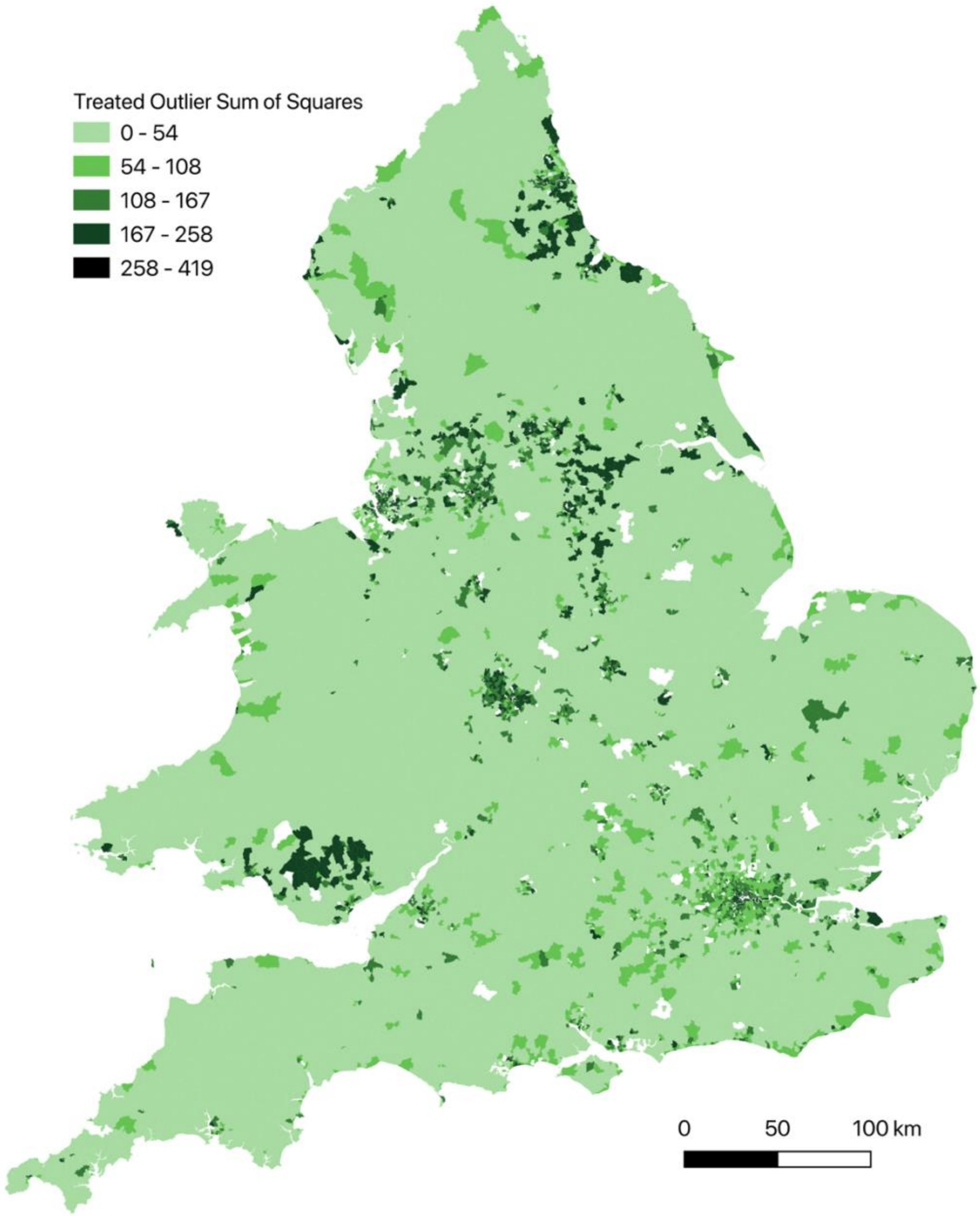


FIGURE 6-4 TREATED OUTLIER SUM OF SQUARE EFFECT ON OVERALL CLUSTER FIT

The final map shows that the cluster fit is good across the majority of England and Wales, however major cities such as London, Manchester, Liverpool and Birmingham are highlighted as a poorer fit of the classification. The Cardiff and Bristol areas around South Wales, as well as rural areas in the north of England around the Yorkshire Dales and Lake District are also highlighted. When investigated, the fit statistics showed that all but one of the highest cluster distances (>259) occur in the "Efficient City Living" Supergroup.

Overall, the classification is a worse fit in those areas where there is likely greater heterogeneity at this spatial scale, as you would expect to find of those living in inner city areas. To reaffirm this, the Rural Urban indicator was appended to the clusters fit measures and investigated, the results are detailed in Figure 6-5 below. It found that the more urban the area, the wider the range of the cluster fit statistic, backing up the notion that there is greater variety in more urban areas. This is also likely to be an effect of the scale at which these results are presented; as a factor of the MAUP, there is likely to be a greater level of heterogeneity at the PCS scale within urban areas due to their denser populations and complex urban environments that is masked here (See Chapter 2) (Openshaw 1984).

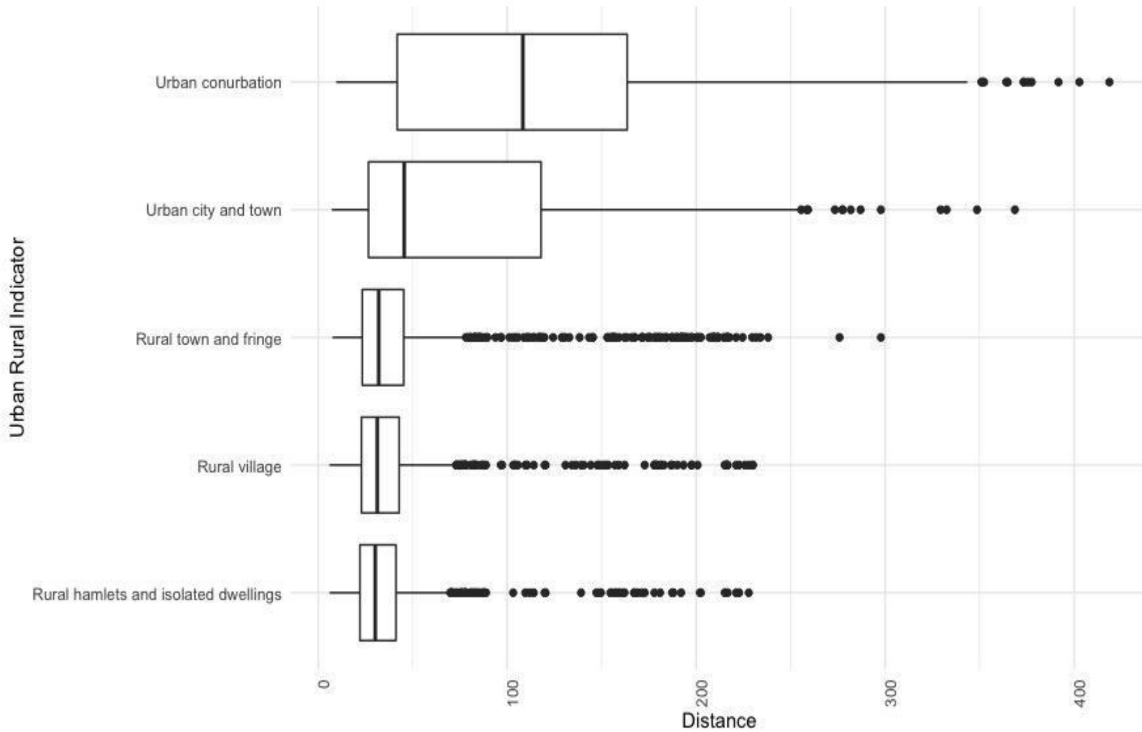


FIGURE 6-5 THE EFFECT OF RURALITY ON CLUSTER DISTANCE

Even after controlling for the disproportionate effect of some variables, a group of areas in the North of England were of particular interest because of their poor cluster fit, and so were more closely inspected to discern if there were any particular characteristics causing this. There were some notable examples, such as 'BD15', in the Local Authority of Bradford. It has been classified into 'Efficient City Living", but an unusually high number of children of school age and their parents in this area has caused a poor fit. The unemployment figure is also notably higher, most likely given the care responsibilities associated with young children - something you would not typically expect to find in an inner-city area.

Property types and their associated fixtures and fittings also account for some of the inaccuracy, especially in these Northern areas, where inner city accommodation can be a mixture of modernised flats and legacy housing. Miles Platting (M4 4) in Greater Manchester for example, is majoritively post war council owned accommodation, meaning that the area has a much larger number of houses with a gas supply than you would expect to see in an inner city area dominated by flats and apartments which are typically not connected to the mains gas supply. Figure 6-6 illustrates the spread of domestic gas supply in Manchester city centre and surrounding areas – it is clear that the inner city areas are characterised by a very high number of properties without a connection.

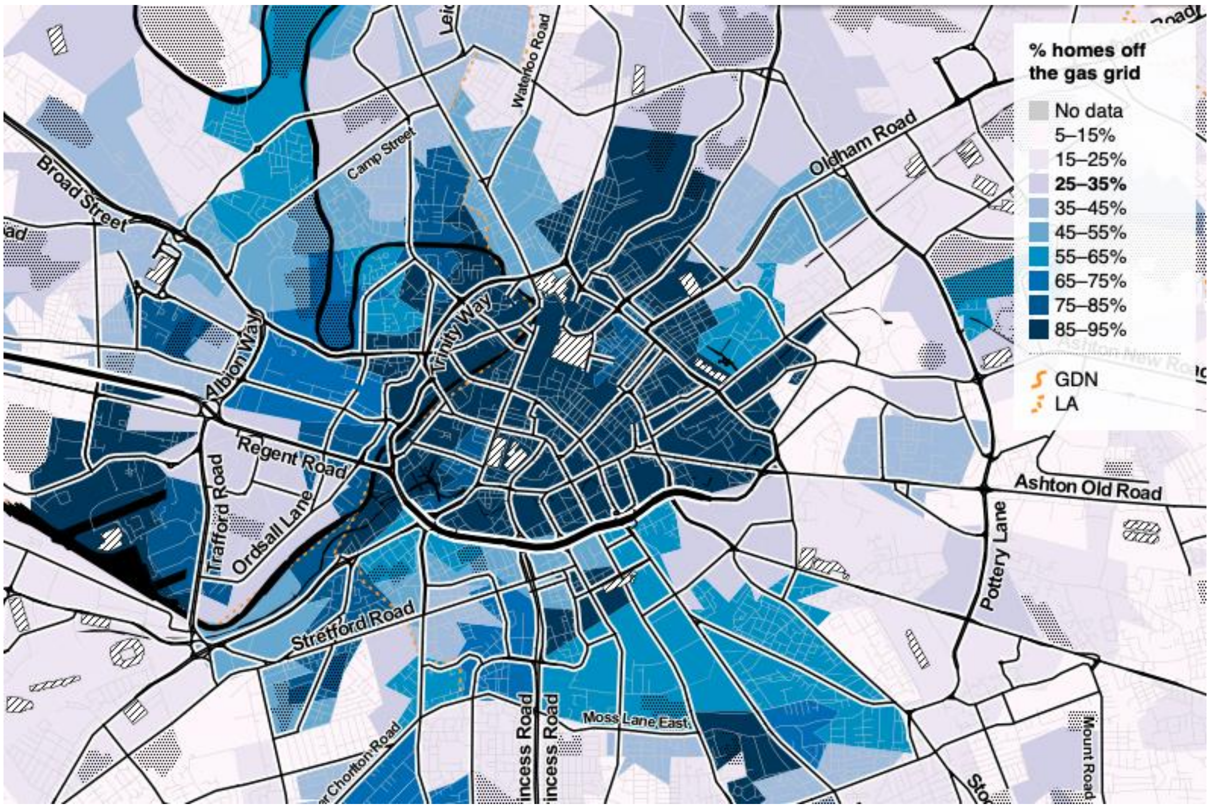


FIGURE 6-6 MANCHESTER CITY CENTRE MAIN GAS SUPPLY (Affordable Warmth Solutions, 2020)

Miles Platting (the area between Oldham Road and Ashton Old Road in the image above) pertains to the 3rd worst cluster fit. The variation appears to be related to the housing types found within this particular area and their occupants; who are mostly families with children of school age, divorced or separated, unemployed or in part time work and living in social tenanted accommodation with traditional fixtures and fittings. Yet its proximity to Manchester City Centre (under 2 miles) and the inclusion of new apartment blocks and rented properties has led to its classification as a member of

Supergroup 3. The image below in Figure 6-7 shows the typical housing in Miles Platting, shadowed by the newly renovated efficient apartment blocks of nearby city centre Manchester.

FIGURE 6-7 MILES PLATTING, MANCHESTER (MANCHESTER EVENING NEWS, 2019)

For those postcode sectors which had a fit statistic in the top decile, the same few variables were consistently responsible for the disproportionately high values. In Supergroup 3 variables relating to family dynamics such as school age children, marriage and owning the accommodation were most likely to skew the result, even when treated for outliers. In the case of those in Supergroup 1, an interesting example is TS 37 with a cluster distance of 277 - upon investigation it was found that this postcode sector, in the centre of Middlesborough, is closely associated with Teesside University, which explains the unusually high level of students, shared accommodation and people aged 20 to 44.

6.3 Mapping the Relationship between the Energy User Classification and Material Deprivation

To give further context to the properties of the Energy User Classification (EUC), additional descriptive statistics were appended to the dataset. Whilst not included in the original clustering model, they do give a detailed view of the characteristics associated with each Supergroup.

The EUC focuses specifically on fuel poverty and consumption characteristics, but as discussed in the literature (Chapter 2), fuel poverty and high level deprivation are intrinsically linked (Frederiks *et al.*, 2015). As such, this section considers the intersection between the Energy User Classification (EUC) and both English and Welsh Indices of Multiple Deprivation (IMD) deciles from 2015.

Following guidance from the Office for National Statistics, where it occurs that postcode sector and LSOA boundaries do not align, only one postcode sector per LSOA is allocated. This is the one which contains the majority of residents as indicated in the 2011 census. This method differs from the previous reweighting which has been employed due to the categorical nature of the dataset. Whilst it may lead to some loss of detail, it is a standardised and recognised methodology (Office for National Statistics, 2016b). Figure 6-8 overleaf shows cluster distribution across the IMD deciles and provides a good overview of the relationship between overall deprivation and cluster assignment.

Once in the IMD decile 4 or above, the proportion of the decile assigned to Supergroup 2 remains fairly constant, but it is Supergroup 3 that is the most notable here. Their pen portrait suggest that they are the highest earners and live in the most modern accommodation, however this figure implies that those in Supergroup 3 are actually likely to live in areas of relatively high deprivation. This could take into account the inner city areas where LSOA and PCS boundaries overlap, but it does help to reinforce the notion that the definition of fuel poverty is too narrow – it is known that deprivation and fuel poverty are intrinsically linked, yet we see here that all four clusters of the EUC are present at all levels of deprivation, albeit at varying degrees, with the exception of decile 1, where Supergroup 2 does not appear.

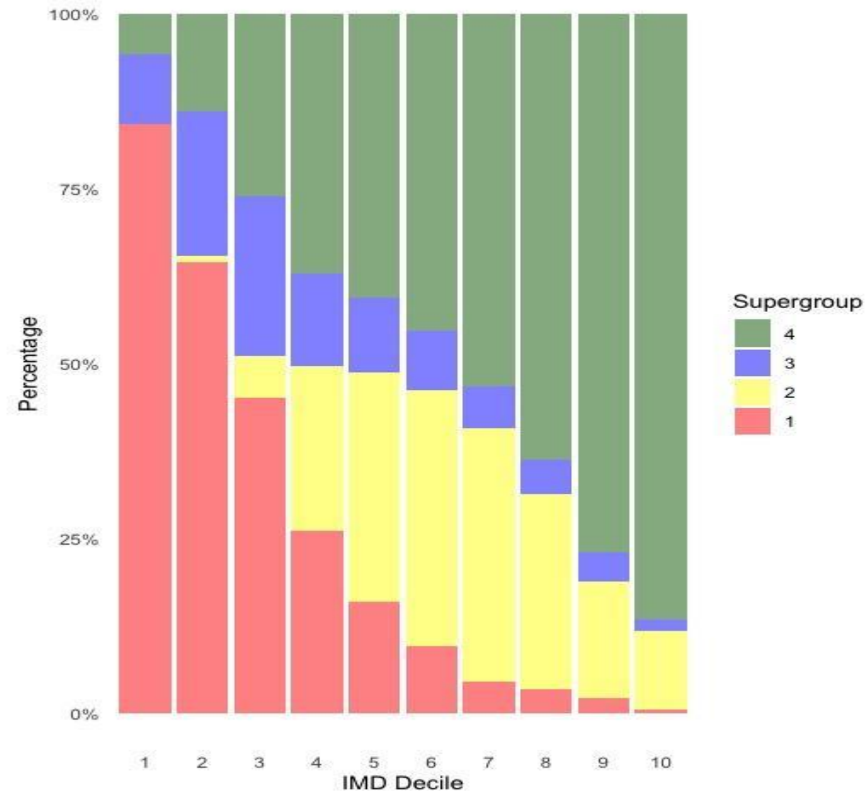


FIGURE 6-8 SUPERGROUP AND IMD INCOME DECILE CROSS-TABULATION

6.3.1 The Energy User Classification, Income and Fuel Poverty

Further to the exploration of the Energy User classification (EUC) and its links to material deprivation, it identifies that demographic indicators as well as the energy efficiency characteristics of homes could be utilised to provide a more nuanced understanding of the geographies of fuel poverty. Prospectively, such a classification might be used to improve targeting of those most vulnerable. Under the current definition low income is one of the prevailing factors in defining the fuel poor; although, as highlighted in previous sections, it is influenced by a range of wider factors. As such, this section considers the intersection between the Energy User Classification and differential levels of income taken from the Small Area Income Estimates, which provides data on before and after housing costs at an MSOA level for the years 2012, 2014 and 2016, giving a longitudinal view of changing household costs. The intersection is then compared to the current fuel poverty distribution to understand the similarities and differences from that which already exist, and to articulate the utility in a multifaceted approach to redefining fuel poverty.

As Figure 6-9 details, it is clear that the 'Cold and Costly' supergroup 1 are the lowest earners by some margin. They are also seeing the gap between BHC and AHC increase over time and whilst they are not the group with the largest financial burden of housing cost, relative to their overall income they are the only group to have seen an increase in the percentage of housing cost, as described by Figure 6-10. Their demographic characteristics lend themselves to also having other significant costs such as those associated with disabilities and expensive energy tariffs, as well as fixtures and fittings such as immersion heaters that make maintaining a level of thermal comfort difficult.

