Understanding Energy Contexts: An Assessment of Emerging Methods for the Thermo-Behavioural Characterization of Residential Households

by

Gordon Stephen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Environmental Studies
in
Environment and Resource Studies

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Unlocking the full potential of residential-sector energy efficiency gains will require the efforts of external agents (whether in the public, private, or not-for profit sectors) engaging with individual homeowners in order to encourage the adoption of energy-saving measures. To achieve this result efficiently and effectively, such agents require an easily-obtained understanding of the "energy context" governing a household's energy use and efficiency investment decisions: factors from the number, characteristics, attitudes, and values of occupants to the physical state of a dwelling to broader geographic, financial, and legal considerations. Continuously-emerging sources of contextual and household-specific data have the potential, if integrated appropriately, to provide this understanding – but to what extent can this be achieved with current methodological tools, and can the state-of-the-art be improved?

This research has attempted to address this question, with an emphasis on the physical characteristics of homes and the behavioural patterns of their occupants. A review of existing characterization techniques in the literature yielded a set of methodological best practises and theoretical shortfalls, which were integrated with physical first principles and empirically-observed statistical trends to develop new modelling approaches to make use of hourly whole-house electricity consumption data, aiming to improve upon the state-of-the-art. A subset of these models (chosen for their speed and stability of parameter estimation) were compared to existing techniques: while one of the novel approaches yielded improved behavioural disaggregation performance and a simpler formulation compared to existing alternatives, there would seem to remain considerable opportunity for improvement, with results suggesting several potentially-promising areas for further research.

Acknowledgements

I am grateful to Ian Rowlands for his frequent and thoughtful feedback on early iterations of this research and manuscipt, to Paul Parker for his feedback on later drafts, and to Jatin Nathwani for his willingness to serve as an external examiner on my thesis committee and provide additional feedback. I am also thankful to my colleagues in the Sustainable Energy Policy group for providing a supportive and stimulating academic environment from which to conduct this work.

This research was supported by the Energy Council of Canada through an Energy Policy Research Fellowship, the Natural Sciences and Engineering Research Council of Canada through an Alexander Graham Bell Canada Graduate Scholarship, and the University of Waterloo through a President's Graduate Scholarship.

Data for this research were provided by the Energy Hub Management System (EHMS) project. The EHMS project was supported by the Ontario Centres of Excellence, Hydro One Networks Incorporated, Milton Hydro Distribution Incorporated and Energent Incorporated. It engaged, among others, graduate students and other members of the Sustainable Energy Policy group at the University of Waterloo, who helped to advance the goals of the project and to inform thinking on these issues more generally.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Energy in society

Access to a stable supply of useful energy has been a key factor in the development of human society throughout history – from the discovery of fire to the agricultural revolution, the coal-fired industrial revolution, the nuclear age, and beyond. The increasing availability of energy to perform work has stimulated growth in material production and consumption, further increasing demand and incentivizing newer and increasingly ecologically-intensive ways of extracting energy from the biophysical systems in which humanity exists.

The role of energy in society and the ecological implications of harnessing it extend across scales and systems, profoundly impacting both local socio-ecological systems and global networks, from the basic provisioning of material sufficiency and employment to the effects of resource extraction activities on health and biodiversity and international conflict stemming from the depletion and scarcity of global sources and sinks.

These impacts are not independent: actions at one scale are linked through a complex system panarchy (Walker et al., 2012) to coupled socioeconomic and biophysical systems at levels both above and below. Human prosperity (as well as that of all other life on Earth) is critically dependent on the state of these interdependent systems. Anthropogenic energy extraction and use is at present a major driving force in moving global system states away from the "safe operating space" defined by its finite biophysical limits (Rockstrom et al., 2009). Indicators such as the Human Appropriation of Net Primary Productivity (Bishop et al., 2010) explicitly demonstrate the increasing impact of human development on global biophysical energy flows.

The single-greatest sustainability challenge posed by energy use may be its contributions to climate change: sixty-nine percent of global anthropogenic greenhouse gas emissions are attributable to extracting and harnessing energy (International Energy Agency, 2014a). However, local energy-related air and water pollution also place significant stresses on the health of individuals in diverse socioecological systems and lead to biodiversity loss, while political instability and global socioeconomic inequalities often relate to energy security and a lack of affordable energy access for economic development (International Institute for Applied Systems Analysis, 2012).