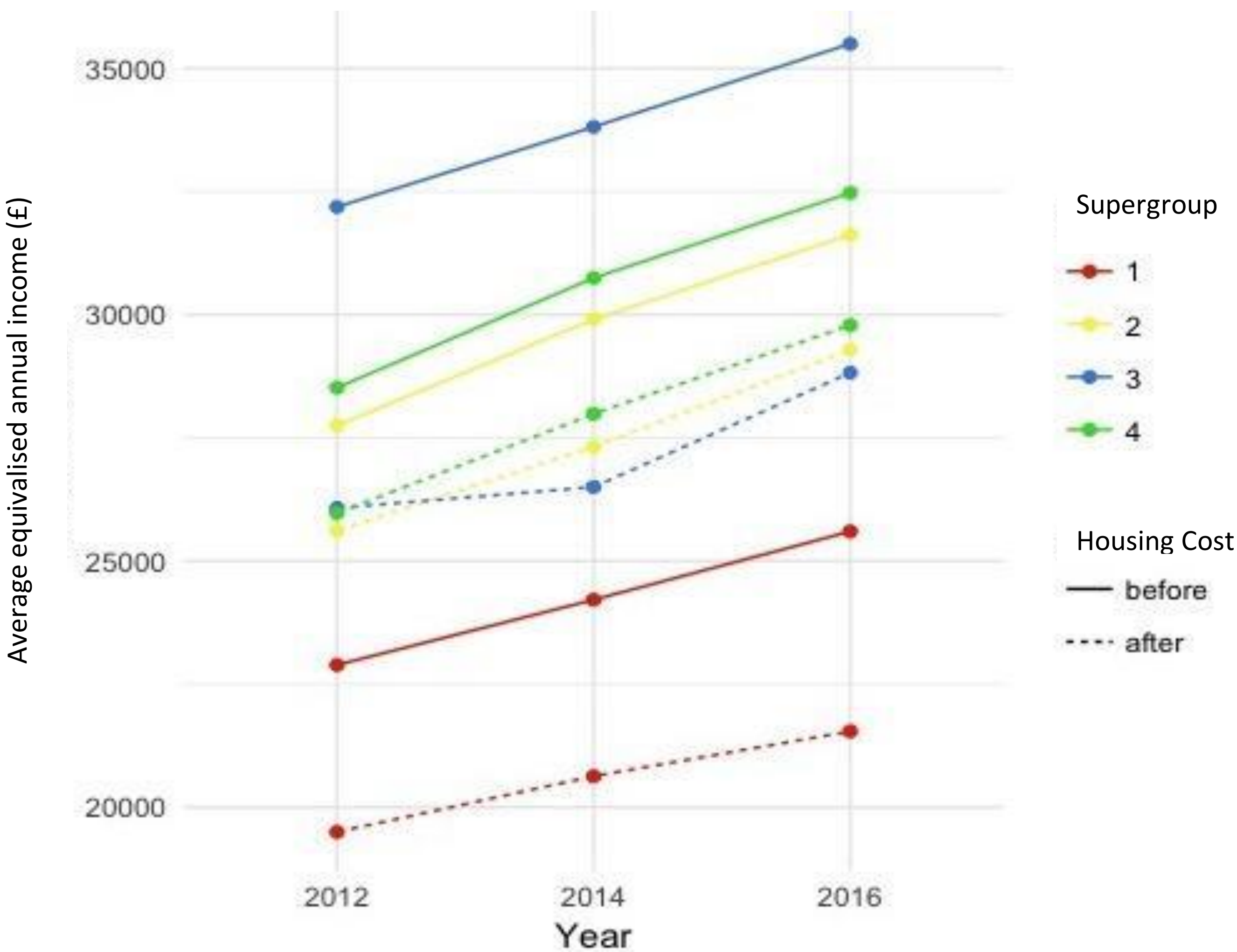


FIGURE 6-9 INCOME CHANGE BEFORE AND AFTER HOUSING COSTS BY SUPERGROUP 2012 - 2016

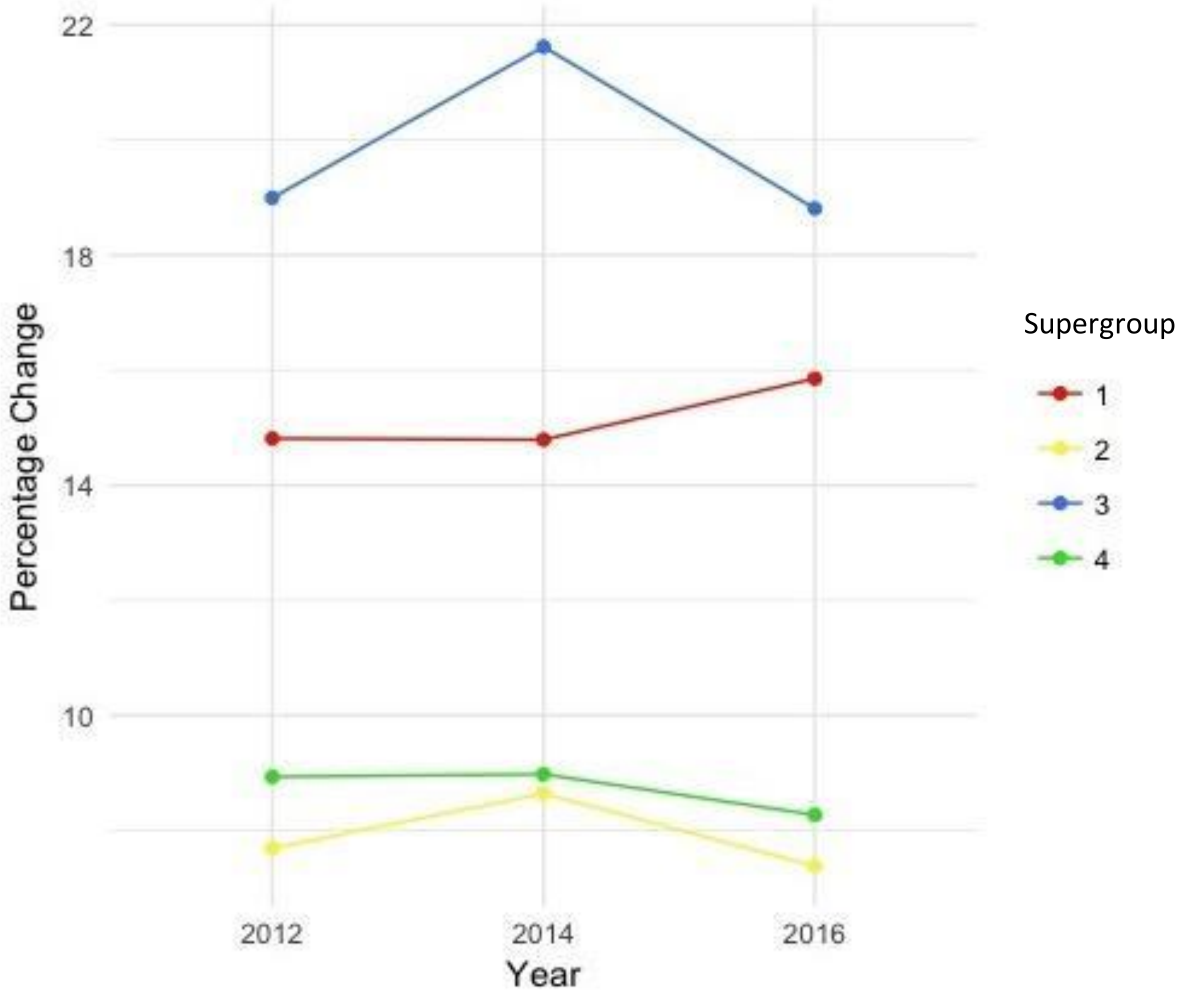


FIGURE 6-10 PERCENTAGE CHANGE IN INCOME AFTER HOUSING COSTS BY SUPERGROUP 2012 - 2016

Supergroup 3 'Efficient City Living' find themselves paying a substantially bigger proportion of their income towards housing cost due to their location in major towns and cities where rentals and sales command considerably higher prices. They have however seen the gap begin to close over time, meaning a larger proportion of their income is now disposable, but they are unlikely to consider using this money for improvements or energy bills as they already live in efficient housing, and do not generally struggle to cover their energy costs. Furthermore, they are more likely to rent their properties than the other Supergroups, and so would be unlikely to invest in any housing improvements as their returns on investment will be low.

Both the 'Off Grid Owners' (Supergroup 2) and 'Typical Tariff' (Supergroup 4) consumers have seen a small decrease in terms of percentage of income accounted for by housing cost, and a steady, comparable rise in both income and housing cost overall leads them to be the most stable of the clusters, possibly able to react better to both housing and energy cost price changes. The slight percentage decrease in housing cost may allow those who previously would have been in short term fuel poverty to plan for seasonality and reduce their energy bills by making some investment in efficiency measures; for example, by considering replacing boilers or the prevalent single glazing, or by moving away from solid fuels. They may also be in a position to reduce or clear outstanding debt with their energy provider, leaving them able to shop around for better deals and switching suppliers to save even more money.

While Figure 6-8 clearly details the relationship between each PCSs cluster assignment and its income decile, the jitter plot in Figure 6-11 below includes a third dimension to illustrate the varying degrees at which fuel poverty is present in these segments by the current definition. This allows for detailed insight into the relationship between the current fuel poverty definition and income, and also allows rationalizing of the demographic characteristics associated with each cluster and its relative level of fuel poverty as it stands. Each postcode sector is represented by a dot, the colour of which is dictated by its current level of fuel poverty. The more densely populated the grid square, the more Postcode Sectors are represented.

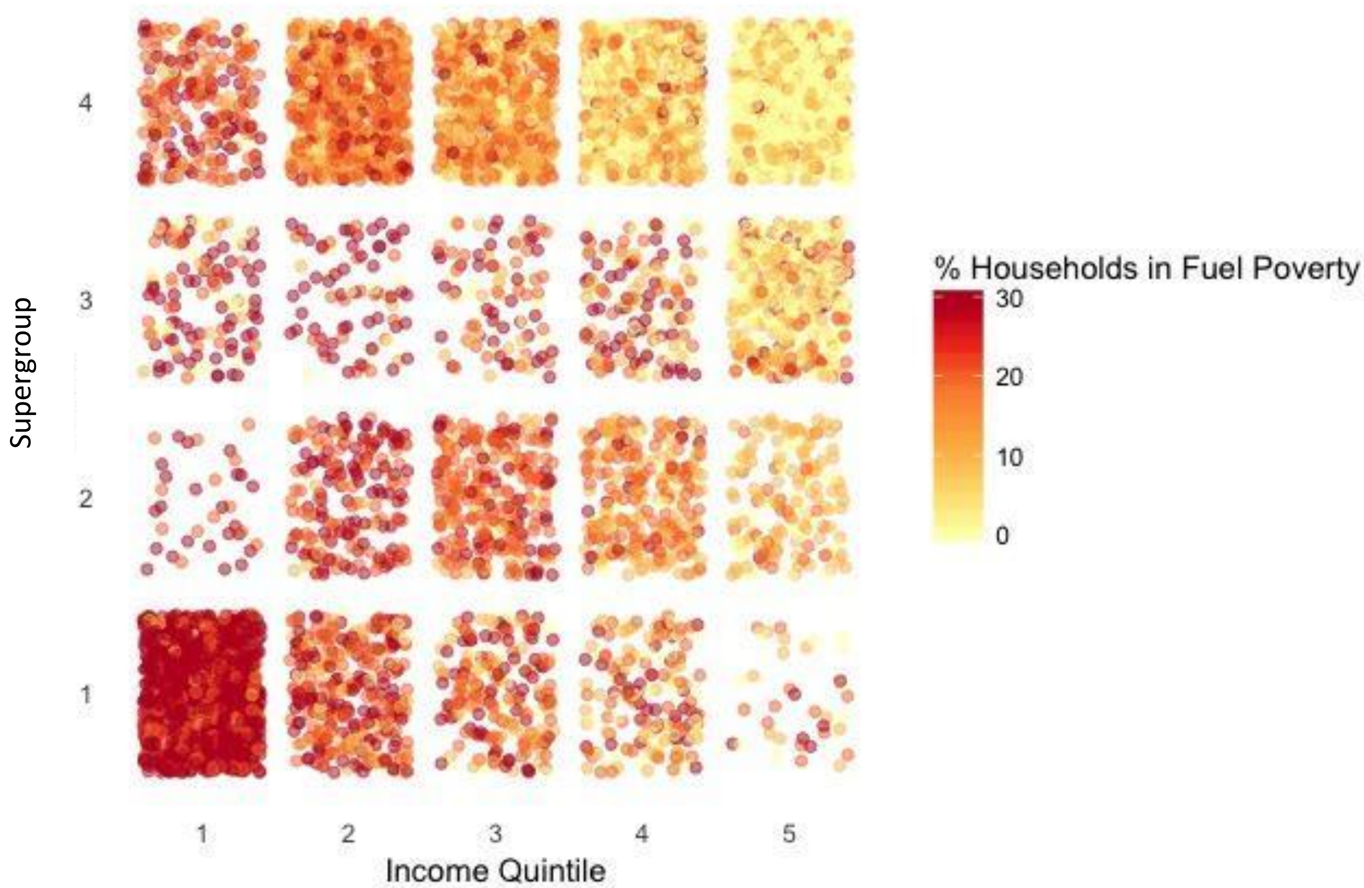


Figure 6-11 Current Fuel Poverty Levels by Income and Supergroup Assignment

Within Supergroup 1 the majority of the PCSs assigned fall into the lowest income quintile, as can be seen by the density of the points in the lower left hand corner. However, it is important to notice that where areas do fall into the higher income quintiles, (even though they are fewer) there is still evidence of them suffering from a high percentage of homes considered to be in fuel poverty, as detailed by the darker coloured points. The Supergroup characteristics suggest that these areas are characterized by typically hard to heat, inefficient homes occupied by young renting families, so even if they are some of the highest earners their costs are still unaffordably high. Their tenancy arrangements are also more likely to be socially or privately rented, meaning they may struggle to access cheaper tariffs or make home improvements to reduce their costs because of restrictions imposed by their landlords.

As discussed above, the even spread of PCSs in Supergroup 2 across the income quintiles suggests that fuel poverty in this cluster in particular is more multidimensional than the current definition encompasses. Figure 6-11 clearly shows that levels of fuel poverty under the current definition are still prevalent in the middle income quintiles. The cluster characteristics suggest it is much more likely to be caused by old, inefficient and under occupied buildings. The occupants are middle aged or retired and are less likely to be connected to the mains gas, instead relying on coal, oil or wood as their main fuel source. Whilst these households may appear most prevalently in the higher income quintiles, they also accept higher energy costs and so are ignored under the current definition, but nonetheless may struggle to heat their homes to a comfortable temperature, even after they accept an increased cost.

Supergroup 3 is particularly interesting in our argument for expanding the current fuel poverty definition as this cluster is generally considered to be the highest earning. Whilst those in the highest income quintile suffer less fuel poverty than others, the demographic characteristics associated with this cluster still appear at other income levels, which do show very high levels of fuel poverty. This could be representative of those who live in inner city areas but are low earners struggling to meet the additional cost of city centre living, despite living in relatively energy efficient accommodations. This corroborates our previous analysis of areas such as Miles Platting. Those in the highest income quintile are as previously discussed, most likely to live in new and refurbished efficient properties with a low floor area and access to the cheapest tariffs, so not only are they able to take advantage of low bills, but also have the most disposable income to meet unexpected costs without becoming temporarily fuel poor.

Given the occurrence of areas classified as Supergroup 4 in current fuel poverty across the income quintiles in this, the most average category, it would suggest that they are on some level at risk of falling into fuel poverty at any time and only those with the highest wages and the lowest costs are unlikely to struggle, but one or the other does not guarantee thermal comfort, regardless of their housing, tenancy or family arrangements.

6.4 Demographic Constraints on Smart Meter Adoption Rates

Understanding the relationship between the Energy User Classification (EUC) Supergroup to which an area is assigned and its smart meter adoption rate (SMAR) allows us to infer the demographic constraints associated with smart meter adoption, as well as how this then impacts the household's energy profile. It is well reported in existing literature (Chapter 2) that the value of a smart meter varies greatly depending on the rationale for having one but can in the right circumstances decrease consumption by 14%; smart meter users on a pre-payment tariff were especially motivated to make a saving through engagement with their smart meters. Other studies found an average 7% decrease in consumption where smart meters were adopted (Faruqui *et al.*, 2010; Ehrhardt-Martinez and John, 2010). It is unsurprising then that Supergroup 3 exhibits the lowest average adoption rates of the four clusters, as described in Table 6-1. As previously discussed, the residents of this cluster are likely to be higher earners and the least likely to need help in reducing either their bills or consumption - either because they are already low, or because they can afford the cost of high usage - the two main advantages to the consumer of having a smart meter installed.

TABLE 6-1 SMART METER ADOPTION RATES BY CLUSTER

Supergroup	Mean Adoption Rate
1	2.69
2	1.63
3	1.46
4	2.58

Supergroup 1 have the highest average adoption rates overall. This might be indicative of this group taking active steps to reduce their energy outgoings by having a smart meter installed; this Supergroup have been characterised as the most at risk of fuel poverty given their energy and demographic characteristics and so would benefit from reduced energy costs. It is worth noting that whilst Supergroup 1 presents the highest average adoption rate overall, relatively speaking, 2.7% is still extremely low in regard to the Government directive to place one inside every household.

Supergroup 3s particularly low SMAR is likely to be as a result of their tenancy or the physical properties of their accommodation – first generation smart metering technology is more likely to fail in apartment blocks because of their centralised metering systems and restricted access (Section 2.2.5). Rental agreements reduce the autonomy of the tenant and restrict their ability to make changes to the fixtures and fittings of their accommodation (Hope and Booth, 2014). Supergroup 2 also show a very low SMAR, likely as a result of their lower number of connections to the mains supply. Furthermore, there are characteristics not reflected in the EUC such as lack of engagement with innovative technologies - those with lower levels of computer literacy are less likely to shop around for energy tariffs, less able to access educational material regarding the benefits of smart meters and even go so far as to be unable to request or book a smart meter appointment, the vast majority of which is now done through online billing accounts. This relationship with technological engagement is explored in greater detail in the following section.

6.5 Case Study

In order to demonstrate the practical viability of utilising the Energy User Classification (EUC) in redefining fuel poverty as a multifaceted phenomenon, the following section examines the socio-demographic characteristics in tandem with the SMAR for each PCS in order to address the likely constraints to improving SMAR and allow policy stakeholders to understand the causal mechanisms of fuel poverty by providing an illustrative case study of how the DEP might utilise tools such as the EUC to optimise the targeting of their energy poverty intervention measures to have a greater social impact.

This case study attempts to take an localised view of the 'Cold and Costly' supergroup, who present as the supergroup with the most socio-demographic indicators of fuel poverty risk. Despite having the highest SMAR of all the clusters, relatively it is still extremely low or zero in some areas. This case study selects an area where there is a cluster of PCSs with especially low SMAR as they present the greatest opportunity for improvement from the DEP. To narrow down suitable case study areas; based on the fallacy that 'birds of a feather flock together' and the notion that greater visibility and word of mouth are helpful tools in increasing the SMAR of an area; people see their friends and family start to achieve savings and learn to trust the technology (Buchanan *et al.*, 2016), clusters of high SMAR which neighbour areas of very low SMAR were identified with the application of a LISA analysis. This first step is detailed in section 6.5.1 and a case study area is chosen based on the spatial proximity of high and low PCSs.

The second stage of the analysis closely examines the characteristics of the case study area; firstly the Supergroup assignments of areas neighbouring those which are Supergroup 1, the SMARs for each PCS within the area and finally, to investigate other socio-demographic constraints to accessing smart metering technology, each areas Internet User Classification. Finally, once all these conditions have been examined, the section concludes by offering suggestions to address and overcome some of the caveats that could be acknowledged by the DEP to increase SMARs, objectively increasing the likelihood of meeting government guidelines, whilst also acknowledging the importance of the lived experience of fuel poverty and utilising smart meters for social good.

6.5.1 Identifying a Suitable Case Study Area

As discussed in the introduction to this section, the evidence thus far suggests that the 'Cold and Costly' Supergroup display the characteristics most closely associated with fuel poverty, and as such, a subset of the dataset was taken to include only them. To identify those areas which have low SMAR but are surrounded by areas with a high SMAR (and so may be more likely to adopt smart metering under the right conditions) a Morans I test was undertaken. The Local Morans I identifies clusters of high-low values as well as low-low and high-high and also indicates to what extent this clustering occurs. The result indicates that areas of high and low SMAR do cluster spatially within the Supergroup (score - 0.463, p-value <0.05). Within this it is also possible to extract a LISA score (Local Spatial Autocorrelation) which is indicative of the extent of the significant spatial clustering around each observation (in this case, for each PCS) (Anselin, 1995).

Of the 1889 Postcode Sectors within the Supergroup, 277 had a significant LISA p value of <= 0.05, indicating that the spatial clustering has not occurred randomly. A subset was extracted of the intersection of those significant areas which also have a value indicative of being a high-low observation (an lI value of between 2.03 and 5.25, resulting in 83 areas - indicating that these areas contributed significantly to a negative global autocorrelation outcome). The areas where the high-low clusters also have a significant LISA value underwent a visual inspection and a several local areas where a cluster of PCSs firstly belonged to the Cold and Costly Supergroup as well as presenting the high-low SMAR LISA characteristic presented themselves. By far the most prevalent clustering occurred in the Wolverhampton area, as detailed in Table 6-2 overleaf. Other areas in England and Wales also met the criteria, but none were located near each other to this extent.

TABLE 6-2 WS AND WV NEIGHBOURING POSTCODE SECTORS

Postcode Sector	District	lI	pval	SMAR
WS20	Willenhall	3.500755	0.0000000	4.95
WS27	Walsall	2.763879	0.0000000	4.59
WS29	Walsall	5.247405	0.0000000	5.71
WS31	Walsall	3.017759	0.0000001	4.86
WS32	Walsall	3.471239	0.0000000	5.55
WS86	Brownhills	2.899146	0.0000202	6.10
WV108	Wolverhampton	2.356741	0.0000012	5.56
WV131	Willenhall	3.737764	0.0000000	4.85
WV132	Willenhall	2.470220	0.0000091	4.47
WV133	Wolverhampton	4.171641	0.0000000	5.84
WV146	Wolverhampton	4.276230	0.0000000	5.49
WV23	Wolverhampton	2.749314	0.0000009	4.58

To avoid having a case study area with missing polygons, and to understand the demographic characteristics of surrounding areas, the case study area consists of all the PCSs within the WV and WS Postcode Areas. Figure 6-12 contextualises the case study area. It is important to remember that the high values are relative to the EUC Supergroup and not the dataset as a whole.

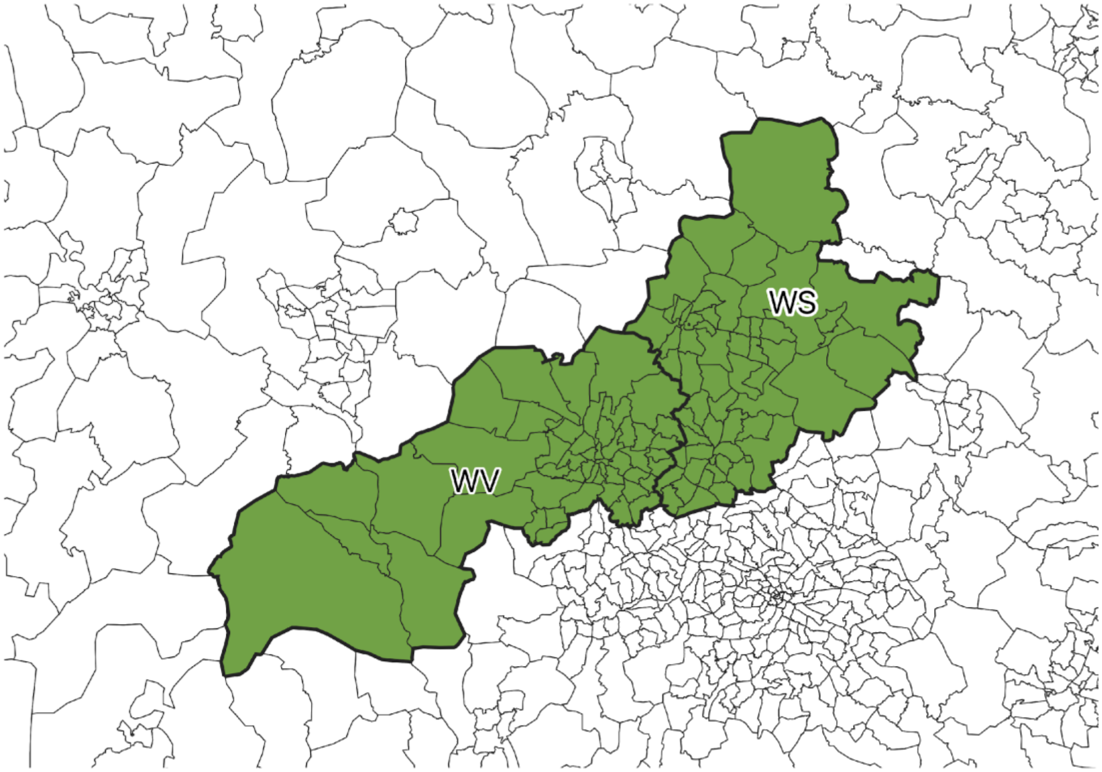


FIGURE 6-12 WOLVERHAMPTON WS AND WV POSTCODE AREAS

To understand the underlying characteristics within the case study area where postcode sectors are not assigned to the 'Cold and Costly' Supergroup, Table 6-3 and Figure 6-13 also detail distribution of the other 3. The majority (51%) of surrounding areas fall into Supergroup 4, suggesting that many areas show characteristics close to the national average. Only 3% are in Supergroup 3 and considered least likely to find themselves suffering from fuel poverty.

Supergroup	**Count**	**Percentage**
1 – Cold and Costly	41	38%
2 – Off Grid Owners	5	4%
3 – Efficient City Living	4	3%

TABLE 6-3	4 – Typical Tariff	55	51%
	Missing	1	<1%

WOLVERHAMPTON CASE STUDY AREA CLUSTER ASSIGNMENTS

Figure 6-13 details the distribution of the Supergroups, showing clusters of Supergroups 1 and 4, with only a few disjoined areas of Supergroup 2 and even fewer of Supergroup 3. This could be indicative of an area that is considered fairly similar to the national average but contains pockets of deprivation. The areas of Supergroup 3 are very close to the city centre and are also home to the University of Wolverhampton student halls of residence.

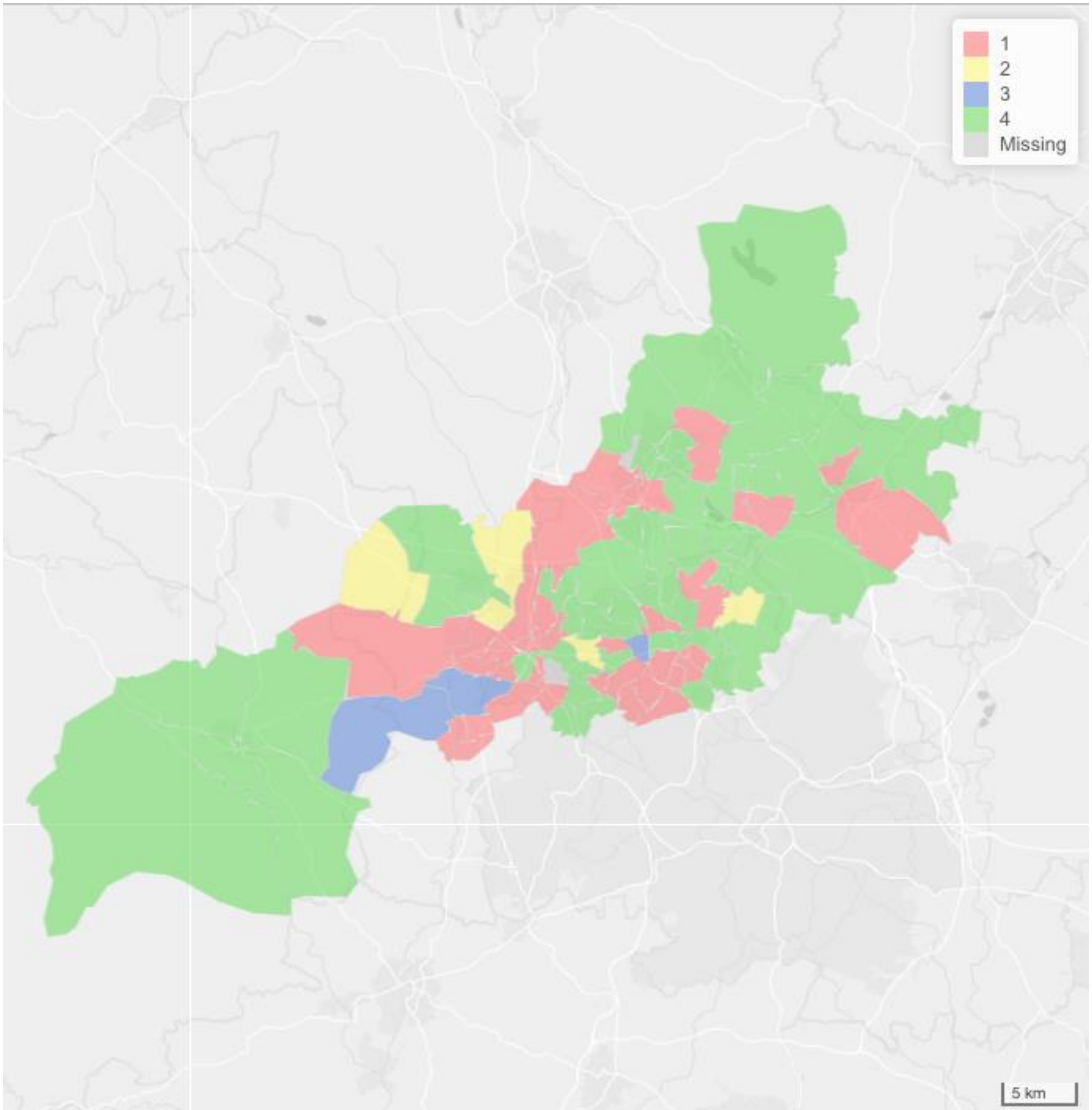


FIGURE 6-13 GEOGRAPHIC DISTRIBUTION OF SUPERGROUP ASSIGNMENTS IN THE CASE STUDY AREA

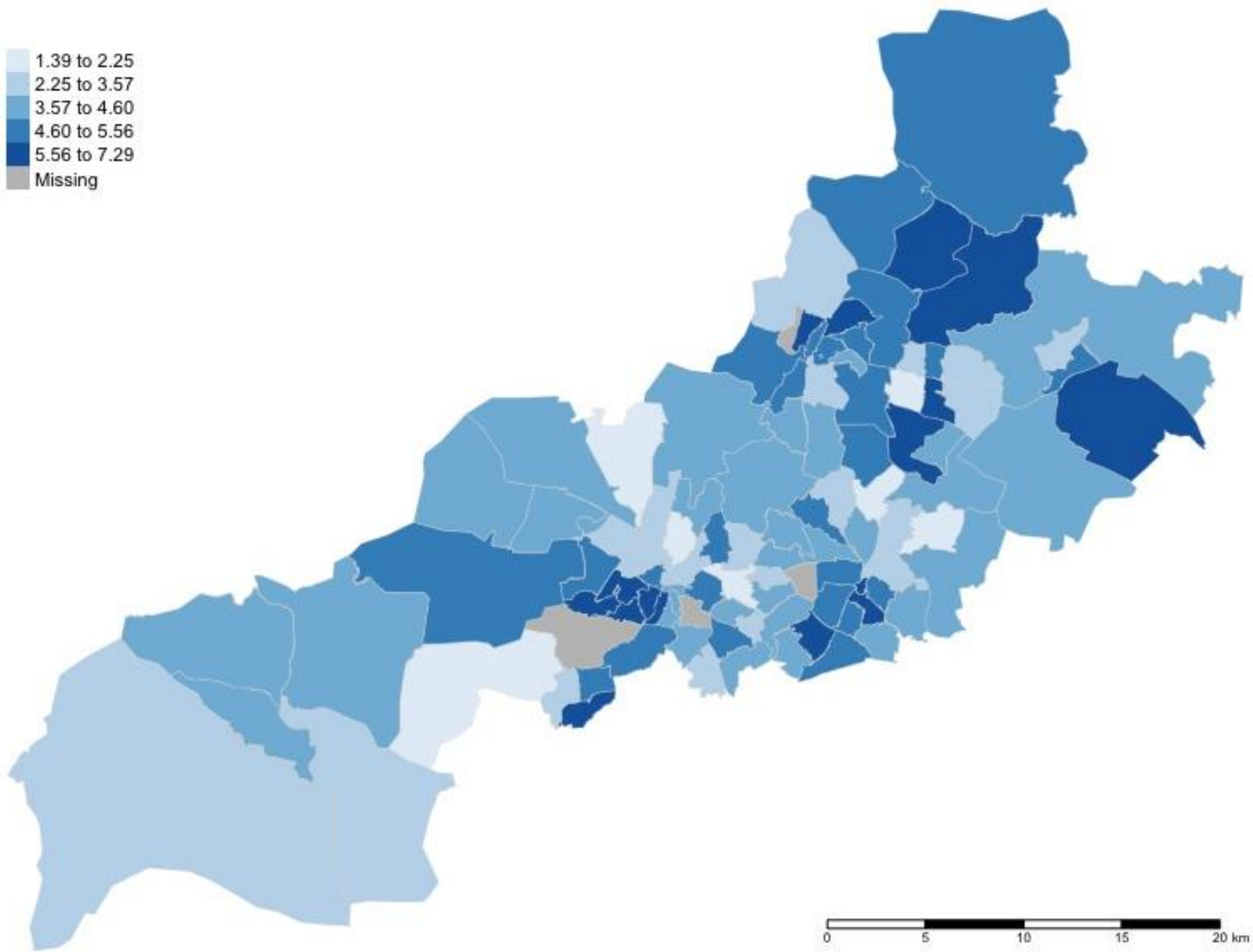


FIGURE 6-14 SMART METER ADOPTION RATES IN WOLVERHAMPTON

Figure 6-14 shows the distribution of SMARs in Wolverhampton, with dispersed areas of high adoption and more prevalent clusters of low to middling SMARs.

6.5.2 Wolverhampton and The IUC

The links between SMARs and the education surrounding them was discussed in the literature at the beginning of this chapter and the relationship between the ability to access educational materials and the Internet access are implicit. The utility of the Internet User Classification (IUC) is discussed in

Section 3.4.2; it is rich in information relating to understanding how a community engages with the Internet and new technologies more generally, as well as how they access information and services. Here the intersection between the areas in the case study and the CDRCs IUC have been investigated; by understanding the levels of internet engagement within the PCSs included in the case study area, it may reveal other factors that act as constraints to improving SMARs beyond those already discussed.

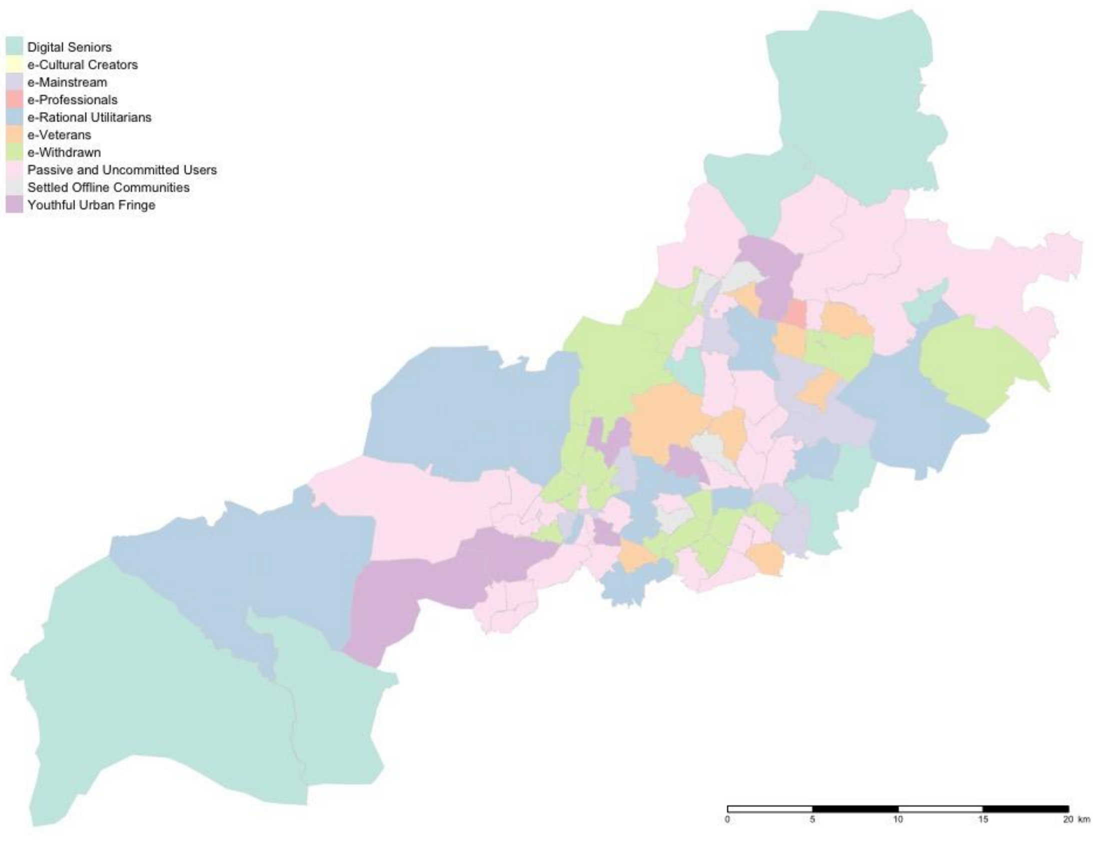


FIGURE 6-15 CASE STUDY AREA IUC SUPERGROUPS

TABLE 6-4 EUC AND IUC INTERSECTIONAL CHARACTERISTICS

Supergroup	IUC Group	Count	SMAR	Mean % of homes in current fuel poverty
1	Digital Seniors	1	3.40	12.75
1	e-Mainstream	1	3.36	13.69
1	e-Withdrawn	17	4.57	18.05
1	Passive and Uncommitted Users	21	4.68	15.15
1	Youthful Urban Fringe	1	5.36	15.41
2	e-Rational Utilitarian’s	5	2.40	13.65
3	e-Withdrawn	1	NA	15.54
3	Youthful Urban Fringe	2	1.39	15.57
4	Digital Seniors	6	4.26	10.53
4	e-Mainstream	8	4.61	11.60
4	e-Professionals	1	3.09	9.57
4	e-Rational Utilitarian’s	16	4.24	11.48
4	e-Veterans	8	3.88	9.28
4	Passive and Uncommitted Users	10	4.81	13.02
4	Settled Offline Communities	4	5.36	11.34
4	Youthful Urban Fringe	2	3.69	15.63

The distribution of IUC groups and their intersection with both the Energy User Classification (EUC) and smart meter adoption rates (SMAR) is detailed in Table 6-4 and Figure 6-15

The most prominent IUC cluster within the case study area is the "Passive and Uncommitted Users" (31 areas). Residents within these areas will typically have "limited or no interaction" with the Internet suggesting an overall disconnect from technology in general. More interestingly, their very low score for using the Internet for information seeking and financial services suggests that these groups are the least likely to undertake household management online; this includes online banking but also managing utility accounts online or receive bills by email, and therefore are not exposed to the smart metering advertisements and opportunities to request installation online. They are also less likely to seek out educational material or conduct research into the smart meter benefits and their very low score for having broadband access within the home may indicate an overall reluctance or distrust in new technology. However, both their current fuel poverty score and their SMAR falls into the upper quartiles, which may suggest that they are prepared to take necessary actions to reduce their consumption or bills. 21 of the 31 areas are also in Supergroup 1, suggesting they are more likely to be at home during the day due to their employment characteristics, and may be responsive to smart meter installations through doorstep targeting, where information and demonstrations, as well as follow up appointments can be carried out and arranged face to face.

The second most overrepresented group are the "e-Rationale Utilitarian's", who are characterized within the IUC as residents of areas with high demand for the Internet despite poor infrastructure with low broadband speeds, and fewer mobile devices. They use the Internet for utilities such as Internet banking and information seeking, perhaps meaning they are managing their energy bills via an online account and engaging in the educational material around the benefits of smart meters. In this case study areas, the majority are also categorized into Supergroup 4, who display the least distinctive socio-economic characteristics and are close to the national average, but the intersection with the IUC implies that these areas in particular are more likely to be retired homeowners who are again likely to be at home during the day to enable smart meter installations to take place. As homeowners they are also less likely than those in Supergroup 1 to be constrained by tenancy rules but may also feel that a smart meter isn't currently required as they do not find their energy bills particularly difficult to manage.

The group displaying the highest SMARs overall are the members of Supergroup 1 and the "Youthful Urban Fringe". They also have one of the highest rates of current fuel poverty. This group are likely tech savvy and highly engaged, utilizing mobile apps for household management tasks - they appear to be taking action to reduce their consumption by engaging with smart metering technology. Their SMAR is likely to be constrained by living in rented accommodation and they are unlikely to be able to afford to replace older appliances with energy efficient ones, especially if they have been provided for them by as part of a tenancy agreement.

Of particular interest is the group within Supergroup 1, who are also classed as "e-Withdrawn". They have a below average SMAR and the highest percentage of houses in current fuel poverty. This group are of particular relevance when attempting to understand the constraints to SMAR and understanding the limitations of the current definition of fuel poverty. They are disengaged, financially constrained and both currently in fuel poverty and considered at risk of fuel poverty under the EUC. They have the highest rates of social housing, suggesting that these households do not have the autonomy to make energy efficiency decisions. This group shows the lowest rates of engagement with information seeking and services in the IUC, and in the EUC they are the most likely to be pre-payment meter customers. This amalgamation of characteristics leads to this group being trapped in expensive energy tariffs, unable to search for and take a cheaper energy deal; they do not have the opportunity to engage with switching and comparison websites. This group are the most likely to benefit from interventions from the Housing Associations (HAs) - the DEP should consider working in tandem with HAs to provide a service which enables this group to alleviate their fuel poverty through smart meter installations which are pre-agreed with the Housing Association to combat the difficulties faced by tenants in getting changes to their properties approved and making efforts to move their tenants away from pre-payment tariffs. It is important however that educational efforts are also made in order to ensure the tenants adopt long term habitual changes to see the most benefit.

Conversely, but unsurprisingly some of the lowest SMARs and levels of fuel poverty occur in the most digitally engaged - the e-Professionals and e-Veterans, especially those who intersect Supergroup 4. They are characterised by working age people, usually highly qualified and experienced Internet users. Their willingness to engage with new technologies, coupled with their higher earning power and relative disposable income (Figure 6-10 and **Error! Reference source not found.**) could suggest that they are the main group who are prepared to research and invest more heavily in home efficiency improvements, such as triple glazing and modern appliances. They are likely to have the financial ability to purchase up to date, energy efficient appliances, hence their low fuel poverty propensity. They are also the most likely to utilise the Internet for information and services, suggesting they frequently use the Internet for household management tasks such as energy switching and seeking out the cheapest tariffs. This results in lower bills, and a smart meter may be surplus to their requirements as they do not struggle to meet their consumption costs, nor do they feel they need to reduce their consumption for other reasons, such as environmental concerns.

From this case study of Wolverhampton, it is clear that there are constraints and facilitators acting on SMARs and fuel poverty levels. The young and digitally engaged who find themselves in fuel poverty are quick to act and uptake of smart metering technologies is relatively high, however, those with a similar level of fuel poverty but categorised as e-Withdrawn do not engage with smart meter technologies to the same degree. It is clear then that digital awareness and access to technology plays a part in SMARs. The lower SMARs could be as a result of less household management taking place online in this group, an inability to access educational materials or simply a distrust of the technology. This is something the DEP could seek to minimise by utilising a variety of marketing and educational channels, such as doorstep targeting, leaflets and in-home demonstrations.

The least constrained, those who sit in the intersection of Supergroup 4 and the e-Mainstream and e-Professionals also suffer from low SMARs, but not high fuel poverty levels, suggesting that there is a reason aside from technological awareness for their lack of engagement with smart metering technologies. It is possible that they simply do not need to monitor their consumption - they are likely to have efficient homes and appliances, as well as the disposable income to cover both their usual and unexpected costs. The DEP might consider an improved household management approach with this group to justify the installation of a smart meter, as they are likely to do much of it online - time of use tariffs, engaging in “If This Then That” technologies and the ability to have a fully integrated “Smart Home” might appeal to the very digitally aware.

It is important to recognise the difficulties faced by those in social and privately rented accommodations when making these recommendations; the DEP could make a concerted effort to engage landlords and Housing Associations in order to educate them on the benefits of allowing and encouraging their tenants to install smart meters; this is likely to be the biggest challenge to overcome when considering how to increase SMARs as a whole.

6.6 Conclusions

Throughout this chapter, the overarching aim has been to show the utility of the Energy User Classification with regard to understanding the complex geographies and socio-economic indicators of fuel poverty above and beyond the current definition, as well as the geographies of smart meters and finally the intersection between the two.

A validation of the EUC showed a reasonable level of cluster fit across England and Wales with major cities showing more variation; Supergroup 3 contained the highest outliers and when the Urban Rural indicator was appended to the dataset, urban conurbations showed the widest range in cluster fit. This is to be expected in major cities, where a greater density of people with a larger variety of characteristics live in much closer proximity to one and other. The classification is by no means perfect, and some of the areas where cluster fit was at its worst were investigated more thoroughly to understand their characteristics. This found that variables related to property type and family life stage were most likely to result in inaccuracy or poor cluster fit; young families living in semi-detached accommodation in typically inner city areas in particular.

Additional characteristics were appended to the classification to provide additional context and the inclusion of the IMD decile and income data confirmed the notion that the current fuel poverty definition is far too narrow, by revealing that demographic characteristics of fuel poverty are present at all levels of income, albeit at varying degrees. This is exemplified in Figure 6.11, which clearly visualises the relationship between income and fuel poverty across the four Supergroups.

To create a holistic view of the intersection between smart meter adoption rates and fuel poverty, the independent analyses from Chapters 4 and 5 were considered in tandem, revealing some clear demographic constraints to accessing smart metering technologies; in particular; accommodation types and tenancy agreements. Finally, to address the low SMARs as a whole and to prove the utility of the EUC in addressing fuel poverty levels, a case study area was used to offer some recommendation to the DEP as to where they would likely have the greatest social impact through undertaking targeted smart meter installations.

This detailed practical validation found that for a successful potential increase in targeted installations, the key characteristics of an area should be that it has an already low SMAR and should also have persistently high levels of fuel poverty under the current government definition, but importantly are also engaged with current technology to a good standard. This equips them with the autonomy to make an informed decision and seek out educational material around smart metering and consumption reduction techniques. It is not imperative that they have the most energy efficient devices, as there will be savings to be made based on the consumption behaviours. In order for successful installations to take place, the household should not be limited by constraints imposed by landlords, suggesting homeowners and those with mortgages are the ideal candidates. That is not to say that other areas would be unsuccessful, if consideration were given to the constraints and facilitators acting on people's likelihood to engage. It is important to consider that those who are digitally disengaged may accept a smart meter if they are approached on the doorstep - they may simply not have been exposed to the digital marketing, as they are less likely to undertake their household management tasks online. Furthermore, constraints imposed by landlords should be given special consideration by the DEP in order for SMARs to increase. By working with housing associations and landlords rather than tenants, the benefits to both parties can be properly disseminated, and the issue of tenants making unwanted changes to a property outside of their agreement can be overcome. However, issues of privacy must be properly addressed in order that the tenant retains autonomy once the device is installed and the landlord does not interfere unnecessarily with the tenant's usage.

In conclusion, the current fuel poverty definition is too narrow in scope; it does not account for the demographic factors confirmed here; precarious tenancy agreements, digital disengagement, poor quality accommodation and fixtures and fittings, the inability to move away from expensive tariffs and suppliers. This leaves people increasingly vulnerable to perpetual fuel poverty, which is exacerbated by the lack of smart meter installations taking place in the most at risk locations, due to disempowered tenants, lack of education and limited technological understanding. In order to see SMARs rise, which will empower consumers to make more informed decisions around their energy consumption, the DEP must engage with these constraints and work to alleviate them. Increased communication and education will increase the trust in the devices, but for those who are disempowered and living in with precarious tenancy agreements, working with landlords and housing associations is likely to have the largest overall impact. By working in this way, it simultaneously addresses the fact that these groups are typically the ones that also suffer the highest levels of disengagement, disempowerment and fuel poverty and stand to see the greatest social impact from the DEPs interventions either thermally or financially.

7 Discussion, Recommendation and Research Prospects

7.1 Introduction

This thesis has brought together multiple data sources to investigate fuel poverty, energy consumption and smart meter adoption rates across England and Wales. The aims were twofold and addressed policy challenges relating to the existing fuel poverty definitions and the disjointed framework of energy stakeholders in England and Wales, paying particular regard to the utility of big data in resolving these challenges.

Studies of consumer energy consumption have previously been limited by a lack of data, or rather a lack of access to rich spatio-temporally granular data sources. The commercial sensitivity of such data necessitates that these are accessed only under the most rigorous circumstances to avoid commercial competitive advantages being lost and potentially sensitive customer data being misused. To fully capitalise on access to previously unseen commercial data and apply it within a social science research setting, the aim of this thesis has been to provide a thorough exploration of the geography of energy consumption and those factors that contextualise differentiated access to, and consumption of, both gas and electricity in England and Wales. By combining innovative big data and traditional open data sources, a holistic view of the current geographies of energy consumption can be presented, and our understanding of the lived experience of fuel poverty can be enriched beyond the current definition by including demographic characteristics. A literature review covering energy and energy policy at a variety of scales, material deprivation and the fuel poverty vernacular is followed by a chapter summarising the data and methodologies. The three empirical chapters firstly provide validation of the energy data utilised within the thesis, followed by their integration within a geodemographic framework to provide new insights into fuel poverty and barriers to smart meter adoption.

This chapter concludes this thesis by drawing together the findings to consolidate the contributions of this thesis by discussing the applications and implications of these work both in the context of geodemographics but also more broadly in terms of the various stakeholders in the energy landscape, such as those of researchers, policy makers and energy suppliers. A reflection on the data and methodologies employed and acknowledging the known and discovered limitations follows this. The discussion is concluded by delineating the key findings from the results of this research and identifying pathways for future work to build on what has been achieved here.

7.2 Implication and Application

This thesis provided many valuable insights for the integration of energy data into social science research and for understanding the potential uses of different kinds of consumer data to address societal problems. It is widely known that the understanding of the causal mechanisms of material deprivation has been improved by the inclusion of non-monetary indicators and is now widely considered to be a multidimensional phenomenon, characterised by a range of domains encompassing finance, health, education and crime amongst others and importantly, is a consequence of a lack of income *and other resources* (Payne and Abel, 2012). Given that the populace living with deprivation and fuel poverty are likely overlapping, and given that many of the drivers of deprivation also relate to fuel poverty, it is hoped that this analysis may encourage stakeholders to take a broader view in regard to understanding fuel poverty.