Fundamental shifts in the nature and scale of societal energy use are necessary to bring material (and thus energy) consumption into balance with the planet's biophysical limits (Jackson, 2009) in the face of increasing world population and individual levels of affluence. Unfortunately, navigating the transition to a sustainable global energy system is by no means a trivial exercise. Addressing the depth and breadth of energy issues and the far-reaching implications of radical change in a system so fundamentally linked to society's socioeconomic prosperity requires conscientious and systems-aware innovation that spans domains of knowledge (Wickson et al., 2006) and scales of governance (Leach et al., 2012). As with almost any global system, the diverse priorities and often-entrenched interests of the stakeholders involved have historically prevented quick and decisive action.

While the specifics of energy transition strategies vary widely and are frequently controversial, all can be fundamentally reduced to two core components: reducing the negative impacts of harnessing energy (moving to **sustainable energy production**) and reducing the amount of energy required to deliver value to end-users (increasing **conservation and energy efficiency**). While both will be vital and complementary elements of ongoing energy system transitions (International Energy Agency, 2014b), this thesis will focus on the latter: demand-side efforts into reducing the raw inputs required to deliver satisfactory energy services represent high-impact, immediate, and economically-attractive contributions to addressing energy sustainability challenges. However, the means by which efficiency gains can be made are highly variable across energy end-use contexts, and require case-specific considerations. Such considerations are discussed in the next section.

1.2 End-use energy efficiency

1.2.1 Energy use across sectors

Societal energy use is spread across numerous sectors, most notably transport, industry, commercial/institutional, and residential. Figure 1.1 presents total energy end-use across these sectors at three scales: within Canada, across OECD (wealthly, developed) countries, and globally. Differing social, economic, geographic, and technical factors drive widely varying energy consumption patterns across and within each of these sectors: for example, energy delivery in the transport sector is dominated by portable, high-energy-density oil products, while large, stationary industrial processes draw on a more diverse supply mix including lower-cost coal. Households in developing countries obtain more energy services from simpler biofuels such as wood, while residential and commercial consumers in developed countries draw on a more convenient supply mix dominated by high-quality electricity and natural gas (International Energy Agency, 2013a).

This heterogeneity across nations and sectors results in highly variable requirements and motivations for achieving energy efficiencies: there is no "one-size-fits-all" approach to reducing energy use while maintaining energy services. Perhaps the most important "non-panacea" to recognize is the free market: while highly-energy-intensive industries may be

Figure 1.1: Share of energy use by sector in Canada, OECD countries, and globally for the year 2013. Total consumption for each jurisdiction is reported in exajoules. (Data from International Energy Agency (2013a))

strongly motivated by economic incentives for efficiency gains, other profit-driven commercial energy users may be less sensitive to reduced energy costs relative to non-energy expenses. Energy markets that fail to fully price the environmental and social externalities of consumption further weaken economic incentives for efficiency. While the invisible hand has driven significant innovation historically, it may not be a reliable or sufficiently expedient tool for achieving the widespread efficiency gains required to mitigate and adapt to ongoing and potentially irreversible environmental degradation. This market failure may be most pronounced in the residential sector, where individual actors are rarely rational profit-maximizers (Wilson and Dowlatabadi, 2007).

1.2.2 Challenges and opportunities in the residential sector

Challenges to residential efficiency

The residential sector constitutes a significant share of society-wide energy use (over 19% in Canada and 25% globally - see Figure 1.1), with correspondingly significant opportunities for efficiency gains, it has historically proven more resistant to the implementation of effective conservation initiatives. Figure 1.2 illustrates one Ontario electricity savings estimate that identifies roughly equal technical (theoretical) potential for future energy savings across the industrial, commercial, and residential sectors. In spite of this theoretical parity, estimated achievable residential potential is considered to make up a much smaller proportion of total achievable savings in both the short- and long-term (ICF Marbek, 2014).