By proving the utility of demographic and energy data in the identification of at risk populations, this work begins to establish a new fuel poverty vernacular which diversifies the definition away from a purely income based metric and provides positive evidence of the contextual approach already applied to material deprivation having significant utility within the fuel poverty domain. As discussed in the literature review, current policy mandates for the defining and alleviation of fuel poverty are fragmented and feature an imprecise focus on low income households with an overarching obligation to tackle carbon mitigation, whilst also requiring that alleviation measures are both cost effective and targeted to the most vulnerable.

Succeeding this, the contextual approach proves the value of achieving affective data linkage from multiple data sources. At a high level, this generates substantial opportunity to incorporate new forms of data in support of existing population datasets to enable a more detailed representation of populations that those it is possible to generate from traditional data sources. This thesis has highlighted the multifaceted and multiscale nature of energy poverty; driving a necessity for new frameworks within which this societal issue can be examined and addressed. The representation of energy poverty requires insight from multiple different perspectives, which necessitate synthesis across different data to draw out a comprehensive understanding of this complex geography. This thesis argues strongly for the adoption of more comprehensive and multidimensional aggregate measures of energy poverty, and that addressing this issue through monotopical measures is limited. For example, whilst the Energy Performance Certificates are reasonably complete in terms of representing the general population, the data mostly relate to the fixtures and fittings of the household in which people live, and not of the people themselves. Whilst this allows us to understand the conditions in which certain demographics are more likely to live, it can only tell us so much about the lived experiences of the population and offer prospects for supplementing annualised fuel poverty statistics.

Notable applications relate mainly to data quality and representation challenges within social science research, but also pertain to the value of releasing data for academic research and the value in cohesive collaborations between academia and commercial entities. Historically there has been a fundamental lack of understanding in regard to consumer data more generally, due to the self-selection bias and the often commercially sensitive nature of the data. This work provides a framework for the analysis of nationally extensive smart meter data for future applications regarding the general population and provides ample evidence for the potential benefits of the incorporation of consumer big data into social science research. Novel elements of this work such as the cadence of the DEP smart meter data provide insight into the temporal usage patterns of consumers which were not possible with traditional data sources, as well as the treatment of gas and electricity consumption in tandem, allowing for a more comprehensive overview of temporal consumption trends than previously possible.

This work has made clear those inherent biases and necessary treatment required to extract insight from these commercial data. A related contribution was the development of effective data linkage across multiple energy data and other ancillary sources. Through the documentation of these, this thesis provides a point of reference for researchers, industry and policy makers, that ensures the aforementioned biases may be recognised and addressed in future work. It is clear that mechanisms which eliminate the risk of personal re-identification need to be in place as the prospect of increasingly reliable data linkage brings about further ethical and disclosure challenges.

It is hoped that this research underpins the utility of consumer data as an asset to social science as well as being a profitable commercial entity with this thesis acting as a catalyst for increased collaboration with organisations as they realise the benefits of allowing researchers with broader timescales and fewer commercial constraints controlled access to their datasets in order that insights aside from those which drive revenue be generated for public and social good. Under the condition of careful controls and recognised and respected collaborations such as the CDRC it is clear that such future endeavours could be hugely beneficial for both social science research and commercial data providers alike. Insights from the high level smart meter data exploration and the Energy User Classification demonstrate both the value in the validation of these big energy data and the importance of developing the frameworks which enable cohesive collaborative efforts between commercial entities and academic research by highlighting implications of optimised smart meter roll out targeting both from a commercial and social perspective. Given that energy policy is increasingly enacted through a public-private partnership and the onus of fuel poverty alleviation is now largely the responsibility of the energy providers, tools such as the classification developed in this work which optimise decision making are invaluable in ensuring the effectiveness of fuel poverty policy mandates.

Considering the practical applications of this work, it begins to identify areas where fuel poverty could potentially be alleviated to an extent through the installation of smart meters; it demonstrated how such analysis can identify areas where adoption rates are particularly low, despite being geographically close to areas where uptake is high (and so, roll-outs are in operation in those areas) as well as identifying low uptake areas with similar demographic characteristics to those of high uptake, where it is suggested that roll-out schemes would likely have the most success, thus aiding policy-makers understanding of the broader fuel poverty vernacular and driving tangible policy decision making.

7.3 Reflection on Methods

The majority of data used within this thesis are derived from non-traditional sources which are not privy to the same scientific approaches to data collection that are employed in traditional datasets (e.g. a survey). Working with such "transactional" data therefore required additional consideration to ensure that they were robust and fit for purpose. The complexity of addressing such issues are exacerbated within the context of arguably "big data": notably in this case, including those data derived from smart meter readings.

A first challenge was the manipulation of available measures between overlapping geographical boundaries, as first presented in Chapter 3, and then applied in Chapter 4. This population based dasymetric mapping methodology was deemed to best resolve the issue of different geographies, given its computational efficacy, and provided a better reflection of underlying population structure relative to more simple area based apportionment. For the purposes of end user utility, the target geography was selected as postcode sectors; however, on reflection, it is likely that transforming the geography of the smart meter dataset to be in line with the supplementary datasets (predominantly census based geography) may have returned more optimal results. However, despite this limitation the implemented dasymetric mapping approach was robust and effectively generated input for the Energy User Classification.

The use of clustering within Chapter 5 to explore the multidimensional characteristics of energy consumption as a more nuanced indicator of fuel poverty policy was argued as a positive step in illustrating the utility of "big" commercial data for policy discourse. Although Clustergrams were innovatively used to extract an optimal value of *k* in the k-means algorithm, classifications of this nature can be criticised for a lack of geographic sensitivity to local conditions. This might be argued as a valid critique given that no explicit spatial associations were encoded within this model, however, the approach taken is akin to many other standard geodemographic classification that have wide use and have been assured through successful application. Although there has been some work in this area (Alexiou, 2016), this is still far from conclusive as to whether such additions bring significant enhanced descriptive power to classifications. There is some evidence to suggest that in very different geographical settings this may prove more important (Longley and Singleton, 2014); which provides a direction that might be explored in future work. The validation of the classification created in Chapter 5 and presented and validated across Chapters 5 and 6 demonstrated both internal and external strength and through practical evaluation in Chapter 6 demonstrated utility for the energy sector within a policy driven decision making application.

Finally, much of the analysis presented and interpreted has been descriptive in nature; albeit framed within robust theoretical and applied framework. Such methods were necessary as valuable means of summarising complex interactions within the new big consumer datasets. The descriptive insights identified and presented have provided great insight into the multi-dimensional characteristics of energy poverty, and developed a useful tool for applications with commercial of policy objectives. The utility of such analysis in the communication of complex geographic patterns should not be understated, however this also provides a useful basis upon which future work might explore some of the causal relationships underlying these patterns.

7.4 Limitations

This thesis presents a broad range of insights, not only in terms of aggregate energy consumption, but also with regard to the characteristics that constrain or compel population groups in their usage. However, as with any analysis, this work has its limitations. In this instance, limitations are primarily as a result of the recognised uniqueness of the data, but which may also have been exacerbated by the relative infancy of the smart metering technology.

There are various areas where uncertainty was a factor. Firstly, as was addressed in Chapter 3, there were issues of data quality, which it was not always possible to cross-validate. In the majority of cases, findings have been cross-validated through triangulation with ancillary data sources, but there were cases where it was not possible, given the uniqueness of the data. Examples include entry errors such as invalid PCSs and processing errors which generated "0" readings in the smart meter data. Due to the uniqueness of this data, it is impossible to know whether they appear as a result of failure of the physical technology (or the technology being switched off if the household switch suppliers), a true zero where no consumption has taken place, a rounding error or a data processing error. In addition to this uncertainty, the total kWh at various temporal granularities presented results which appear to be below the national average consumption of even the lowest levels. Without further validation it is not possible to know if this is as a result of poor data quality, deliberate targeting of DEP customers who systematically under-consume to be the first to receive the new physical meters or because the transformation applied to the raw data in watts has resulted in an underestimation. There is no literature that the author knows of which elucidates on the methodology of calculating a kilowatt-half hour. Furthermore, because of the anonymisation of the dataset, it is not possible to know which, if any meters changed ownership during the timeframe of the dataset, which should be considered if future applications intended to utilise the individual level meter readings.

Speaking to the utility and limitations of the EPC data, chapter 3 recognised that the free text elements of the data capture process introduced uncertainty through entry and processing errors which cannot be cross-validated because of language barriers. A secondary limitation of the EPC data pertains to the modernity of its availability at the time of undertaking this work. EPC data was first made openly available in 2016, making this thesis one of the first piece of work to utilise it. This meaning that there was no pre-existing benchmark as to its accuracy. The volume of the EPC data meant that even with the data cleaning measures applied in order to account for error and attempt to eliminate it, when aggregated to postcode sector level, the data still presented a complete picture. However, it is important to acknowledge that proceeding this preliminary investigation into the veracity of the EPC data, a thorough evaluation of the quality of the EPC data was undertaken by Hardy and Glew (2019) which ultimately suggests that up to 27% of EPC records have at least one error, largely caused by the EPC assessor, as was suspected in this work. They recognise that many of the variables where the highest incidence of error occurs are those where the answer is left open to interpretation. They also note a geographical disparity in error rates, with higher error rates in the Greater London Authority than the rest of the country, which can be attributed to the increased number of flats (the property type with the highest propensity to contain an error). Access to this evaluation may have aided the data cleaning methodology and lead to a more thorough and informed data cleaning process, and should be considered for future applications of the EPC dataset within fuel poverty investigations both when making individual assessments and producing bulk statistics from the dataset as a whole. As they rightly state, the policy implications to an individual of having their home incorrectly categorised could lead to them qualifying or not for fuel poverty assistance. However, they also consider the entirety of the EPC dataset, and many errors they highlight relate to variables which were not considered pertinent in this work and they conclude by stating that some errors such as those pertaining to variables of floor, wall and insulation type have little to no impact on the final SAP rating that a property was given.

Leading on from this, another aspect of limitation is down to the relative infancy of both the technology and the dataset; the data provided by the DEP is only a fraction of the data which is collected and is from a period in time when smart metering technology was in its infancy. Whilst what they provided was a comprehensive and detailed source of data with temporal referencing, there are a number of areas where access to longitudinal comparison datasets may have provided valuable insights or opportunity for quantifying errors. For example, as previously noted, "0" consumption readings were difficult to corroborate – access to longitudinal data may have firstly made it possible to detect customers who were inactive for extended periods of time and could be deemed to have switched suppliers, but with time the quality and reliability of the physical apparatus also improved, which may have led to more consistent meter readings, making true zeros easier to recognise amongst the noise. Furthermore, not only did the technology improve, but even during the period of this dataset it was noted that active customer numbers did not remain static; and so access to an updated data extract from the DEP would provide enriched data over an longer timeframe, with an increased sample size. Additionally, an updated extract could help to shed light on the longer term impacts that having a smart meter can have on a consumer's ability or willingness to change their consumption habits and provide data driven insight into the utility of smart meters in decreasing consumption or cost for those in fuel poverty (see Chapter 4). Despite this, the dataset does provide a sample size much larger than previously available, and at a highly granular cadence.

Further uncertainty is apparent when considering issues of representativeness, which is a prominent consideration for consumer data which suffers from "self-selection bias". In chapter 3 efforts were made to understand the extent of this self-selection, through cross-validation with existing national statistics. The smart meter energy data were more heavily drawn from households in the north of England, where installations were much more prevalent. Issues such as this are unavoidable when working with data provided by one of the 'big six' domestic energy providers as historically, they tend to have a regional bias to their customer distribution. Such issues are exacerbated given that the roll-out programme is of unknown design (Chapter 3); however, as discussed earlier, are not uncommon when utilising secondary ancillary data for social enquiry. Furthermore, even within the DEPs customer base, there is a second layer of self-selection into those who do and do not have smart metering technology installed. As outlined in Chapters 2 and 4, smart meter adoption rates are affected by concerns over privacy, accessibility and opportunity, but the particular behavioural reasons for abstaining or not are well reported in the literature (Chapter 2) and substantiating these behavioural differences between up-takers and abstainers is not possible within the scope of the dataset. It is likely that qualitative data sources would need to be incorporated to understand the full extent of the resistance to smart metering technologies, if the aim was to understand the exhaustive extent of this element of self-selection. Furthermore, the infancy of the installation programme inherently limits the opportunity to participate; as discussed above, greater access to more longitudinal data may also have addressed some of the geographical disparities that are noted here; by allowing the installation programme to mature and more households be given the opportunity to receive a smart meter, the representation is likely to improve, once again inevitably enriching insights.

To this point, it is important to remember the most fundamental aspect of this bias on the overall population when you consider that only one of around 60 UK energy providers data is examined. Their customer base may be inherently skewed toward a certain demographic, but without competitor datasets for comparison, and using only common knowledge, it is not possible to know to what extent. Thus, the outputs in this thesis cannot be considered fully representative of the population.

A final, theoretical limitation of working with the smart meter dataset arises from the fact that much of the literature reviewed here is rightly disparaging of the use of technology as a suitable proposal for alleviating fuel poverty. It has not gone unnoticed that smart meters are an inherently technical solution, however, what they offer that other technical solutions such as increased insulation and the installation of solar panels do not is a bottom up approach which engages the user without the initial expense of the other methods. Smart meters offer immediate control over a household's consumption and increase the visibility of the cost of consumption in a tangible and accessible way. The other, more expensive methods rely on having the capital or being accepted into an energy efficiency scheme to make the changes and accept that the benefit of them is long term; something which many households living in fuel poverty cannot afford in either a financial or emotional sense. Smart meters are available to all, free to the user and require a minimal amount of disruption during installation.

With regard to geodemographics as a framework, there are methodological limitations imposed by subscribing large numbers of individuals to generalised profiles, potentially engendering ecological fallacy (see Chapter 2). This represents a commonly recognised issue when implementing classifications and could be criticised for its tendency to simplify trends as a factor of the modifiable areal unit problem and the effect of the scale at which the final classification was produced. But, as the aims of this work were to understand the patterns that we may be able to extract from energy data, an attempt had to be made to summarise the complexities in order to utilise the findings in a way that could be applied at a large geographical scale and was considered sufficient enough to quantify these representations.

Physical and environmental limitations also presented themselves in the early stages of the work; given the necessary security restrictions placed on users of the smart meter dataset, physical access to the data was limited to the opening hours of the University facilities due to its being held in a secure laboratory. These restrictions also dictate what can be extracted for use outside of the secure laboratory, and the process for extraction can be time consuming – it is not possible to receive feedback until the data is outside of the laboratory. Furthermore, it is a particularly difficult working environment with no Internet access, meaning that all code must be prepared without having seen the data, or has to be compiled from memory.

From a wider perspective and considering everything that has been discussed in both this and the previous section, these limitations only serve to highlight the need to support insights with evidence from alternative data sources. In summary, though this thesis champions the utility of consumer data as an indicator for social and spatial phenomena, it is an important consideration that the outputs and insights that can be derived will be necessarily limited by the scope of the available data.

7.5 Future Prospects and Closing Remarks

In developing a new framework that envelops a new and wider and contextual definition of energy poverty this provides great utility to explore patterns and define policy interventions to mitigate these issues. In doing so, the analysis of this thesis have necessitated consideration of "what is", rather than "what if". As such, it is argued that there is great potential to develop further insights from the presented work through more causal frameworks, that could further explore those drivers of the observed patterns.

Given the recency of the national smart meter roll out programme, and those generated data utilised by this thesis, there is a related challenge that extends this work through exploration of the longitudinal implications of having a smart meter. Notably this could extend the application of this theoretical and methodological framework to enable new understanding of the longer term impact of smart metering technologies on efficiency practises.

It is clear that those demographic groups with a propensity to being at home during the day due to unemployment or caring responsibilities are more likely to have been recipients of smart meters in the initial stages of the roll-out programme. Further work is now needed to understand the issue of smart meter inequity through policy which addresses the causal limitations which lead to low adoption rates as outlined in this thesis.

The constraints imposed on this work necessitated that its focus was at an aggregate geographical scale. Future endeavours may find utility in a more granular approach in order to understand how individual effects differ from area level findings to better understand the causal mechanisms of fuel poverty and implications of over and under consumption by accounting for unique and individual circumstances which are masked by the area level aggregations.

In closing, this thesis provides positive evidence for considering fuel poverty as a multifaceted societal phenomenon, which thereby allows energy policy stakeholders to re-evaluate the current understanding of the causal mechanisms behind it. This thesis provides clear evidence that demographic and efficiency indicators play an important role in the uncovering of populations not currently recognised as fuel poor by a monotopical monetary definition. This thesis supports and proves the utility of consumer data as a valuable tool in the social science realm, as well as highlighting the benefits of a collaborative relationship between researchers and commercial stakeholders to optimise effective decision making.

8 Bibliography

Abrahamse, W. & Steg, L. (2011) Factors Related to Household Energy Use and Intention to Reduce It: The Role of Psychological and Socio-Demographic Variables. *Human Ecology Review*, 18, 30–40. Society for Human Ecology.

Affordable Warmth Solutions (2020) *The Non-Gas Map*. Available at: https://www.nongasmap.org.uk

Alahakoon, D. & Yu, X. (2013) *Advanced analytics for harnessing the power of smart meter big data*.

Alexiou, A. (2016) *Putting Geo into Geodemographics: Evaluating the performance of national classification systems within regional contexts*. The University of Liverpool (United Kingdom).

Alexiou, A. & Singleton, A. (2018) *Indicators of Internet Use and Engagement*.

Anselin, L. (1995) Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27(2), 93–115.

Ardakanian, O. et al. (2014) *Computing Electricity Consumption Profiles from Household Smart Meter Data*.

Ástmarsson, B. et al. (2013) Sustainable renovation of residential buildings and the landlord/tenant dilemma. *Energy Policy*, 63, 355–362.

Barnett, M. (2015) *The £100m campaign you've never heard of*. Available at: https://www.marketingweek.com/the-100m-campaign-youve-never-heard-of/

BBC (2019) *Can my landlord lock my thermostat in a box?* Available at: https://www.bbc.co.uk/news/uk-50294404

Beatty, T.K.M. et al. (2014) Is there a 'heat-or-eat' trade-off in the UK? *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 177(1), 281–294.

BEIS (2017) *Smart Meters, Great Britain. Quarterly Report*.

BEIS (2019) 2018 UK greenhouse gas emissions, provisional figures. *National Statistics*.

Birchfield, V. & Duffield, J. (2011) *Toward a Common European Union Energy Policy: Problems, Progress, and Prospects*. Palgrave Macmillan US.

Bird, S. & Hernández, D. (2012) Policy options for the split incentive: Increasing energy efficiency for low-income renters. *Energy Policy*, 48, 506–514.

Boardman, B. (1991) Fuel Poverty is Different. *Policy Studies*, 12(4), 30–41.

Boardman, B. (2013) Fixing fuel poverty: Challenges and solutions. In: Taylor & Francis *Fixing Fuel Poverty: Challenges and Solutions*. [Online]. Taylor & Francis.

Boarini, R. & D'Ercole, M.M. (2006) Measures of material deprivation in OECD countries. *Social, Employment and Migration Working Papers*. OECD.

Bollier, D. & Firestone, C.. (2010) The Promise and Peril of Big Data. In: *Communications and Society Program*. [Online].

Bouzarovski, S. & Petrova, S. (2015) A global perspective on domestic energy deprivation: Overcoming the energy poverty-fuel poverty binary. *Energy Research and Social Science*, 10, 31–40.

Boyd, D. & Crawford, K. (2012) Six Provocations for Big Data. *SSRN Electronic Journal*.

Bradshaw, J. & Hutton, S. (1983) Social policy options and fuel poverty. *Journal of Economic Psychology*, 3(3–4), 249–266.

Bradshaw, M. (2013) Sustainability, Climate Change, and Transition in Global Energy. In: John Wiley & Sons Ltd Oxford, UK *The Handbook of Global Energy Policy*. John Wiley & Sons Ltd Oxford, UK.

Bridgeman, T. et al. (2018) *Tackling fuel poverty, reducing carbon emissions and keeping household bills down: tensions and synergies*.

Brounen, D. et al. (2012) Residential energy use and conservation: Economics and demographics. *European Economic Review*, 56(5), 931–945.

Brunsdon, C. & Singleton, A. (2015) *Geocomputation: A Practical Primer*. London: Sage.

Buchanan, K. et al. (2016) The British public's perception of the UK smart metering initiative: Threats and opportunities. *Energy Policy*, 91, 87–97.

Buchmann, E. et al. (2013) Re-identification of smart meter data. *Personal and ubiquitous computing*, 17(4), 653–662. Springer.

Carpenter, D.B. et al. (1955) Social Area Analysis: Theory, Illustrative Application and Computational Procedures. *American Sociological Review*, 20(4), 497.

Carroll, J. et al. (2014) Reducing household electricity demand through smart metering: The role of improved information about energy saving. *Energy Economics*, 45, 234–243.

Chen, M. et al. (2014) Big data: A survey. *Mobile Networks and Applications*, 19(2), 171–209. Springer.

Committee on Fuel Poverty (2018) *Commitee on Fuel Poverty Third Annual Report - 2018*.

Dalton, C.M. & Thatcher, J. (2015) Inflated granularity: Spatial "Big Data" and geodemographics. *Big Data and Society*, 2(2).

Darby, S. (2008) *Why, What, When, How, Where and Who? Developing UK Policy on Metering, Billing and Energy Display Devices*. In: 2008.

Darby, S. (2010) Smart metering: what potential for householder engagement? *Building Research & Information*, 38(5), 442–457. Taylor & Francis.

Darby, S. & Pisica, I. (2013) *Focus on electricity tariffs: experience and exploration of different charging schemes*. In: 2013.

Data Governance Institute (2014) *Data Governance Definition*. Available at: http://www.datagovernance.com/adg_data_governance_definition/ [Accessed: 16 August 2018].

Department for Business Energy and Industrial Strategy (2018) *Energy Consumption in the UK*.

Department for Business Energy and Industrial Strategy (2019) *Sub-Regional Fuel Poverty in England, 2019*.

Department of Energy and Climate Change (2013) *Fuel Poverty Report*.

Department of Energy and Climate Change (2014a) *Quantitative Research into Public Awareness, Attitudes, and Experience of Smart Meters: Wave 4*.

Department of Energy and Climate Change (2014b) *Smart meter roll-out for the domestic and small and medium non- domestic sectors (GB)*.

Department of Energy and Climate Change (2014c) *Understanding the behaviours of households in fuel poverty*.

Department of Energy and Climate Change (2015) *Cutting the cost of keeping warm: A fuel poverty strategy for England*.

Department of Energy and Climate Change (2017) Annual Fuel Poverty Statistics Report. *Department of Energy and Climate Change*.

Doane, D.P. & Seward, L.E. (2011) Measuring skewness: A forgotten statistic? *Journal of Statistics Education*, 19(2).

Drax (2016) *7 of the biggest TV moments in UK electricity history*. Available at: https://www.drax.com/technology/7-of-the-biggest-tv-moments-in-uk-electricity-history/ [Accessed: 2 June 2020].

Druckman, A. & Jackson, T. (2008) Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model. *Energy Policy*, 36(8), 3167–3182.

Dubash, N.K. & Florini, A. (2011) Mapping global energy governance. *Global Policy*, 2(1), 6–18.

Ehrhardt-Martinez, K. & John, A. (2010) *People-centered initiatives for increasing energy savings*.

Einav, L. & Levin, J. (2014) The Data Revolution and Economic Analysis 1. *Stanford Univerity and NBER*, 14(1), 1–24.

Ekins, P. et al. (2015) *Global energy: Issues, potentials, and policy implications*. Oxford University Press.

El-Hawary, M.E. (2014) The Smart Grid—State-of-the-art and future trends. *Electric Power Components and Systems*, 42(3–4), 239–250.

El-Pro-Cus (2019) *Overview of Smart Grid Technology And Its Operation and Application (For Existing Power System)*.

Elwood, S. & Leszczynski, A. (2011) Privacy, reconsidered: New representations, data practices, and the geoweb. *Geoforum*, 42(1), 6–15.

Energy Saving Trust (2017) *TOU tariffs - are energy bill savings all about timing?* Available at: http://www.energysavingtrust.org.uk/blog/tou-tariffs-are-energy-bill-savings-all-about-timing

Energy Saving Trust (2020) *Energy Performance Certificates*.

European Commission (2010) *Energy 2020 A strategy for competative, sustainable and secure energy*.

European Environment Agency (2020) *Trends and projections in Europe 2020*.

Eurostat (2019) *Energy consumption in households*. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_consumption_in_households#Use_of_energy_products_in_households_by_purpose

Faiers, A. et al. (2007) Towards a contemporary approach for understanding consumer behaviour in the context of domestic energy use. *Energy Policy*, 35(8), 4381–4390.

Faruqui, A. et al. (2010) The impact of informational feedback on energy consumption—A survey of the experimental evidence. *Energy*, 35(4), 1598–1608. Elsevier.

Fell, M.J. et al. (2014) Exploring perceived control in domestic electricity demand-side response. *Technology Analysis and Strategic Management*, 26(10), 1118–1130.

Field, A. et al. (2012) *Discovering Statistics Using R*. [Online]. SAGE Publications.

Firth, S. et al. (2008) Identifying trends in the use of domestic appliances from household electricity consumption measurements. *Energy and Buildings*, 40(5), 926–936.

Flowerdew, R. (2011) How serious is the modifiable areal unit problem for analysis of English census data? *Population trends*, 145(1), 102–114.

Fouquet, R. (2010) The slow search for solutions: Lessons from historical energy transitions by sector and service. *Energy Policy*, 38(11), 6586–6596.

Frederiks, E.R. et al. (2015) The socio-demographic and psychological predictors of residential energy consumption: A comprehensive review. *Energies*, 8(1), 573–609.

French, N. et al. (2018) *Minimum Energy Efficiency Standards (MEES) Impact on UK property management and valuation Title*.

Fuerst, F. et al. (2016) Energy performance ratings and house prices in Wales: An empirical study. *Energy Policy*, 92(1), 20–33.

G20 (2016) *G20 Energy Efficiency Leading Programme*.

Gale, C.G. et al. (2016) Creating the 2011 area classification for output areas (2011 OAC). *Journal of Spatial Information Science*, 12, 1–27.

Gandomi, A. & Haider, M. (2015) Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. Elsevier.

Gharavi, H. & Ghafurian, R. (2011) *Smart grid: The electric energy system of the future*. In: 2011.

Goldthau, A. (2016) *The handbook of global energy policy*. John Wiley & Sons.

Goldthau, A. & Witte, J.M. (2010) *Global energy governance: The new rules of the game*. Brookings Institution Press.

Goodchild, M.F. (2013) Prospects for a Space-Time GIS: Space-Time Integration in Geography and GIScience. *Annals of the Association of American Geographers*, 103(5), 1072–1077.

Government Statisical Services (2014) *GSS/GSR Disclosure control guidance for Microdata produced from social surveys*.

Graham, M. & Shelton, T. (2013) Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography*, 3(3), 255–261.

Grubler, A. (2012) Energy transitions research: Insights and cautionary tales. *Energy Policy*, 50, 8–16. Elsevier.

Guerreiro, S. et al. (2015) Making energy visible: sociopsychological aspects associated with the use of smart meters. *Energy Efficiency*, 8(6), 1149. Springer Science & Business Media.

Guo, J. et al. (2010) Impact of inter-sectoral trade on national and global CO2 emissions: An empirical analysis of China and US. *Energy Policy*, 38(3), 1389–1397.

Haben, S. et al. (2016) Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE transactions on smart grid*, 7(1), 136–144. IEEE.

Hardy, A. & Glew, D. (2019) An analysis of errors in the Energy Performance certificate database. *Energy Policy*.

Hargreaves, T. et al. (2010) Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors. *Energy policy*, 38(10), 6111–6119. Elsevier.

Harris, R. et al. (2005) *Geodemographics, GIS and Neighbourhood Targeting*. John Wiley & Sons.

Harris, R. & Jarvis, C. (2014) *Statistics for geography and environmental science*. Routledge.

Hausfather, Z. (2019) Analysis: Why the UK's CO2 emissions have fallen 38% since 1990. *Carbon Brief*.

Hazas, M. et al. (2011) Look back before leaping forward: Four decades of domestic energy inquiry. *IEEE Pervasive Computing*, 10(1), 13–19.

Hills, J. (2012) *Getting the measure of fuel poverty: Final Report of the Fuel Poverty Review*.

HM Government (2009) The UK Low Carbon Transition Plan: National strategy for climate and energy. In: *Department of Energy & Climate Change*. [Online].

Holloway, D. & Bunker, R. (2006) Planning, Housing and Energy Use: A Review. *Urban Policy and Research*, 24(1), 115–126.

Hope, A.J. & Booth, A. (2014) Attitudes and behaviours of private sector landlords towards the energy efficiency of tenanted homes. *Energy Policy*, 75, 369–378.

House of Commons Science and Technology Committee (2016) *Evidence check: smart metering of electricity and gas*.

Hu, S. et al. (2017) A survey on energy consumption and energy usage behavior of households and residential building in urban China. *Energy and Buildings*, 148, 366–378.

Information Commisioners Office (2018) *What is personal data?* Available at: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/

Jawurek, M. et al. (2011) *Plug-in privacy for smart metering billing*. In: Springer Verlag, 2011. Springer Verlag. Available at: doi:10.1007/978-3-642-22263-4_11

Jiawei, H. et al. (2012) *Data Mining: Concepts and Techniques*. [Online]. Elsevier.

Jonas, M. et al. (2019) Quantifying greenhouse gas emissions. *Mitigation and Adaptation Strategies for Global Change*, 24(6), 839–852.

Jones, R. V. & Lomas, K.J. (2015) Determinants of high electrical energy demand in UK homes: Socio-economic and dwelling characteristics. *Energy and Buildings*, 101, 24–34.

Kelling, S. et al. (2009) Data-intensive Science: A New Paradigm for Biodiversity Studies. *BioScience*, 59(7), 613–620.

Kitchin, R. (2014a) Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). SAGE Publications.

Kitchin, R. (2014b) *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Kwac, J. et al. (2013) *Utility customer segmentation based on smart meter data: Empirical study*. In: 2013.

Laney, D. (2001) 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*.

Langsdorf, S. (2011) EU Energy Policy : From the ECSC to the Energy Roadmap 2050. *Green European*

Foundation, 9.

Lansley, G. (2014) *Evaluating the utility of geo-referenced Twitter data as a source of reliable footfall insight*.

Laurila, J.K. et al. (2013) *From big smartphone data to worldwide research: The Mobile Data Challenge*. In: Elsevier B.V., 2013. Elsevier B.V.

Lavin, A. & Klabjan, D. (2015) Clustering time-series energy data from smart meters. *Energy Efficiency2*, 8(4), 681–689.

Lenormand, M. & Ramasco, J.J. (2016) Towards a Better Understanding of Cities Using Mobility Data. *Built Environment*, 42(3), 356–364. Alexandrine Press.

Levy, P.S. & Lemeshow, S. (2013) *Sampling of populations: methods and applications*. John Wiley & Sons.

Lewis, D. & Kerr, J. (2014) *Not too clever: will Smart Meters be the next Government IT disaster?* London.

Liddell, C. & Morris, C. (2010) Fuel poverty and human health: A review of recent evidence. *Energy Policy*, 38(6), 2987–2997.

Lisovich, M.A. et al. (2010) Inferring personal information from demand-response systems. *IEEE Security and Privacy*, 8(1), 11–20.

Logica (2007) *Turning Concern into Action; Energy Efficiency and the European Consumer*. Available at: http://www.logica.com/energy+efficiency+and+the+european+ consumer/400008607

Lomax, R.G. & Hahs-Vaughn, D.L. (2013) *An Introduction to Statistical Concepts*. Routledge.

Longhurst, N. & Hargreaves, T. (2019) Emotions and fuel poverty: The lived experience of social housing tenants in the United Kingdom. *Energy Research and Social Science*, 56.

Longley, P. et al. (2018) *Consumer Data Research*. UCL Press.

Longley, P.A. (2007) Some challenges to geodemographic analysis and their wider implications for the practice of GIScience. *Computers, Environment and Urban Systems*.

Longley, P.A. et al. (2015) The geotemporal demographics of twitter usage. *Environment and Planning A*, 47(2), 465–484.

Longley, P.A. (2017) Geodemographic Profiling. In: John Wiley & Sons *The International Encyclopedia of Geography*. John Wiley & Sons.

Longley, P.A. & Singleton, A.D. (2014) *London Output Area Classification Final Report*.

Lopes, M. et al. (2014) *Facilitating Energy Behaviours Transition To More Sustainable Patterns: Findings From A Case Study*.

Löschel, A. et al. (2010) Indicators of energy security in industrialised countries. *Energy Policy*, 38(4), 1665–1671.

Lynham, J. et al. (2016) Why does real-time information reduce energy consumption? *Energy Economics*, 54, 173–181.

Ma, Z. et al. (2017) The role of data analysis in the development of intelligent energy networks. *IEEE Network*, 31(5), 88–95. IEEE.

McKenna, E. et al. (2012) Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy*, 41, 807–814. Elsevier.

McLoughlin, F. et al. (2015) A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied energy*, 141, 190–199. Elsevier.

MHCLG (2018) *English Housing Survey 2017: stock condition*.

MHCLG (2019) *English Housing Survey. Headline Report, 2017-18*.

Middlemiss, L. (2017) A critical analysis of the new politics of fuel poverty in England. *Critical Social Policy*, 37(3), 425–443.

Middlemiss, L. et al. (2018) Plugging the gap between energy policy and the lived experience of energy poverty: Five principles for a multidisciplinary approach. In: *Advancing Energy Policy: Lessons on the Integration of Social Sciences and Humanities*. [Online].

Middlemiss, L. et al. (2019) Energy poverty and social relations: A capabilities approach. *Energy Research and Social Science*, 55, 227–235.

Middlemiss, L. & Gillard, R. (2015) Fuel poverty from the bottom-up: Characterising household energy vulnerability through the lived experience of the fuel poor. *Energy Research and Social Science*, 6, 146–154.

Miller, H.J. (2010) The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science*, 50(1), 181–201.

Miller, H.J. & Goodchild, M.F. (2015) Data-driven geography. *GeoJournal*, 80(4), 449–461. Springer.

Moore, R. (2012) Definitions of fuel poverty: Implications for policy. *Energy Policy*, 49, 19–26.

Morabito, V. (2015) Big data governance. In: Springer *Big data and analytics*. Springer.

Mould, R. & Baker, K.J. (2017) Documenting fuel poverty from the householders' perspective. *Energy Research and Social Science*, 31, 21–31.

National Right to Fuel Campaign (2013) *Right2FuelUK Home*. Available at: https://www.right2fueluk.com

Nejat, P. et al. (2015) A global review of energy consumption, CO2emissions and policy in the residential sector (with an overview of the top ten CO2emitting countries). *Renewable and Sustainable Energy Reviews*, 43, 843–862.

Nicol, S. et al. (2016) The cost of poor housing in the European Union. *University of Warwick. United Kingdom*.

OECD (2014) Society at a Glance 2014: OECD Social Indicators. In: *Society at a Glance*. [Online].

OECD & The International Energy Agency (2012) OECD Green Growth Studies: Energy. *Director*.

Office for National Statistics (2016a) *2011 rural/urban classification*. Available at: https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification [Accessed: 12 April 2018].

Office for National Statistics (2016b) *A Guide to ONS Postcode Products*.

Office for National Statistics (2016c) *Postcode Geography*. Available at: https://www.ons.gov.uk/methodology/geography/ukgeographies/postalgeography [Accessed: 8 February 2021].

Office for National Statistics (2019a) *National Statistics Postcode Lookup (May 2019)*.

Office for National Statistics (2019b) *ONS Postcode Directory (May 2019).*

Ofgem (2009) *Ofgem Consumer First Panel - Billing Information and Price Metrics*.

Ofgem (2018) *Feed in Tariffs*.

Ofgem (2019a) *About Us*. Available at: https://www.ofgem.gov.uk/about-us/our-priorities-and-objectives [Accessed: 4 February 2020].

Ofgem (2019b) *Electricity supply market shares by company: Domestic (GB) | Ofgem*. Available at: https://www.ofgem.gov.uk/data-portal/electricity-supply-market-shares-company-domestic-gb [Accessed: 1 May 2019].

Ofgem (2021) *Smart Meter Rollout: Open letter on Energy Suppliers' Delivery of the Rollout and Regulatory Obligations*.

Oltra, C. et al. (2013) A qualitative study of users' engagement with real-time feedback from in-house energy consumption displays. *Energy Policy*, 61, 788–792.

Openshaw, S. et al. (1980) A critique of the national classifications of OPCS/PRAG. *Town Planning Review*, 51(4), 421–439.

Openshaw, S. (1984) Ecological fallacies and the analysis of areal census data (UK, Italy). *Environment & Planning A*, 16(1), 17–31.

Pachauri, S. & Spreng, D. (2003) Energy use and energy access in relation to poverty. *Economic and Political Weekly*, 25, 1–15.

Park, R.E. & Burgess, E.W. (1925) The City: Suggestions for Investigation of Human Behavior in the Urban Environment. In: *Social Forces*. [Online].

Payne, R. a & Abel, G. a (2012) UK indices of multiple deprivation - a way to make comparisons across constituent countries easier. *Health statistics quarterly / Office for National Statistics*.

Petrova, S. (2018) Encountering energy precarity: Geographies of fuel poverty among young adults in the UK. *Transactions of the Institute of British Geographers*, 43(1), 17–30.

Polonetsky, O.T.& J. (2012) Privacy in the Age of Big Data: A Time for Big Decisions. *Stanford Law Review Online*, 3(2), 19–28.

Practical Computing Applications (2017) *Cluster Analysis*.

Ramos, S. & Vale, Z. (2008) *Data mining techniques to support the classification of MV electricity customers*. In: 2008.

Rhodes, J.D. et al. (2014) Clustering analysis of residential electricity demand profiles. *Applied Energy*, 135, 461–471.

Riddlesden, D. (2016) *Expanding the Big E-Society: A Contemporary Model of Internet Access, Exclusion and Investment*. University of Liverpool.

Roberts, D. et al. (2015) Fuel poverty in the UK: Is there a difference between rural and urban areas? *Energy Policy*, 87, 216–223.

Roberts, S. & Redgrove, Z. (2011) *The smart metering programme: a consumer review*.

Robinson, C. et al. (2018a) 'Getting the measure of fuel poverty': The geography of fuel poverty indicators in England. *Energy Research and Social Science*, 36, 79–93.

Robinson, C. et al. (2018b) Underrepresenting neighbourhood vulnerabilities? The measurement of fuel poverty in England. *Environment and Planning A*, 50(5), 1109–1127.

Rosenow, J. et al. (2013) Fuel poverty and energy efficiency obligations - A critical assessment of the supplier obligation in the UK. *Energy Policy*, 62, 1194–1203.

Royston, S. & Guertler, P. (2013) Fact-file: Families and fuel poverty. *London Association for the Conservation of Energy*.

Sanchez Nicolas, E. (2020) *Why is EU off track for 2020 energy efficiency target?*

Schipper, L. et al. (1989) Linking Life-Styles And Energy Use: A Matter Of Time. *Annual Review of Energy and the Environment*, 14(1), 273–320.

Schonlau, M. (2004) Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams. *Computational Statistics*, 19(1), 95–111.

Simcock, N. et al. (2016) Fuel poverty in the UK: beyond heating? *People Place and Policy Online*, 10(1), 25–41.

Singleton, A.D. & Longley, P.A. (2009) Geodemographics, visualisation, and social networks in applied geography. *Applied Geography*, 29(3), 289–298.

Singleton, A.D. & Spielman, S.E. (2014) The past, present, and future of geodemographic research in the United States and United Kingdom. *The Professional Geographer*, 66(4), 558–567. Taylor & Francis.

Sleight, P. (2007) An introductory review of geodemographic information systems. *Journal of Targeting, Measurement and Analysis for Marketing*, 12(4), 379–388.

Smart Energy GB (2014) *Consumer attitudes to the energy market and smart meters. Wave 1.*

Smart Energy GB (2018a) *2020 Energy Strategy*. Available at: https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union/2020-energy-strategy

Smart Energy GB (2018b) *Britain's Smart Grid*. Available at: https://www.smartenergygb.org/en/smart-future/britains-smart-grid [Accessed: 19 April 2018].

Smil, V. (2010) *Energy transitions: history, requirements, prospects*. ABC-CLIO.

Soares, S. (2012) Big data governance. *Information Asset, LLC.*

Soares, S. (2013) *IBM InfoSphere: A platform for Big Data governance and process data governance*. MC Press.

Sovacool, B.K. (2015) Fuel poverty, affordability, and energy justice in England: Policy insights from the Warm Front Program. *Energy*, 93, 361–371.

Stern, S.M.C.-K. (2011) Smart-Grid: Technology and the Psychology of Environmental Behavior Change. *Law Review*, 86(1), 139.

Swan, W. et al. (2013) Adoption of sustainable retrofit in UK social housing. *Structural Survey*.

The Committee on Climate Change (2019) *Reducing the UKs Emissions*. Available at: https://www.theccc.org.uk/our-impact/reducing-the-uks-emissions/ [Accessed: 3 December 2019].

The International Energy Agency (2017) Key World Energy Statistics 2017. In: *Key World Energy Statistics 2017*. [Online].

Thomson, H. & Snell, C. (2013) Quantifying the prevalence of fuel poverty across the European Union. *Energy Policy*, 52, 563–572.

Tobler, W.R. (1970) A Computer Movie Simulation Urban Growth in Detroit Region. *Economic Geography*, 46(1), 234–240.

UK Committee on Climate Change (2010) *Fourth Carbon Budget*. London.

UK Parliament (2000) *Warm homes and energy conservation act*. London.

Ushakova, A. et al. (2018) Interpreting Smart Meter Data of UK Domestic Energy Consumers. In: *Consumer Data Research*.

USwitch (2017) *White Meters Explained*. Available at: https://www.uswitch.com/gas-electricity/guides/white-meters/

USwitch (2018) *Gas and electricity bills: What is average usage and what type of energy user are you?* Available at: https://www.uswitch.com/gas-electricity/guides/gas-and-electricity-user-groups/ [Accessed: 10 August 2018].

Vickers, D. et al. (2005) Creating the national classification of census output areas: Data, methods and results. *Working Paper of the University of Leeds, School of Geography*.

Vickers, D. & Rees, P. (2007) Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society. Series A: Statistics in Society*.

Viegas, J.L. et al. (2016) Classification of new electricity customers based on surveys and smart metering data. *Energy*, 107, 804–817. Elsevier.

Voas, D. & Williamson, P. (2001) The diversity of diversity: A critique of geodemographic classification. *Area*, 33(1), 63–76.

Wang, S. et al. (2005) *Attribute Uncertainty in GIS Data*. In: Springer Berlin/Heidelberg, 2005. Springer Berlin/Heidelberg.

Wang, S. et al. (2016) The relationship between economic growth, energy consumption, and CO2 emissions: Empirical evidence from China. *Science of the Total Environment*, 542, 360–371.

Watson, D. & Maitre, B. (2015) Is fuel poverty in Ireland a distinct type of deprivation? *Economic and Social Review*, 46(2), 267–291.

Webber, R.. (1978) Making the most of the census for strategic analysis. *The Town Planning Review*, 49(3), 274–284.

Webber, R.J. (1980) A response to the critique of the OPCS/PRAG national classifications (UK). *Town Planning Review*, 51(4), 440–450.

Webber, R.J. et al. (2015) Adoption of geodemographic and ethno-cultural taxonomies for analysing Big Data. *Big Data and Society*, 2(1), 20–53.

Weiss, M. et al. (2012) *Leveraging smart meter data to recognize home appliances*. In: 2012.

Wilkinson, P. et al. (2007) A global perspective on energy: health effects and injustices. *Lancet*, 379(9591), 965–978.

Winzer, C. (2012) Conceptualizing energy security. *Energy Policy*, 46, 36–48.

Wyatt, P. (2013) A dwelling-level investigation into the physical and socio-economic drivers of domestic energy consumption in England. *Energy Policy*, 60, 540–549.

Yohanis, Y.G. et al. (2008) Real-life energy use in the UK: How occupancy and dwelling characteristics

affect domestic electricity use. *Energy and Buildings*, 40(6), 1053–1059.
Zarsky, T.Z. (2017) Incompatible: The GDPR in the Age of Big Data. *Seton Hall Law Review*, 47, 995.

9 Appendices

9.1 Ethics Approval

From: **Ethics** ethics@liverpool.ac.uk
Subject: RE: Ethics Requirement query
Date: 27 February 2019 at 08:43
To: Talbot, Ellen sgetalbo@liverpool.ac.uk

E

Hello Ellen

Apologise for the late reply.

From what you have described you will not need ethical approval as the data will be fully anonymised

Best wishes
Fran

Frances Thomason
Research Ethics & Integrity Officer
Research Support Office
University of Liverpool
2nd Floor Block C
Waterhouse Building

Tel: 0151 795 7666
Email: F.I.Thomason@liverpool.ac.uk

-----Original Message-----
From: Talbot, Ellen <sgetalbo@liverpool.ac.uk>
Sent: 19 February 2019 16:14
To: Ethics <ethics@liverpool.ac.uk>
Subject: Ethics Requirement query

Good afternoon,

I am writing to confirm whether or not my research requires ethics approval. The data is in its full form individual records of household energy consumption but is anonymised. It is stored on the secure data server held at CSD. To remove anything from this server you must pass a safe researcher check and data must be aggregated. My thesis will contain only non-disclosive visualisations (mainly maps) and will not identify any individual.

From what I can read online, it doesn't sound as though I do require ethics approval, but if this is the case it would be helpful to have confirmation in writing.

Many thanks.