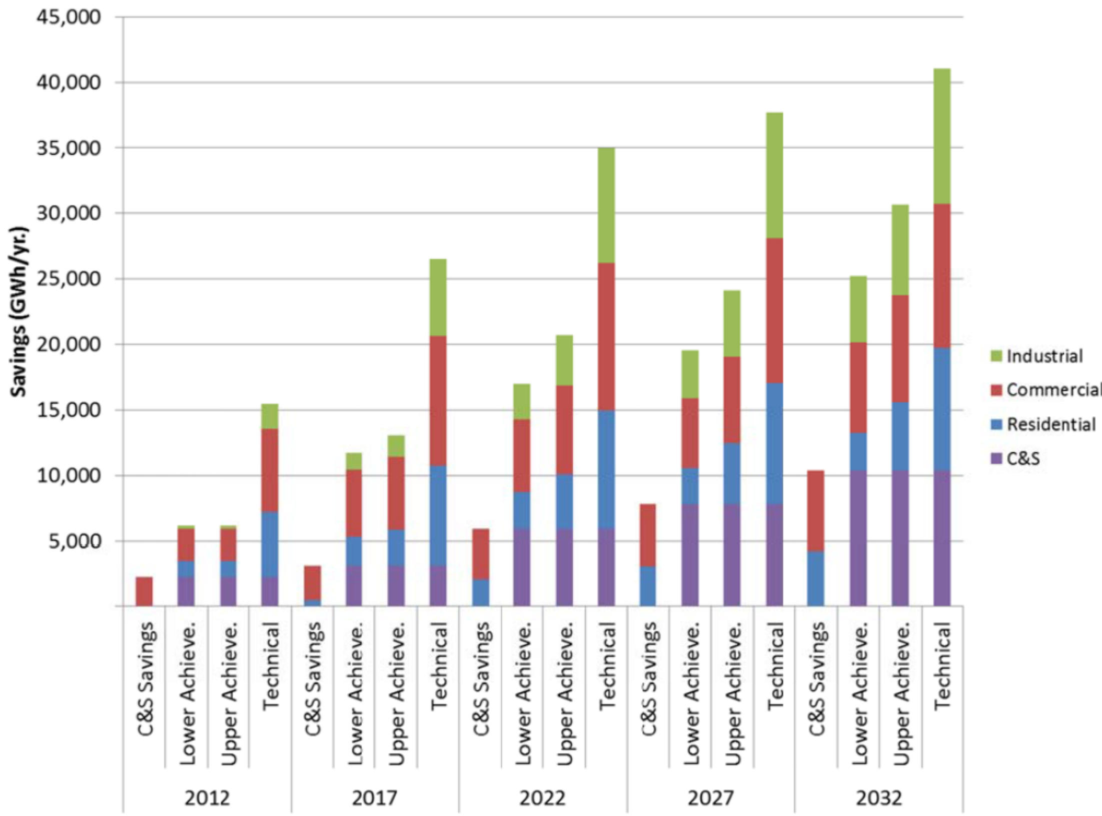

Figure 1.2: Technical vs achievable potential projections across Ontario sectors (ICF Marbek, 2014).

While the purely financial case for energy efficiency and conservation investments in commercial and industrial settings has been made (demonstrated by the existence of energy service companies, virtual energy audit services, etc), a significant "efficiency gap" (Hirst and Brown, 1990) exists between economically attractive efficiency investments in the residential sector and demonstrated adoption. Many homeowners may weigh social norms or environmental considerations much more heavily than favourable economic arguments when considering energy efficiency investments - if such investments are even considered in the first place.

A lack of energy education, access to capital, time, and trust on the part of homeowners may present barriers to adopting energy efficiency investments (Parker et al., 2003). From the perspective of private sector entrepreneurs, higher customer acquisition costs, increased variability in property and occupant characteristics (and thus ability to deliver savings) and lower quality of readily available data can present a more challenging business environment relative to serving the commercial and industrial sectors (Wesoff, 2014).

Co-benefits to improved residential efficiency

While economic arguments alone may be insufficiently effective in incentivising widespread adoption of residential energy efficiency measures, such measures can provide numerous co-benefits beyond the direct benefits of reduced energy demand in homes. Financial savings are leveraged into social benefits for economically vulnerable households who may struggle to pay energy bills while providing for other basic necessities. Energy retrofitting represents both a long term community investment and stimulates inherently local economic activity and job creation. Reducing consumption allows homes and communities to obtain a greater proportion of their energy needs from locally-generated sources, increasing resilience to exogenous supply influences and keeping funds in the community.