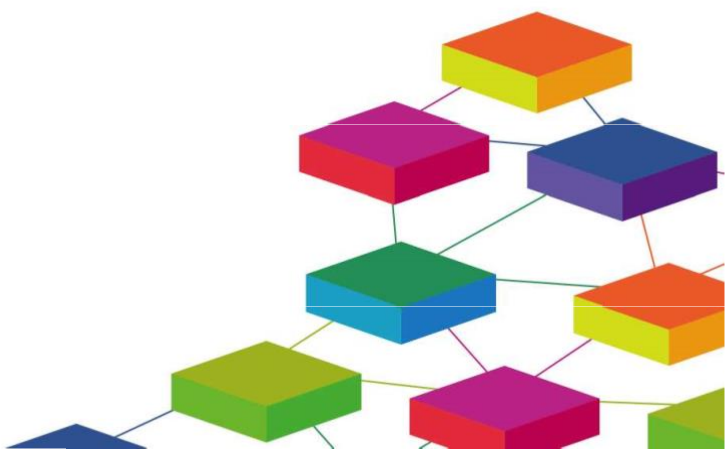
Ellen Talbot

(201155605)

9.2 CDRC User Guide

CONSUMER DATA RESEARCH CENTRE DATA SERVICE
USER GUIDE

Version: 6.0



Consumer Data Research Centre

An ESRC Data Investment

Introduction

The Consumer Data Research Centre (CDRC or Centre) is an academic led, multi-institution
laboratory which discovers, mines, analyses and synthesises consumer-related datasets from around
the UK. The CDRC forms part of the ESRC-funded Big Data network and offers a data service aimed

at providing researchers with access to a wide range of consumer data to address many societal challenges. CDRC's key areas of interest include retail, transport, health, crime, housing, energy,
mobility and sustainable consumption. We support the acquisition and analysis of data in these areas and others to achieve benefits for the 'public good'.
The purpose of this guide is to describe the Centre's data services and how researchers can access them. It identifies the different types of data the Centre holds and the service tiers through which
these data sets are available. For data that are not publicly available, the guide details how researchers can register or apply for access and the kinds of support that is available to them.

CDRC Data Services

The CDRC provides data with three different levels of access. These correspond to the data levels described in the UK Data Service's three tier access policy:

- • Open data: data which are freely available to all for any purpose. Data includes open datasets where CDRC have added value and non-sensitive and aggregated data and derivative products produced by the CDRC. Examples might include geodemographic data derived from the Census. Open data are accessed through the CDRC service via basic registration and download.
- • Safeguarded data: data to which access is restricted due to licence conditions, but where data are not considered 'personally-identifiable' or otherwise sensitive – an example might include data from retail companies on store turnover. Access to safeguarded CDRC data is via a remote service that requires users to submit a project proposal. This proposal must receive approval from the Centre's Research Approvals Group (RAG) (see below) before access to the data will be authorised. Users are able to retrieve data after authentication and authorisation by the service.
- • Controlled data: data which need to be held under the most secure conditions with more stringent access restrictions, including data which are 'personally-identifiable' and therefore subject to Data Protection legislation or are considered commercially sensitive. Examples might include data on individual consumer purchases. Access to CDRC controlled data is provided through the CDRC-secure service. This service requires that individuals gain project approval through the RAG and visit one of our secure facilities at either the University College London, University of Leeds or University of Liverpool.

  Finding Data

  All data available through the CDRC are accompanied by metadata that enable both attributes and geography to be searched.

Research Approvals Process

Access to both safeguarded and controlled data requires a process by which individuals submit project proposals for assessment and approval. The approval process is overseen by an independent Research Approvals Group (RAG) which comprises representation from the Data Partner(s) and the social science academic community. The Group may also draw upon the expertise from a social science ethics practitioner. The CDRC Senior Management Team provides comment on resource implications of a proposal. The composition ensures that the RAG has expertise in research design, analysis and impact, while also considering any commercial sensitivities a project may have. The RAG review process is overseen by the Chair of RAG.

For full details of the Research Approvals Process please see the Research Approvals Guidelines at www.cdrc.ac.uk/data-services/using-our-data/.

Criteria for Approval

These criteria align with CDRC objectives and cover the following:

- • Scientific advancement – how the project has the potential to advance scientific knowledge, understanding and/or methods using consumer data;
- • Public good – how the project has the potential to provide insight and/or solutions that could benefit society;
- • Privacy and ethics – the potential privacy impacts or risks, and wider ethical considerations relating to the project
- • Project Design and Methods – how the project will be conducted and who will be involved with a focus on demonstrating project feasibility.
- • Cost and resources issues – what impact the project is likely to have on CDRC resources, including CDRC staff time and use of infrastructure, as well as any data acquisition costs. Resource requirements should be justified.

  The RAG typically considers applications remotely and is designed to be lightweight but robust, enabling timely decisions on user applications.

  Approval will not be granted without evidence that the user has acquired ethical approval for the research through their institution, or supplied evidence that it is not applicable. For non-academic projects, where there is no approval process in place the CDRC will assist the user with acquiring this.

Safe Researcher Training and Training and Development

Safe Research Training

Users, both academic and non-academic stakeholders, wishing to access controlled data and on occasion safeguarded data are required to have completed a safe researcher course, as offered by the Administrative Data Research Network (ADRN), HM Revenue and Customs (HMRC), Office for National Statistics (ONS) or the UK Data Service (UKDS). Evidence of valid accreditation for the duration of access to the data will be required. If the user has not previously completed such training the CDRC will offer access to training courses.

Training and Development

In addition to providing data services, the CDRC has a range of training courses and materials available. Many of these will be of benefit to those who wish to use our facilities, as they are aimed at enhancing capacity in data analytics and data visualisation methods. Full details of the training available can be found at cdrc.ac.uk/training-capacity-building/ and online training tutorials at data.cdrc.ac.uk/tutorial. Our programme includes training in the following areas:

- • Working on Big Data: introductory courses that explain the growing importance of Big Data; the importance of analytics and protocols; and standards for data management.
- • Introductory and advanced courses in data analysis and visualisation, including courses in R.
- • Introductory and advanced courses in Geographical Information Systems, including ArcGIS and Q-GIS.
- • Advanced courses in microsimulation and geo-temporal demographics.
- • Courses on how insights from Big Data analytics can enhance business.
- • Visualisation.

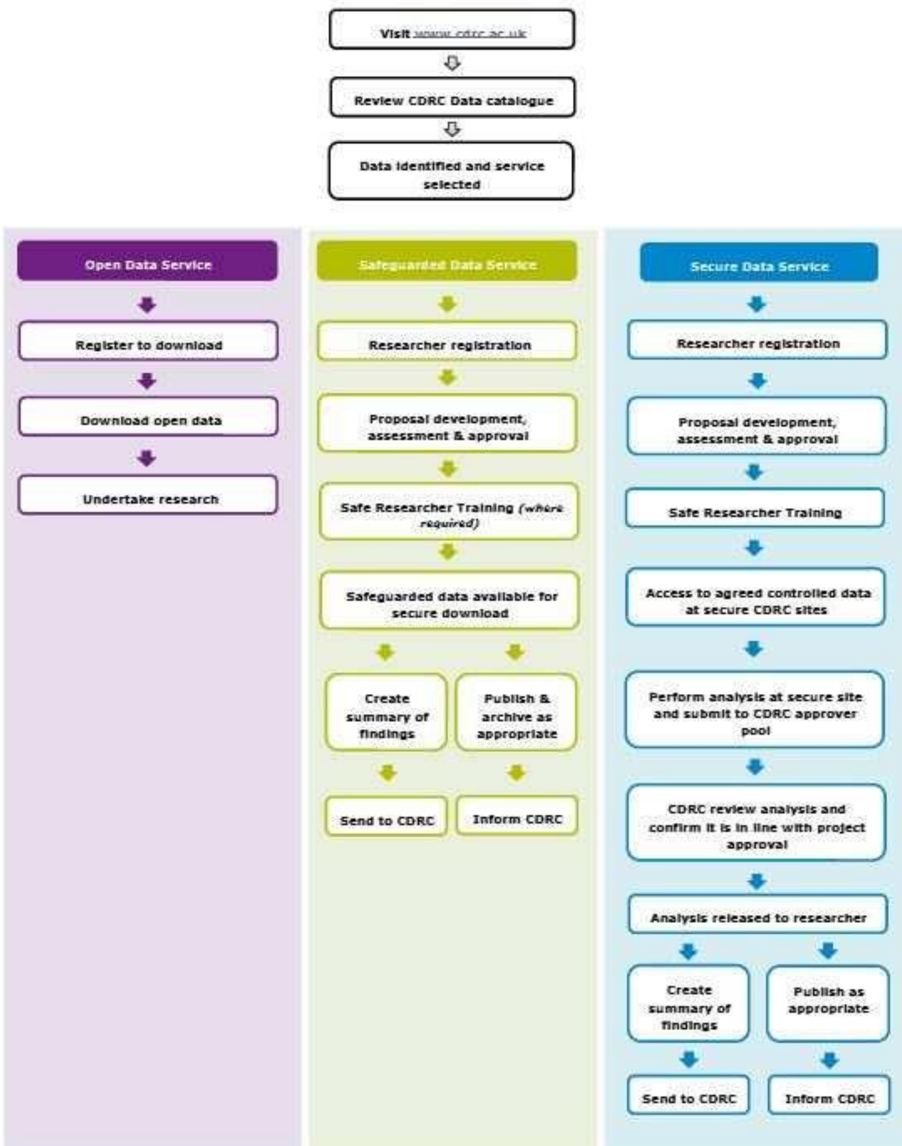
Charges for CDRC Services

While a service will be provided to the academic community and stakeholders free of charge, researchers may need to apply for funding to cover the costs of additional data acquisitions, or be charged for access to certain, licensed software.

Consumer Data Research Centre

An ESRC Data Investment

CDRC Services Overview and User Journey



E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL | UCL | UNIVERSITY OF LIVERPOOL | UNIVERSITY OF OXFORD | UNIVERSITY OF LEEDS

Consumer Data Research Centre

An ESRC Data Investment

CDRC Website: A Single Point of Entry into the CDRC Data Services

The CDRC website, www.cdrc.ac.uk, is designed to provide a single point of entry into our services and these are clearly linked from the homepage.

CDRC Data

Our data portal, CDRC Data, provides a complete listing of data available through the three tiers of the service and enables the dissemination of open data and application for access to safeguarded and controlled data.

Accessing data from CDRC Data data.cdrc.ac.uk Open Service:

Access to the Open Service requires:

1) Registration

Users will be required to provide contact details including a valid email address prior to download. This is to enable the CDRC to monitor the use of the resource. Data will then be available to the user to download for unrestricted use.

Safeguarded Service:

Access to the Safeguarded Service requires that users to obtain formal approval.

1) Initial Proposal

An approach is made to the CDRC by the user through completion of an online form, www.cdrc.ac.uk/data-services/using-our-data/. This initial proposal is processed and assessed by the Senior Management Team to see if it fits within the remit of the Centre. If not, the proposal may be referred to another Centre in the Big Data Network. Proposals that do not fit into either of these categories will be turned down at this stage.

2) Proposal Development

If the initial proposal fits within the Centre's remit, the user is supplied with the 'Safeguarded Data Project Proposal Form', and assigned to a CDRC data scientist who can advise on the technical aspects of the formal application. The aim is to co-produce an acceptable project proposal. Proposals will comprise:

1. a) Research motivation and purpose
2. b) Research impact
3. c) Plannedoutputs

4. d) Research team
5. e) Data requested
6. f) Data linkage
7. g) Duration of access

Consumer Data Research Centre

An ESRC Data Investment

h) Ethical approval from user's institution[1]

3) RAG Assessment and Approval

Once an application has been completed it is considered by the RAG against agreed criteria that are published on our website, www.cdrc.ac.uk/data-services/using-our-data/. The number of rejected approvals will be minimised through initial interaction with the data scientists. Where approval is withheld, applications are referred back to the user for revision, and clear guidance will be given regarding those areas requiring clarity. If such amendments are agreeable by RAG, approval will be given. If minor, the user may be asked to make further revisions, however, if issues are still considered to be major the RAG may decide to make a final decision to reject the proposal. Following approval, the user and their institution are required to agree to the CDRC User Agreement, including stipulations made by the Data Partner(s) and RAG.

4) Data Access

Access to a secure download of the agreed data is made available. This process requires that users telephone the CDRC to obtain a further password to unlock the encrypted download files. Once the user has downloaded the encrypted file, they are solely responsible for the data and its analysis.

5) Outputs

Users can use results of their analyses in publications, reports and presentations provided they abide by the terms and conditions with particular reference to the data partner publication terms. There is no screening of outputs by CDRC staff.

6) Completion, Reporting and Acknowledgement

Users are required to deposit copies of working papers, peer-reviewed journal articles, logs of impact and other publications for access with the CDRC site wherever copyright permits. Where this is not possible, full references to research outputs are required for CDRC audit purposes. Please email publications@cdrc.ac.uk when publications are ready for deposit or logging. The commitment to produce specified outputs is normally a condition of the data approval process. The terms of service require that published outputs include an acknowledgement stating: “The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC xxx, ES/L011840/1; ES/L011891/1”. The acknowledgement will make further reference to the use of specific datasets according to the wishes and needs of individual data partners. After the project end date is reached, the CDRC will contact the user to confirm the destruction of the data and to document any outputs to date. The CDRC will contact users normally at 6 and 12 months after the project end date to request a log of any further publications or impact logs.

[1] If the user’s institution does not have a system for data protection and ethics approval then the CDRC will assist with gaining ethical review if required.

7) Undergraduate and Postgraduate Student Applications

Undergraduate and Masters Students requesting access to data will be required to submit a proposal in the normal way including their academic supervisor as a named applicant.

CDRC Secure Service

Access to CDRC controlled data is via our Secure Service at one of three secure facilities located at University College London, the University of Liverpool and the University of Leeds. Independent analysis of secure data can be undertaken at all of our secure facilities. If users require bespoke guidance and support with analytics, this service is provided at the University of Leeds only.

Use of the CDRC-Secure service requires registration and project approval, with an additional step of booking into one of the secure facilities and meeting any site specific secure facility requirements. The user will be informed of these once the site to be visited has been selected.

Accessing data from CDRC secure sites

Access to this service requires that users obtain formal approval.

1) Initial Proposal

An approach is made to the CDRC by the user through completion of an online form, www.cdrc.ac.uk/data-services/using-our-data/. This initial proposal is processed and assessed by the Senior Management Team to see if it fits within the remit of the Centre. If not, the proposal may be referred to another Centre in the Big Data Network. Proposals that do not fit into either of these categories will be turned down at this stage.

2) Proposal Development

If the initial proposal fits within the Centre's remit, the user is supplied with the 'Controlled Data Project Proposal Form', and assigned to a CDRC data scientist who can advise on the technical aspects of the formal application. The aim is to co-produce an acceptable project proposal. Proposals will comprise:

1. a) Research motivation and purpose
2. b) Research impact
3. c) Plannedoutputs
4. d) Research team
5. e) Data requested
6. f) Data linkage
7. g) Access requirements
8. h) Ethical approval from user's institution[2]

3) RAG Assessment and Approval

[2] If the user's institution does not have a system for data protection and ethics approval then the CDRC will assist with gaining ethical review if required.

Once an application has been co-produced it is considered by the RAG against agreed criteria that are published on our website www.cdrc.ac.uk/data-services/using-our-data/. The number of rejected approvals will be minimised through initial interaction with the data scientists. Where approval is withheld, applications are referred back to the user for revision, and clear guidance will be given regarding those areas requiring clarity. If such amendments are agreeable by RAG, approval will be given. If minor the user may be asked to make further revisions, however if issues are still considered to be major the RAG may decide to make a final decision to reject the proposal. Following approval, the user and their institution are required to agree to the CDRC User Agreement, including stipulations made by the Data Partner(s) and RAG.

4) Data Access

Following approval, the allocated CDRC data scientist arranges access for the registered user. Dates are booked to use the secure facility at either UCL, University of Liverpool or University of Leeds. Users will receive a document informing them of site specific secure facility requirements and instructions of use.

5) Data Analysis

The user works on the data only within the secure environment. If users wish to combine controlled data with other less sensitive data (open or safeguarded), then it will be necessary to have obtained consent for this from RAG as part of the project proposal. This supporting data will then be made available to the user in the secure facility. The same applies to software required for analysis. CDRC staff provide limited support through the advanced analytics service. At the University of Leeds, a supported analytics service is available which provides the user with bespoke guidance and support in both accessing and analysing data.

6) Outputs

All outputs that the user wants to take out of the secure environment must be vetted and cleared by the CDRC before they can be released. Source data do not leave the secure facility. Users can take results of their analyses for use in publications, reports and presentations provided they abide by the terms of the User Agreement and with particular reference to the data partner publication terms.

After completion of analysis the user informs the data scientist that the analysis is complete and that their files are now ready for vetting. For full details of the output process please see the CDRC site specific 'Secure Lab Data Import/Export Procedures'.

a) Outputs will be checked by two CDRC data scientists to ensure that they conform to CDRC control criteria.

i. Outputs requested should be 'finished outputs' i.e. the finished statistical analyses that you intend to present to the public, must be easy to read and interpret and how they are to be used explained and must be non-disclosive.
ii. The CDRC team will ensure that the outputs are the same specification as those agreed in the approved project proposal.

iii. The user is informed of the outputs vetting outcome within 5 working days and if successful with details about how the data extracts or analysis will be returned to them.

iv. Extracts that match approval are transferred by the CDRC team to a secure server from where outputs can be downloaded under the same arrangements as safeguarded data or transferred to the user on an encrypted USB/hard drive.
v. Whereextractsaredeemednottomatchtherequiredcriteria,theuserisinformed.
    i. Where there are issues with a part of the output, if feasible the user will be allowed to revisit the secure facility to rectify the problem.
    ii. Major transgressions may be permanently deleted and the remaining output is returned to the CDRC approver pool.
vi. Once the user has completed all their analysis or their agreed lab access time has been reached all passes or electronic fobs are returned and access to the secure facility is immediately revoked.

7) Completion, Reporting and Acknowledgement

Users are required to deposit copies of working papers, peer-reviewed journal articles, logs of impact and other publications for access with the CDRC site wherever copyright permits. Where this is not possible, full references to research outputs are required for CDRC audit purposes. Please email publications@cdrc.ac.uk when publications are ready for deposit or logging. The commitment to produce specified outputs is normally a condition of the data approval process. The terms of service require that published outputs include an acknowledgement stating "The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC xxx, ES/L011840/1; ES/L011891/1". The acknowledgement will make further reference to the use of specific datasets according to the wishes and needs of individual data partners. After the project end date is reached, the CDRC will contact the user to confirm the destruction of the data and to document any outputs to date. The CDRC will contact users normally at 6 and 12 months after the project end date to request a log of any further publications or impact logs.

8) Undergraduate and Postgraduate Student Applications

Undergraduate and Masters Students requesting access to data will be required to submit a proposal in the normal way including their academic supervisor as a named applicant.

9) Request for data not currently available through CDRC

It is possible to request access to data variables or datasets not currently available through the CDRC. To submit a request please complete an initial proposal form cdrc.ac.uk/data-services/initial-proposal-form/ and we will contact you to discuss further.

9.3 EPC Data Glossary Extract

POSTCODE `POSTCODE`

The postcode of the property

CURRENT ENERGY RATING `CURRENT_ENERGY_RATING`

Current energy rating converted into a linear 'A to G' rating (where A is the most energy efficient and G is the least energy efficient)

CURRENT ENERGY EFFICIENCY `CURRENT_ENERGY_EFFICIENCY`

Based on cost of energy, i.e. energy required for space heating, water heating and lighting [in kWh/year] multiplied by fuel costs. (£/m2/year where cost is derived from kWh).

PROPERTY TYPE `PROPERTY_TYPE`

Describes the type of property such as House, Flat, Mansion, Maisonette etc. This is actually the type differentiator for Property but only a limited number of property types, notably Apartment and Apartment Block, have any specific characteristics and warrant their own definition.

BUILT FORM `BUILT_FORM`

The building type of the Property e.g. Detached, Semi-Detached, Terrace etc. Together with the Property Type, the Build Form produces a structured description of the property

TRANSACTION TYPE `TRANSACTION_TYPE`

Type of transaction that triggered EPC. For example, one of: marketed sale; non- marketed sale; rental; not sale or rental; assessment for Green Deal; following Green Deal; FIT application; none of the above; RHI application; ECO assessment. Where the reason for the assessment is unknown by the energy assessor the transaction type will be recorded as 'none of the above'. Transaction types may be changed over time.

ENVIRONMENT IMPACT CURRENT

`ENVIRONMENT_IMPACT_CURRENT`

The Environmental Impact Rating. A measure of the property's current impact on the environment in terms of carbon dioxide (CO_2) emissions. The higher the rating the lower the CO_2 emissions. (CO_2 emissions in tonnes / year)

ENERGY CONSUMPTION CURRENT

`ENERGY_CONSUMPTION_CURRENT`

Estimated total energy consumption for the Property in a 12 month period. Value is Kilowatt Hours per Square Metre (kWh/m2)

LIGHTING COST CURRENT `LIGHTING_COST_CURRENT`

GBP. Current estimated annual energy costs for lighting the property.

HEATING COST CURRENT `HEATING_COST_CURRENT`

GBP. Current estimated annual energy costs for heating the property.

HOT WATER COST CURRENT `HOT_WATER_COST_CURRENT`

GBP. Current estimated annual energy costs for hot water

TOTAL FLOOR AREA `TOTAL_FLOOR_AREA`

The total useful floor area is the total of all enclosed spaces measured to the internal face of the external walls, i.e. the gross floor area as measured in accordance with the guidance issued from time to time by the Royal Institute of Chartered Surveyors or by a body replacing that institution. (m2)

ENERGY TARIFF `ENERGY_TARIFF`

Type of electricity tariff for the property, e.g. single.

MAINS GAS FLAG `MAINS_GAS_FLAG`

Whether mains gas is available. Yes means that there is a gas meter or a gas-burning appliance in the dwelling. A closed-off gas pipe does not count.

FLOOR LEVEL `FLOOR_LEVEL`

Flats and maisonettes only. Floor level relative to the lowest level of the property (0 for ground floor). If there is a basement, the basement is level 0 and the other floors are from 1 upwards

FLAT TOP STOREY `FLAT_TOP_STOREY` Whether the flat is on the top storey

FLAT STOREY COUNT `FLAT_STOREY_COUNT`

The number of Storeys in the Apartment Block.

MAIN HEATING CONTROLS `MAIN_HEATING_CONTROLS`

Type of main heating controls. Includes both main heating systems if there are two.

GLAZED TYPE `GLAZED_TYPE`

The type of glazing. From British Fenestration Rating Council or manufacturer declaration, give as one of; single; double; triple.

EXTENSION COUNT `EXTENSION_COUNT`

The number of extensions added to the property. Between 0 and 4.

NUMBER HABITABLE ROOMS `NUMBER_HABITABLE_ROOMS`

Habitable rooms include any living room, sitting room, dining room, bedroom, study and similar; and also a non-separated conservatory. A kitchen/diner having a discrete seating area (with space for a table and four chairs) also counts as a habitable room. A non-separated conservatory adds to the habitable room count if it has an internal quality door between it and the dwelling. Excluded from the room count are any room used solely as a kitchen, utility room, bathroom, cloakroom, en-suite accommodation and similar; any hallway, stairs or landing; and also any room not having a window.

HOTWATER DESCRIPTION `HOTWATER_DESCRIPTION` Overall description of the property feature

WINDOWS DESCRIPTION `WINDOWS_DESCRIPTION` Overall description of the property feature

SECONDHEAT DESCRIPTION `SECONDHEAT_DESCRIPTION`

Overall description of the property feature

MAINHEAT DESCRIPTION `MAINHEAT_DESCRIPTION`

Overall description of the property feature

MAIN FUEL `MAIN_FUEL`

The type of fuel used to power the central heating e.g. Gas, Electricity

WIND TURBINE COUNT `WIND_TURBINE_COUNT`

Number of wind turbines; 0 if none.

PHOTO SUPPLY `PHOTO_SUPPLY`

Percentage of photovoltaic area as a percentage of total roof area. 0% indicates that a Photovoltaic Supply is not present in the property.

SOLAR WATER HEATING FLAG `SOLAR_WATER_HEATING_FLAG`

Indicates whether the heating in the Property is solar powered

9.4 Variable Description Table

Variable ID	Variable Description	Domain	Distribution
PCS	Postcode sector	Geolocator	NA
CURRENT_ENERGY_EFFICIENCY	Based on cost of energy, in kWh/year multiplied by fuel costs.	Energy Info	Approximately Symmetrical
ENVIRONMENT_IMPACT_CURRENT	The Environmental Impact Rating. The higher the rating the lower the CO_2 emissions. (CO_2 emissions in tonnes / year)	Energy Info	Approximately Symmetrical
ENERGY_CONSUMPTION_CURRENT	Estimated total energy consumption for the Property in a 12 month period. Value is Kilowatt Hours per Square Metre (kWh/m^2)	Energy Info	Deleted due to missingness
LIGHTING_COST_CURRENT	GBP per anum	Energy Info	Deleted due to missingness
HEATING_COST_CURRENT	GBP per anum	Energy Info	Deleted due to missingness
HOT_WATER_COST_CURRENT	GBP per anum	Energy Info	Deleted due to missingness
TOTAL_FLOOR_AREA	The total useful floor area is the total of all enclosed	Energy Info	Approximately Symmetrical

	spaces measured to the internal face of the external walls (m^2)		
FLAT_STOREY_COUNT	The number of storeys in the Apartment Block.	Energy Info	Deleted due to missingness
NUMBER_HABITABLE_ROOMS	Habitable rooms include any living room, sitting room, dining room, bedroom, study and similar.	Energy Info	Deleted due to missingness
CER_A	Proportion of certificates with an A rating within the PCS	Energy Info	Approximately Symmetrical
CER_B	Proportion of certificates with an B rating within the PCS	Energy Info	Approximately Symmetrical
CER_C	Proportion of certificates with an C rating within the PCS	Energy Info	Approximately Symmetrical
CER_D	Proportion of certificates with an D rating within the PCS	Energy Info	Approximately Symmetrical
CER_E	Proportion of certificates with an E rating within the PCS	Energy Info	Approximately Symmetrical
CER_F	Proportion of certificates with an F	Energy Info	Approximately Symmetrical

	rating within the PCS		
CER_G	Proportion of certificates with an G rating within the PCS	Energy Info	Approximately Symmetrical
CER_INVALID!	Proportion of certificates with an Invalid rating within the PCS	Energy Info	Approximately Symmetrical
PROP_TYPE_Bungalow	proportion of properties of type : bungalow within the PCS	Physical Attribute	Approximately Symmetrical
PROP_TYPE_Flat	proportion of properties of type : Flat within the PCS	Physical Attribute	Approximately Symmetrical
PROP_TYPE_House	proportion of properties of type : House within the PCS	Physical Attribute	Approximately Symmetrical
PROP_TYPE_Park_home	proportion of properties of type : Park Home within the PCS	Physical Attribute	Approximately Symmetrical
BUILT_FORM_Detached	Proportion of properties of building type : Detached within the PCS	Physical Attribute	Approximately Symmetrical
BUILT_FORM_Other	Proportion of properties of building type : other within the PCS	Physical Attribute	Approximately Symmetrical

BUILT_FORM_Semi-Detached	Proportion of properties of building type : semi-detached within the PCS	Physical Attribute	Approximately Symmetrical
BUILT_FORM_Terrace	Proportion of properties of building type : Terrace within the PCS	Physical Attribute	Approximately Symmetrical
TRANS_TYPE_new_build	Proportion of properties given an EPC as a New Build within the PCS	Energy Info	Approximately Symmetrical
TRANS_TYPE_rental_private	Proportion of properties given an EPC when rented privately within the PCS	Energy Info	Approximately Symmetrical
TRANS_TYPE_rental_social	Proportion of properties given an EPC when rented socially within the PCS	Energy Info	Approximately Symmetrical
TRANS_TYPE_sale	Proportion of properties given an EPC when sold within the PCS	Energy Info	Approximately Symmetrical
TRANS_TYPE_unknown	Proportion of properties given an EPC for an unknown reason within the PCS	Energy Info	Approximately Symmetrical
TRANS_TYPE_upgrade_assessment	Proportion of properties given an EPC when assessed	Energy Info	Approximately Symmetrical

	for an upgrade within the PCS		
ET_24_hour	Proportion of properties on a 24 hour tariff within the PCS	Energy Info	Approximately Symmetrical
ET_dual	Proportion of properties on a dual fuel tariff within the PCS	Energy Info	Approximately Symmetrical
ET_offpeak	Proportion of properties on an off-peak tariff within the PCS	Energy Info	Approximately Symmetrical
ET_single	Proportion of properties on a single tariff within the PCS	Energy Info	Approximately Symmetrical
ET_standard	Proportion of properties on a standard tariff within the PCS	Energy Info	Approximately Symmetrical
ET_unknown	Proportion of properties on an unknown tariff within the PCS	Energy Info	Approximately Symmetrical
MGF_FALSE	Proportion of properties with a false mains gas flag within the PCS	Physical Attribute	Approximately Symmetrical
MGF_NA	Proportion of properties with an NA mains gas flag within the PCS	Physical Attribute	Approximately Symmetrical

MGF_TRUE	Proportion of properties with a true mains gas flag within the PCS	Physical Attribute	Approximately Symmetrical
FL_basement	Proportion of properties on the basement level within the PCS (flats only)	Physical Attribute	Approximately Symmetrical
FL_ground	Proportion of properties on the ground floor within the PCS (flats only)	Physical Attribute	Approximately Symmetrical
FL_middle_floor	Proportion of properties on any middle floor within the PCS (flats only)	Physical Attribute	Approximately Symmetrical
FL_top_floor	Proportion of properties on the top floor within the PCS (flats only)	Physical Attribute	Approximately Symmetrical
FL_unknown	Proportion of properties on an unknown level within the PCS (flats only)	Physical Attribute	Approximately Symmetrical
FTS_N	Proportion of of flats which are not on the top storey within the PCS	Physical Attribute	Approximately Symmetrical
FTS_NA	Proportion of flats without applicable data within the PCS	Physical Attribute	Approximately Symmetrical

FTS_Y	Proportion of flats which are on the top storey within the PCS	Physical Attribute	Approximately Symmetrical
GLAZED_double	Proportion of properties with double glazing within the PCS	Fixtures and Fittings	Approximately Symmetrical
GLAZED_secondary	Proportion of properties with secondary glazing within the PCS	Fixtures and Fittings	Approximately Symmetrical
GLAZED_single	Proportion of properties with single glazing within the PCS	Fixtures and Fittings	Approximately Symmetrical
GLAZED_triple	Proportion of properties with triple glazing within the PCS	Fixtures and Fittings	Approximately Symmetrical
GLAZED_unknown	Proportion of properties with unknown glazing within the PCS	Fixtures and Fittings	Approximately Symmetrical
EXT_0	Proportion of properties with no extensions within the PCS	Physical Attribute	Approximately Symmetrical
EXT_1	Proportion of properties with 1 extension within the PCS	Physical Attribute	Approximately Symmetrical
EXT_2	Proportion of properties with 2 extensions	Physical Attribute	Approximately Symmetrical

	within the PCS		
EXT_3	Proportion of properties with 3 extensions within the PCS	Physical Attribute	Approximately Symmetrical
EXT_4	Proportion of properties with 4 extensions within the PCS	Physical Attribute	Approximately Symmetrical
EXT_NA	Proportion of properties where extensions are not applicable within the PCS	Physical Attribute	Approximately Symmetrical
HWD_NA	Proportion of houses that have an invalid entry within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_community_scheme	Proportion of houses that have hot water through a community scheme within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_elec_immersion_offpeak	Proportion of houses that have hot water through an electric immersion heater on an off peak tariff within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_elec_immersion_standard"	Proportion of houses that have hot water	Fixtures and Fittings	Approximately Symmetrical

	through an electric immersion on a standard tariff within the PCS		
HWD_electric_instant	Proportion of houses that havehot water through an instant electric system within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_gas_boiler	Proportion of houses that have hot water through a gas boiler within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_gas_other	Proportion of houses that have hot water through other gas means within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_heat_pump	Proportion of houses that have hot water through a heat pump within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_hot_water	Proportion of houses that have an independent hot water system within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_main_system	Proportion of houses that have hot water through their mains system within the PCS	Fixtures and Fittings	Approximately Symmetrical

HWD_none	Proportion of houses that do not have an obvious hot water system within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_oil	Proportion of houses that have hot water through an oil system within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_room_heaters	Proportion of houses that have hotwater through individual room heaters within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_secondary_system	Proportion of houses that have hot water through a secondary system within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_solid_fuel	Proportion of houses that have hot water through solid fuel within the PCS	Fixtures and Fittings	Approximately Symmetrical
HWD_unclear_origin	Proportion of houses that have hot water through any other means within the PCS	Fixtures and Fittings	Approximately Symmetrical
SHD_NA	Proportion of properties where secondary heating is not applicable	Fixtures and Fittings	Approximately Symmetrical

	within the PCS		
SHD_community_scheme	Proportion of properties where secondary heating is through a community scheme within the PCS	Fixtures and Fittings	Approximately Symmetrical
SHD_gas	Proportion of properties where secondary heating is through a gas system within the PCS	Fixtures and Fittings	Approximately Symmetrical
SHD_hot_water_only	Proportion of properties where secondary heating is hot water only within the PCS	Fixtures and Fittings	Approximately Symmetrical
SHD_none	Proportion of properties where secondary heating is not installed within the PCS	Fixtures and Fittings	Approximately Symmetrical
SHD_other	Proportion of properties where secondary heating is delivered by any other means within the PCS	Fixtures and Fittings	Approximately Symmetrical
SHD_room_heaters_electric	Proportion of properties where	Fixtures and Fittings	Approximately Symmetrical

	secondary heating is through individual electric room heaters within the PCS		
SHD_room_heaters_gas	Proportion of properties where secondary heating is through gas room heaters within the PCS	Fixtures and Fittings	Approximately Symmetrical
SHD_room_heaters_other	Proportion of properties where secondary heating is through individual room heaters with other fuel sources within the PCS	Fixtures and Fittings	Approximately Symmetrical
SHD_room_heaters_wood	Proportion of properties where secondary heating is through wood fueled room heaters within the PCS	Fixtures and Fittings	Approximately Symmetrical
SHD_secondary_heating	Proportion of properties where secondary heating is installed but type is unclear within the PCS	Fixtures and Fittings	Approximately Symmetrical
SHD_underfloor_heating	Proportion of properties	Fixtures and Fittings	Approximately Symmetrical

	where secondary heating is through underfloor heating within the PCS		
WTC_FALSE	Proportion of properties that do not have wind turbines within the PCS	Physical Attribute	Approximately Symmetrical
WTC_NA	Proportion of properties where wind turbine data is not applicable within the PCS	Physical Attribute	Approximately Symmetrical
WTC_TRUE	Proportion of properties that do have wind turbines within the PCS	Physical Attribute	Approximately Symmetrical
SWHF_N	Proportion of properties that do not have a solar water heating flag within the PCS	Physical Attribute	Approximately Symmetrical
SWHF_NA	Proportion of properties where a solar water heating flag is not applicable within the PCS	Physical Attribute	Approximately Symmetrical
SWHF_Y	Proportion of properties that do have a solar water heating flag within the PCS	Physical Attribute	Approximately Symmetrical

PHOTO_FALSE	Proportion of properties that do not have any photovoltaics on their roof within the PCS	Physical Attribute	Approximately Symmetrical
PHOTO_NA	Proportion of properties where photovoltaics are not applicable within the PCS	Physical Attribute	Approximately Symmetrical
PHOTO_TRUE	Proportion of properties that do have a percentage of their roof covered by photovoltaics within the PCS	Physical Attribute	Approximately Symmetrical
MF_Community_scheme	Proportion of properties where the main fuel source is from a community scheme within the PCS	Energy Info	Approximately Symmetrical
MF_No_Data	Proportion of properties where there is no data for the main fuel source within the PCS	Energy Info	Approximately Symmetrical
MF_biofuel	Proportion of properties where the main fuel source is biofuels	Energy Info	Approximately Symmetrical

	within the PCS		
MF_coal	Proportion of properties where the main fuel source is coal within the PCS	Energy Info	Approximately Symmetrical
MF_electric	Proportion of properties where the main fuel source is electricity within the PCS	Energy Info	Approximately Symmetrical
MF_gas	Proportion of properties where the main fuel source is gas within the PCS	Energy Info	Approximately Symmetrical
MF_oil	Proportion of properties where the main fuel source is oil within the PCS	Energy Info	Approximately Symmetrical
MF_wood	Proportion of properties where the main fuel source is wood within the PCS	Energy Info	Approximately Symmetrical
ACCOM_unshared_flat_converted_building	Proportion of properties which are flats in converted buildings	Physical Attribute	Approximately Symmetrical
ACCOM_unshared_flat_commercial	Proportion of properties which are flats	Physical Attribute	Approximately Symmetrical

	in commercial buildings		
ACCOM_shared_dwelling	Proportion of properties which are houses of multiple occupancy	Physical Attribute	Approximately Symmetrical
ACCOM_unshared_house_detached	Proportion of properties which are unshared detached houses	Physical Attribute	Approximately Symmetrical
ACCOM_unshared_house_semi	Proportion of properties which are unshared semi-detached houses	Physical Attribute	Approximately Symmetrical
ACCOM_unshared_house_terrace	Proportion of properties which are terrace houses	Physical Attribute	Approximately Symmetrical
ACCOM_unshared_flat_purposebuiltblock	Proportion of properties which are flats in purpose built blocks	Physical Attribute	Approximately Symmetrical
ECOACT_active_unemployed	Proportion of people who are economically active but consider themselves unemployed.	Demographic	Approximately Symmetrical
ECOACT_active_pt	Proportion of people who are in part time employment	Demographic	Approximately Symmetrical
ECOACT_active_ft	Proportion of people who are in full time employment	Demographic	Approximately Symmetrical

ECOACT_active_student	Proportion of people who are economically active students	Demographic	Approximately Symmetrical
ECOACT_inactive_retired	Proportion of people who are retired	Demographic	Approximately Symmetrical
ECOACT_inactive_student	Proportion of people who are economically inactive students	Demographic	Approximately Symmetrical
ECOACT_inactive_carer	Proportion of people who are economically inactive through being a carer	Demographic	Approximately Symmetrical
ECOACT_inactive_LTS_disabled	Proportion of people who are economically inactive through having a disability.	Demographic	Approximately Symmetrical
ECOACT_inactive_other	Proportion of people who are economically inactive for an undisclosed reason	Demographic	Approximately Symmetrical
ECOACT_unemp_never	Proportion of people who are have never been employed	Demographic	Approximately Symmetrical
ECOACT_lt_unemp	Proportion of people who are long term unemployed	Demographic	Approximately Symmetrical

ECOACT_active_selfemp_ft	Proportion of people who are in full time self-employment	Demographic	Approximately Symmetrical
ECOACT_active_selfemp_pt	Proportion of people who are in part time self-employment	Demographic	Approximately Symmetrical
NSSEC_1	Proportion of people who are in higher managerial and professional occupations	Demographic	Approximately Symmetrical
NSSEC_2	Proportion of people who are in lower managerial and professional occupations	Demographic	Approximately Symmetrical
NSSEC_3	Proportion of people who are in Intermediate occupations	Demographic	Approximately Symmetrical
NSSEC_4	Proportion of people who are small employers and own account workers	Demographic	Approximately Symmetrical
NSSEC_5	Proportion of people who are in lower supervisory and technical occupations	Demographic	Approximately Symmetrical
NSSEC_6	Proportion of people who are in semi-routine occupations	Demographic	Approximately Symmetrical

NSSEC_7	Proportion of people who are in routine occupations	Demographic	Approximately Symmetrical
NSSEC_8	Proportion of people who have never worked or are long term unemployed	Demographic	Approximately Symmetrical
CENTHEAT_NONE	Proportion of homes with no central heating	Fixtures and Fittings	Approximately Symmetrical
CENTHEAT_GAS	Proportion of homes with gas central heating	Fixtures and Fittings	Approximately Symmetrical
CENTHEAT_ELECTRIC	Proportion of homes with electric central heating	Fixtures and Fittings	Approximately Symmetrical
CENTHEAT_OIL	Proportion of homes with oil central heating	Fixtures and Fittings	Approximately Symmetrical
CENTHEAT_SOLID	Proportion of homes with solid fuel central heating	Fixtures and Fittings	Approximately Symmetrical
CENTHEAT_OTHER	Proportion of home with central heating that uses an 'other' fuel type	Fixtures and Fittings	Approximately Symmetrical
CENTHEAT_TWOORMORE	Proportion of homes with central heating powered by two or more fuel types	Fixtures and Fittings	Approximately Symmetrical
SINGLE	Proportion of households who consider themselves single	Demographic	Approximately Symmetrical

MARRIED	Proportion of households who consider themselves married	Demographic	Approximately Symmetrical
CIVIL	Proportion of households who consider themselves in a civil relationship	Demographic	Approximately Symmetrical
SEPARATED	Proportion of households who consider themselves separated	Demographic	Approximately Symmetrical
DIVORCED	Proportion of households who consider themselves divorced	Demographic	Approximately Symmetrical
WIDOWED	Proportion of households who consider themselves widowed	Demographic	Approximately Symmetrical
1_ROOM	Proportion of homes with this number of rooms.	Physical Attribute	Deleted due to missingness
2_ROOMS	Proportion of homes with this number of rooms.	Physical Attribute	Deleted due to missingness
3_ROOMS	Proportion of homes with this number of rooms.	Physical Attribute	Deleted due to missingness
4_ROOMS	Proportion of homes with this number of rooms.	Physical Attribute	Deleted due to missingness
5_ROOMS	Proportion of homes with this number of rooms.	Physical Attribute	Deleted due to missingness

6_ROOMS	Proportion of homes with this number of rooms.	Physical Attribute	Deleted due to missingness
7_ROOMS	Proportion of homes with this number of rooms.	Physical Attribute	Deleted due to missingness
8_ROOMS	Proportion of homes with this number of rooms.	Physical Attribute	Deleted due to missingness
9_ORMORE_ROOMS	Proportion of homes with this number of rooms.	Physical Attribute	Deleted due to missingness
TENURE_owned_outright	Proportion of homes that are owned outright	Demographic	Approximately Symmetrical
TENURE_mortgaged	Proportion of homes that are mortgaged	Demographic	Approximately Symmetrical
TENURE_social_rented	Proportion of homes that are social rented	Demographic	Approximately Symmetrical
TENURE_private_rented	Proportion of homes that are private rented	Demographic	Approximately Symmetrical
AGE0_4	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE5_7	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE8_9	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE10_15	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE16_17	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical

AGE18_19	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE20_24	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE25_29	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE30_44	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE45_59	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE60_64	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE65_74	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE75_84	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE85_89	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical
AGE90_over	Proportion of people of this age bracket.	Demographic	Approximately Symmetrical

9.5 Initial Cluster Summary Tables

Measure	Normal Distribution 1	2	3	4	5	Box Cox Distribution 1	2	3	4	5
CURRENT_ENERGY_EFFICIENCY	0.447284793	0.154078838	-0.125075031	0.197058058	-1.223251772	-0.913453364	0.164468135	0.1254402	0.516372148	-0.11449795
ENVIRONMENT_IMPACT_CURRENT	0.513445447	0.151765103	-0.196091032	0.156388643	-1.263867549	-0.906413619	0.137011535	0.130468133	0.454946144	-0.088155277
TOTAL_FLOOR_AREA	-0.497791458	-0.427991513	0.424650447	-0.25185092	0.967500878	0.741286126	-0.307101617	-0.131006271	-0.39203268	0.414844226
CER_A	-0.046643889	-0.004371425	0.004503488	-0.007345991	0.06584343	0.422037973	-0.043730838	0.237247292	-0.307212043	0.090393339
CER_B	0.222593739	-0.116057679	-0.141146397	-0.052756162	-0.339782969	0.205816145	0.257044773	0.278356101	0.32274416	0.227488381
CER_C	0.471152756	0.150139505	-0.209775527	0.102478788	-0.729170171	-0.290941548	0.270723766	0.246271569	0.485164914	0.053977517
CER_D	-0.258631689	0.221403749	0.240568828	0.253664535	-0.434821146	-0.207496447	0.247105044	0.278787112	-0.200100062	0.349236737
CER_E	-0.282510096	-0.041992593	0.184857537	-0.08034463	0.591229525	0.484646332	0.132775412	0.160021676	-0.04152249	0.312185491
CER_F	-0.251227919	-0.279197247	0.009996426	-0.265930546	1.180723403	0.60968663	0.140111116	0.17348368	0.118731475	0.275957584
CER_G	-0.170332928	-0.135322495	-0.081667213	-0.189127663	0.781976459	0.553468731	0.23252569	0.215312096	0.159343662	0.247375442
CER_INVALID!	-0.010985465	-0.010951114	-0.01065143	-0.010371971	-0.008676122	0.080507803	0.010408953	0.025825447	0.017688571	-0.035842255
PROP_TYPE_Bungalow	-0.74670584	-0.525555401	0.457537524	-0.159938638	0.981704432	0.627455906	0.055679256	0.361648395	-0.689260564	0.478587128
PROP_TYPE_Flat	1.539654081	0.055448156	-0.429766972	-0.239431746	-0.799683063	-0.336366884	0.30747376	0.172044995	0.751071815	0.044787316
PROP_TYPE_House	-1.365842549	0.200982382	0.256099975	0.350830723	0.412174412	0.295709876	0.134292261	0.253170423	-1.281502691	0.179910619
PROP_TYPE_Park_home	-0.068312657	-0.064196932	0.063630102	-0.018490934	0.05469651	0.365141791	-0.177831914	-0.025684065	-0.22654896	0.151180719
BUILT_FORM_Detached	-0.558952229	-0.829445111	0.537782735	-0.33233114	1.112340357	0.712196098	-0.465246631	0.046240866	-0.127191879	0.496366052
BUILT_FORM_Other	0.349658013	-0.085520233	-0.164499267	-0.095372049	-0.279015565	0.196309495	0.280917998	0.293796935	0.376638885	0.251998125
BUILT_FORM_Semi-Detached	-0.777201627	-0.194846779	0.246049841	0.394037352	-0.057271235	0.196452938	-0.011399778	0.450153083	-0.579662365	0.395346542
BUILT_FORM_Terrace	0.897849778	0.930978024	-0.585079095	0.062848216	-0.83245837	-0.559011851	0.806825629	0.189778971	0.808442226	-0.400221829
TRANS_TYPE_new_build	0.096035251	-0.189521831	-0.131923683	-0.103170934	-0.150716053	0.299768441	0.122174965	0.259581529	0.27367173	0.233453351
TRANS_TYPE_rental_private	1.001012351	-0.153722835	-0.2806951	-0.223503994	-0.257844874	0.164222614	0.183967715	0.174569567	0.56112074	0.157413737
TRANS_TYPE_rental_social	0.098583378	1.01591356	-0.366920306	0.173423064	-0.380617131	0.089314395	0.678786689	0.38946288	0.282671692	-0.024510061
TRANS_TYPE_sale	-0.542574301	-0.709802437	0.683759425	0.117925191	0.502059533	0.524350431	-0.657694968	0.134693851	-0.49729045	0.711535606
TRANS_TYPE_unknown	-0.345983362	0.201807635	-0.077628145	0.042276991	-0.130057662	0.154956983	0.363380277	0.292760598	-0.07604462	0.223487009
TRANS_TYPE_upgrade_assessment	-0.737057541	0.563149946	-0.024999152	0.105199819	0.528432232	0.501631083	0.519961905	0.358420921	-0.676183074	0.261062692
ET_24_hour	0.193150538	-0.031882101	-0.032960947	-0.015329813	-0.001182505	0.138818491	-0.100844938	0.012412544	0.157440305	-0.030848224
ET_dual	0.107021994	-0.304580015	0.050024821	-0.135412676	0.875266274	0.552336194	0.101013398	0.193158422	0.298233315	0.230663309
ET_offpeak	0.043303799	-0.084696142	-0.05454159	-0.057285674	-0.009807736	0.529613523	-0.318190017	0.104203866	0.200063825	0.142588713
ET_single	-0.510574002	0.437853285	0.14127241	0.257973597	-0.23313454	-0.33056032	0.40981144	0.215312208	-0.638235223	0.153005669
ET_standard	0.080018005	-0.175148859	-0.125804255	-0.093886663	-0.159543192	0.303285788	0.108955733	0.264434432	0.267428693	0.235841164
ET_unknown	0.655550908	-0.20275661	-0.109200398	-0.171186652	-0.240133824	0.203520768	0.20670789	0.220256404	0.461723779	0.25759422
MGF_FALSE	0.051584034	-0.450878286	-0.177488677	-0.438208103	2.064279553	0.466129738	0.158955825	0.184932366	0.298589346	0.210202686
MGF_NA	0.094760235	-0.242587032	-0.159385224	-0.126353546	-0.167099633	0.236804807	0.23784423	0.255171645	0.250702543	0.252784738
MGF_TRUE	-0.105542304	0.541206741	0.25464205	0.454763437	-1.645966623	-1.201403252	0.554125354	0.426890057	-0.224011001	0.384863242
FL_basement	0.175127268	-0.039041649	-0.062127381	-0.054919938	-0.070898598	-0.383039634	0.149708834	0.013408529	1.174850752	-0.169961004
FL_ground	0.711369113	0.167658576	-0.207161699	-0.032378825	-0.564138042	-0.153313721	0.389506786	0.308264681	0.593012434	0.200566745
FL_middle_floor	1.430838867	-0.06604178	-0.351205747	-0.270358974	-0.510883985	0.041540079	0.295375682	0.281606903	0.452823509	0.191025791
FL_top_floor	0.741740013	0.079950273	-0.265579371	-0.117891622	-0.576818227	-0.046253798	0.320292732	0.251935961	0.524365453	0.179727388
FL_unknown	-1.514336945	-0.048976924	0.423422881	0.242732555	0.792762302	0.831320318	-0.215061303	0.153027929	-1.481783166	0.327461526
GLAZED_double	-0.547252819	0.406597369	0.212511251	0.322861772	0.072387627	-0.079934891	0.441235552	0.330694089	-0.786220988	0.206348333
GLAZED_secondary	0.293251029	-0.212400421	-0.01870197	-0.184302863	0.110566843	0.470203478	-0.188784003	0.117774363	0.44184984	0.349866383
GLAZED_single	0.364090202	-0.048000921	-0.092992833	-0.117175029	-0.025218819	0.319111072	0.229865886	0.236414649	0.443020919	0.218953793
GLAZED_triple	0.211404869	-0.103106057	0.000102292	-0.09002634	0.140843051	0.367373324	-0.261113891	0.050445223	0.311503446	0.233081227
GLAZED_unknown	0.416290456	-0.363133807	-0.19904854	-0.266101493	-0.1027957	0.196938613	0.119115676	0.149223653	0.27956939	0.165705962
EXT_0	0.50016526	0.432953702	-0.264144005	0.15266082	-0.659761295	-0.686972984	0.436875084	0.061020842	0.465453769	-0.331104338
EXT_1	-0.634496712	-0.067539868	0.381523723	0.065171461	0.620893685	0.556310965	0.058514621	0.247764221	-0.428207911	0.447563947
EXT_2	-0.383137862	-0.342096927	0.346148853	-0.154584507	0.828465767	0.645008053	-0.009408408	0.228917433	-0.065152105	0.477547819
EXT_3	-0.364475072	-0.350999113	0.352924641	-0.210155452	1.097827705	0.767604738	-0.21532975	0.249503323	-0.297120617	0.550822104
EXT_4	-0.361508303	-0.41847176	0.35185033	-0.29182223	1.495431332	0.852123283	-0.590141229	0.144980088	-0.321849765	0.581932014
EXT_NA	0.081369412	-0.235014959	-0.150184616	-0.115374943	-0.182299577	0.241537568	0.242563831	0.259634199	0.250767544	0.254939435
HWD_NA	-0.020537354	-0.016516543	-0.008123138	-0.015161769	0.059191884	0.06173431	-0.003296299	-0.01308518	-0.016784802	-0.001979038
HWD_community_scheme	0.746592914	0.022257547	-0.196126925	-0.13627626	-0.230453052	0.00386285	0.254344539	0.128039862	0.655331433	-0.06840323
HWD_elec_immersion_offpeak	0.544758458	-0.223056507	-0.133443642	-0.178158824	0.539213115	0.438378328	0.168141578	0.229933836	0.409480231	0.216786222
HWD_elec_immersion_standard	0.343475736	0.013212253	-0.205946344	-0.151953113	0.017320098	0.297170997	0.307776393	0.269215521	0.352139593	0.23673847
HWD_electric_instant	0.260278161	-0.020241691	-0.120585899	-0.096536691	-0.041718828	0.305123444	0.194689851	0.235049612	0.50080887	-0.019051563
HWD_gas_boiler	0.128396721	0.227683178	-0.090848222	0.041126712	-0.262687494	-0.413792591	0.480979496	0.37742872	0.358261776	0.266603394
HWD_gas_other	0.038972035	0.024597037	-0.009903399	-0.009555123	-0.032768411	-0.246858402	0.339932186	0.246022013	0.298760595	0.062084725
HWD_heat_pump	0.092248998	0.006327199	-0.000530006	-0.01668257	0.013109657	0.077248917	0.003674706	-0.012059146	0.030045927	0.021288648
HWD_hot_water	-0.017095768	-0.110703492	-0.042429238	-0.053162475	-0.053371877	0.301647349	0.208882993	0.324598063	0.257181445	0.330099387
HWD_main_system	-0.812494871	0.159383355	0.302071935	0.276282087	-0.2224828	-0.217215004	0.092658774	0.252757094	-0.92274868	0.371856952
HWD_none	0.114552744	-0.059886183	0.021097276	0.006989598	-0.076835444	-0.183927872	0.099783566	0.311439415	0.447585764	0.246984326
HWD_oil	-0.076524885	-0.076162578	-0.030185575	-0.072448653	0.370605839	1.660054171	-0.475969189	-0.292946917	-0.467566715	-0.015189346
HWD_room_heaters	-0.118380313	-0.110088837	0.094390859	0.005710669	0.113382457	0.305034338	0.017944705	0.357314757	-0.108681982	0.397169193
HWD_secondary_system	-0.186291901	-0.157377281	-0.038251843	-0.136309198	0.990271445	0.998406868	-0.196631499	0.080633605	-0.464763561	-0.002902548
HWD_solid_fuel	-0.153062014	-0.14591501	-0.024272831	-0.112646308	0.761723072	1.227463086	-0.398886583	-0.065169712	-0.444361635	-0.048264762
HWD_unclear_origin	-0.040974674	-0.018120766	-0.034308354	-0.032089889	0.333728608	0.137388112	-0.035762158	0.010115527	-0.006638334	-0.046565716
SHD_NA	0.002724259	-0.006023576	-0.010938939	-0.000497199	-0.020532225	-0.175833187	0.082617177	0.184173146	0.282344335	-0.000926168
SHD_community_scheme	0.136542802	-0.026872447	-0.025325422	-0.024879435	0.007387945	-0.034924264	-0.033059877	-0.033063778	0.299916034	-0.038240683
SHD_gas	0.051612038	-0.030151094	0.03312031	0.018096405	-0.023314756	-0.07032218	0.007006635	0.048258233	0.067065999	0.016483723
SHD_hot_water_only	-0.011363121	0.017844647	0.027434538	-0.006635969	-0.022167736	-0.018015848	0.094915369	0.029915988	-0.008535317	-0.046014129
SHD_none	0.975402565	0.26709416	-0.380705054	-0.062765422	-0.9566321	-0.772208759	0.295330238	-0.071126428	0.965632042	-0.26433064
SHD_other	0.034605762	0.022929376	-0.024950942	-0.025909478	0.138432621	0.053907593	-0.016376192	-0.024728357	0.103252328	-0.015254237
SHD_room_heaters_electric	-0.155810466	0.107180181	-0.065122008	0.166594516	0.142332428	0.271334206	0.245617837	0.35704462	-0.038128375	0.119028295
SHD_room_heaters_gas	-0.773891565	0.233183613	0.339674183	0.402986967	-0.728532972	-0.107056548	0.339479766	0.489505488	-0.484910713	0.522820198
SHD_room_heaters_other	-0.421456232	-0.538663594	0.212308863	-0.315277076	1.739808737	0.682956551	-0.085801195	0.266415514	-0.039110356	0.413458663
SHD_room_heaters_wood	-0.426230086	-0.484536942	0.250587084	-0.331138857	1.525067655	0.639605467	-0.142542175	0.306992863	-0.19048137	0.4616097
SHD_secondary_heating	-0.017091296	-0.110699016	-0.042434241	-0.053158002	-0.053389937	0.301634726	0.208877341	0.324599622	0.25718437	0.33010342
SHD_underfloor_heating	0.074024864	-0.04366925	0.004279062	-0.035905035	-0.005747249	0.005259644	-0.09881055	-0.055254765	0.303910073	0.047430062
WTC_FALSE	-0.026547563	0.285981064	0.051390798	0.094563157	0.130283535	0.005293787	0.330755676	0.05548693	-0.148301226	-0.027450539
WTC_NA	0.026136661	-0.285463099	-0.050630005	-0.093700624	-0.133522888	0.253313577	0.148694872	0.248591705	0.276539493	0.270569005
WTC_TRUE	0.006324973	-0.012775757	-0.011753371	-0.014015543	0.043404505	0.083637648	0.153610556	0.199789206	0.30383421	-0.023376472
PHOTO_FALSE	0.198857185	0.081931565	0.0903334	0.078940823	-0.102792042	-0.245173487	-0.013853248	-0.007329333	0.1673597	0.024372764
PHOTO_NA	-0.180718368	-0.087772881	-0.09223233	-0.082306708	0.088489264	0.243648531	0.114851832	0.117528853	0.021284002	0.105056672
PHOTO_TRUE	-0.250116014	0.080197911	0.02591938	0.046136022	0.19709215	0.480527328	0.123288838	0.408252072	-0.668074333	0.348511157
MF_Community_scheme	0.056505767	-0.012405654	-0.01699624	-0.019084815	-0.010384193	0.061130338	-0.025938862	-0.045257431	0.222199258	-0.04681683
MF_No_Data	0.228966518	-0.132460604	-0.115035345	-0.098500036	-0.089506114	0.287426792	0.266296779	0.294309242	0.333493535	0.280668691
MF_biofuel	-0.120610458	-0.152796029	-0.062199436	-0.181002103	1.173037004	1.009179226	-0.401495437	0.196991979	-0.453955144	0.09904465
MF_coal	-0.152066589	-0.143886941	-0.004333424	-0.036791105	0.483104244	0.540891569	0.084613395	0.347342231	-0.089451009	0.304454739
MF_electric	0.640058006	-0.243577917	-0.237220209	-0.277526778	0.517848206	0.362070716	0.211287994	0.211792162	0.376207165	0.208392519
MF_gas	-0.228533705	0.483690052	0.237630159	0.456286702	-1.752412866	-1.295942695	0.495515502	0.438435624	-0.331915017	0.382711463
MF_oil	-0.405695499	-0.413017357	-0.057561505	-0.368179671	2.212732549	0.855131272	-0.775881426	0.185552381	-0.397805661	0.467161385
MF_wood	-0.113440472	-0.109936025	-0.031052282	-0.099486132	0.436505052	1.191448728	-0.516242827	0.063018778	-0.634979643	0.180403738
ACCOM_unshared_flat_converted_building	1.720305325	-0.000634374	-0.272851761	-0.230960915	-0.368408092	0.053041691	-0.12162273	0.018956398	0.298765781	-0.140990779
ACCOM_unshared_flat_commercial	1.256058628	-0.002431009	-0.206319968	-0.142641397	-0.331530383	0.073609813	-0.170771904	0.039632703	0.189775568	-0.094336327
ACCO_shared_dwelling	1.364335854	0.054355187	-0.247122779	-0.175990802	-0.30674074	-0.365220404	0.23694409	-0.052534197	1.030790749	-0.349779651
ACCOM_unshared_house_detached	-1.104208807	-0.98213848	0.766104203	-0.286356313	1.347899891	0.880107045	-0.83762421	0.031607671	-1.527921557	0.6545125
ACCOM_unshared_house_semi	-1.364885469	-0.185193884	0.150662227	0.449008334	-0.112294705	-0.083823794	-0.095088909	0.435042148	-1.64348043	0.266787096
ACCOM_unshared_house_terrace	-0.776184581	1.161355202	-0.560742711	0.278729907	-0.655391556	-0.473079001	0.9550157	0.359533916	-0.377380336	-0.587491107
ACCOM_unshared_flat_purposebuiltblock	1.821019268	0.349834648	-0.429752152	-0.148224746	-0.78735751	-0.82953289	0.387852156	0.09552452	0.906840397	-0.232894462
ECOACT_active_unemployed	-0.013085885	1.731255412	-0.676942723	0.079219042	-0.770243075	-0.79000064	1.41283128	0.285324438	-0.040978084	-0.732973544
ECOACT_active_pt	-1.986297816	-0.300030611	0.42517672	0.326984122	0.347942907	0.413421338	-0.387183306	0.337925214	-1.746647051	0.397768701
ECOACT_active_ft	-0.000209023	-0.890939514	-0.068493063	0.490269979	-0.512813304	-0.425893057	-0.786342321	0.407953375	0.149014854	0.232831768
ECOACT_active_student	1.285940665	0.042702408	-0.218329672	-0.125062109	-0.440619698	-0.029864736	0.016312583	0.016030207	-0.04992816	0.016084891
ECOACT_inactive_retired	-1.463253579	-0.774682251	0.75526375	-0.106147473	0.913389025	0.893303672	-0.760726804	0.026322349	-1.552890454	0.610431264
ECOACT_inactive_student	1.530595514	0.14789478	-0.26209492	-0.248713854	-0.337365001	-0.543580672	0.514315544	-0.279161153	1.218665455	-0.187924245
ECOACT_inactive_carer	-0.563827706	1.286108246	-0.227574782	-0.113667631	-0.174962247	-0.119706972	1.08076617	-0.054327461	-0.728509632	-0.193997003
ECOACT_inactive_LTS_disabled	-0.309737863	1.400075186	-0.57845801	-0.136718032	-0.454480258	-0.390124832	1.120183201	0.332332473	-0.417626677	-0.708511725
ECOACT_inactive_other	0.178598071	1.010285133	-0.330355795	-0.13609014	-0.245833297	-0.018156507	0.022201662	0.019680335	-0.034477028	-0.010228199
ECOACT_unemp_never	0.05171093	1.735234457	-0.585434876	-0.068474014	-0.621095786	-0.603223039	1.080026863	0.175806748	0.028065615	-0.53329074
ECOACT_lt_unemp	-0.028939979	1.677176989	-0.649959631	0.068571968	-0.72018394	-0.688357801	1.294649211	0.254032477	-0.043048974	-0.686598346
ECOACT_active_selfemp_ft	-0.279476453	-1.042871629	-0.459163314	0.377359924	1.651597397	1.204595414	-1.149762109	-0.281798747	-0.324127634	0.416377965
ECOACT_active_selfemp_pt	-0.284075293	-0.837888917	0.571493268	-0.474883258	1.393019654	1.100352852	-0.925276033	-0.410135023	-0.30807515	0.5111162
NSSEC_1	1.078841117	-1.08863995	0.494460059	-0.330005921	0.05642181	0.22279003	-1.339673441	-0.314321839	0.92771839	0.696696145
NSSEC_2	0.755719787	-1.466924764	0.605710575	-0.193575696	0.18196687	0.191254744	-1.342993738	-0.303998466	0.834441199	0.757696224
NSSEC_3	-0.75187301	-0.601803075	0.322177221	0.449313867	-0.649228789	-0.499097441	-0.473850899	0.373291693	-0.746071104	0.592318525
NSSEC_4	-0.611035401	-0.755816473	0.381543836	-0.389169162	1.854216499	1.350346714	-0.868003943	-0.276705814	-0.67534156	0.286200996
NSSEC_5	-0.884567391	0.402793884	-0.408655811	0.53490039	-0.206713332	-0.145896605	0.401736155	0.622259946	-0.910705561	-0.543275095
NSSEC_6	-0.86958507	1.092076695	-0.541213932	0.394974539	-0.358518385	-0.277809899	1.028939051	0.497403586	-0.952043738	-0.679137995
NSSEC_7	-0.603408231	1.32640083	-0.631022063	0.311155188	-0.452925102	-0.314983431	1.123091353	0.46690137	-0.744539377	-0.7820383
NSSEC_8	0.387400955	1.726717428	-0.600244099	-0.165977799	-0.619501711	-0.813487663	1.396163247	0.140766763	0.553668262	-0.816424825
CENTHEAT_NONE	0.679621837	0.354242762	-0.465449839	-0.185275872	0.606602579	0.37286145	0.273854752	-0.120385395	0.575339123	-0.615588652
CENTHEAT_GAS	-0.332551184	0.302128425	0.228989297	0.504441231	-2.217708076	-1.562442965	0.276691062	0.508807139	-0.40932534	0.536406185
CENTHEAT_ELECTRIC	1.367791752	-0.096699739	-0.311006432	-0.241453348	0.259594712	0.028441967	0.02791826	0.027732781	0.029263086	-0.094982539
CENTHEAT_OIL	-0.469815979	-0.495282732	0.007704922	-0.415243379	2.418720348	0.578196171	-0.851512196	0.043178616	-0.244803085	0.208390421
CENTHEAT_SOLID	-0.51467399	-0.423177294	-0.046512319	-0.22885204	1.932777738	0.853051181	-0.591519269	0.150055156	-1.025806713	0.077805599
CENTHEAT_OTHER	0.827621263	0.683955566	-0.35472168	-0.267054951	-0.01799463	0.072180864	0.080011625	0.041680876	0.003241855	-0.170451246
CENTHEAT_TWOORMORE	-0.330156367	0.078665839	-0.132185007	-0.490214643	2.108367401	0.017817366	0.016342457	0.015607516	-0.118524849	0.015707643
SINGLE	1.964307368	0.663677235	-0.730276225	-0.069388803	-0.811151918	-1.013717631	0.869140281	0.15701398	1.491192621	-0.812404879
MARRIED	-1.682540675	-0.686029087	0.776347523	-0.003775187	0.897212986	0.915857881	-0.915747948	-0.132289332	-1.446092697	0.863689056
CIVIL	1.228589261	-0.091356633	-0.1949318	-0.17771863	-0.100674132	0.00708441	-0.167222712	-0.035791303	0.664685132	-0.189469231
SEPARATED	0.118372826	1.314476031	-0.603215542	0.126709024	-0.69297218	-0.619841971	1.163362796	0.206325471	0.070119208	-0.64275286
DIVORCED	-0.826291166	0.500917694	-0.278119148	0.349991243	-0.16754229	-0.068955059	0.448851291	0.448194784	-0.841553614	-0.433002195
WIDOWED	-1.340537009	-0.127202288	0.476366289	-0.028557291	0.232887085	0.348779117	-0.140128237	0.173438798	-1.42332159	0.333851676
TENURE_owned_outright	-1.266439574	-1.084793533	0.866220193	-0.170779816	1.032647635	0.974755607	-1.079191763	-0.128921094	-1.231485454	0.812213911
TENURE_mortgaged	-1.34147006	-0.766181444	0.346329477	0.551592555	-0.193047116	-0.14429485	-0.667823972	0.453529362	-1.213695356	0.586988036
TENURE_social_rented	0.400721869	1.394203921	-0.607041843	-0.048610108	-0.535866761	-0.306847814	0.930195309	0.145648808	0.304064273	-0.763371552
TENURE_private_rented	1.964360782	0.249468244	-0.531182507	-0.208370231	-0.352200346	-0.220128909	0.307350049	-0.053251624	1.443832594	-0.727920658
AGE0_4	-0.33120816	1.200357864	-0.533325206	0.289559679	-0.756200269	-0.704635178	1.11309412	0.254545847	-0.322919469	-0.400885879
AGE5_7	-1.036144405	0.923569286	-0.162008813	0.196699925	-0.331714535	-0.313167572	0.90028233	0.113479614	-0.953079833	-0.031373195
AGE8_9	-1.246775351	0.745520375	-0.000649344	0.122889474	-0.06710895	-0.082664369	0.734970491	0.040022825	-1.130221409	0.098139613
AGE10_15	-1.606742498	0.537119968	0.150986014	0.122489776	0.18941648	0.14442052	0.540653208	0.042075371	-1.464223335	0.228177881
AGE16_17	-1.205836896	0.333220941	0.157322464	0.089078264	0.129321444	0.135539066	0.355821303	0.128599839	-1.484801123	0.253478636
AGE18_19	0.894320991	0.107265422	-0.218224522	-0.092258484	-0.220106743	-0.357898035	0.590095507	0.083223689	0.160374098	-0.327905955
AGE20_24	1.65003959	0.238413237	-0.45187321	-0.111835269	-0.512455621	-0.959763016	0.742122576	0.177939791	1.305935699	-0.698173016
AGE25_29	1.842575216	0.494901725	-0.677148197	0.001645191	-0.818373075	-1.079158787	0.741269058	0.242541077	1.424993656	-0.756725905
AGE30_44	1.118210266	0.320323581	-0.559786055	0.249347542	-0.904507337	-0.861000205	0.356713728	0.190400126	1.066944935	-0.400480616
AGE45_59	-1.585213427	-0.66589588	0.578460421	0.026633278	0.949144574	0.916619923	-0.723334806	-0.037269971	-1.413755502	0.58276203
AGE60_64	-1.342436082	-0.853616877	0.685404843	-0.152573274	1.220470677	1.129340513	-0.85919049	-0.085629169	-1.311580601	0.52320255
AGE65_74	-1.304790239	-0.78842093	0.729544198	-0.189275459	1.091932889	1.023017416	-0.789486464	-0.068242616	-1.355083649	0.556100155
AGE75_84	-1.140515126	-0.649237928	0.731860335	-0.145590793	0.591331598	0.658082619	-0.634234374	0.00355378	-1.259134886	0.582152367
AGE85_89	-0.801435446	-0.589945038	0.630755512	-0.100643473	0.279713635	0.445878403	-0.569908882	0.093133626	-1.042479356	0.464733002
AGE90_over	-0.553174725	-0.514307356	0.530331779	-0.106750459	0.203413034	0.366052553	-0.462946286	0.091254259	-0.834213986	0.346887194