1.3 Understanding "energy contexts"

Residential households represent a highly heterogeneous population with diverse priorities and varying potential, willingness, and ability to make and benefit from energy efficiency investments. Successfully engaging with residents in conservation programs requires an understanding of individual situations and values – the "energy context" in which a household operates. Understanding this context can be a significant asset when working to drive efficiency gains, but obtaining such knowledge is not necessarily a trivial undertaking. Thankfully, new tools are emerging to assist in this process.

Here the term "energy context" is defined as a shorthand for the broad set of drivers that influence a household's energy consumption and energy-related decisions, in terms of energy services procured, the commodity energy required to deliver those services, and the willingness and ability of the household to change energy use patterns through technical or behavioural means. These factors can include the number of occupants in a home and the nature of their occupancy, individual occupant characteristics, attitudes, and values, the physical state of a dwelling and the nature of its subsystems and broader geographic context, and financial and legal factors that may constrain or motivate occupant decision-making.

An understanding of a household's energy context can provide insights into the incentives most likely to motivate the household and can help guide occupants to take the most effective actions possible to reduce energy consumption in their specific situation (Wilson and Dowlatabadi, 2007).

These benefits can be operationalized through better targeted household engagement and outreach in residential energy efficiency programs, and tailored suggestions and feedback in behavioural programs, while an understanding of household-level behavioural patterns or physical characteristics can help create more representative control groups in intervention studies. Households can use elements of their own energy contextualizations to better understand their energy use and be better informed about costs and benefits of energy-related investments or participation in future transactive energy markets based on their specific situations.

Society's energy generation and delivery systems are presently undergoing technical transformations from relatively static, centralized distribution channels to integrated, intercommunicating networks of distributed energy resources, producing and consuming energy dynamically in response to shifts in demand and supply – the "smart grid" (Farhangi, 2010) and "smart energy networks" (Belanger and Rowlands, 2013) more generally. These technical innovations are widely expected to result in increased delivery network resilience, allowing deeper integration of renewable generation and energy storage technologies, and enabling energy efficiency gains through a deeper understanding of end-user consumption trends and issues, increasing the overall sustainability of energy systems.

In particular, the increasing availability of household hourly electricity demand data in smart-meter-transitioned jurisdictions (including Ontario) is beginning to provide rich data sources for examining both physical and social household characteristics (Depuru et al., 2011). Taken with other forms of increasingly available data, including weather history, municipal mapping resources, census demographic information, and real-estate or property assessment records, there is unprecedented potential to contextualize and break down a particular household's energy use through rapid, inexpensive, indirect inferential techniques.

1.4 Research question

Limiting the effects of anthropogenic climate change to a 2 °C temperature rise, as targeted in the 2015 Paris Climate Agreement, will require significant gains in worldwide energy efficiency over the next decades (International Energy Agency, 2013b). Given global society's current social, economic, and technological context, it is unlikely that traditional, consumer-driven market forces alone will be sufficient to unlock the full potential of residential sector efficiency gains at the rapid pace required: a broader portfolio of approaches would involve external agents (whether in the public, private, or not-for profit sectors) engaging with individual homeowners in order to actively encourage the adoption of efficiency measures. To achieve this result efficiently and effectively, such agents require an easily-obtained understanding of the context governing a household's energy use and efficiency investment decisions. Continuously-emerging sources of contextual and household-specific data may have the potential, if integrated appropriately, to provide this understanding or "energy context" – but to what extent can this be achieved with current methodological tools? Is the state-of-the-art sufficient, or can it be significantly improved?

This manuscript will attempt to address these questions, with a particular focus on the scalable inference of physical characteristics and occupant behavioural trends in residential homes. Chapter 2 will review existing means of household characterization and provide more detailed comparisons and critiques of approaches judged to be the most promising in the context of real-world social, economic, and technical constraints. Chapter 3 will combine the insights from this theoretical assessment with empirical results from the analysis of a representative dataset to inform the development of a novel technique for household characterization. Chapter 4 will provide an empirical assessment of the performance of this characterization approach relative to the existing literature. Chapter 5 will discuss the implications of these results with reference to possible applications, and Chapter 6 will summarize insights from the completed work and outline areas for future research.

Chapter 2

Literature Review

There are a wide range of possible approaches to gaining information about a household's energy context (as defined previously in Section 1.3). This chapter will review historical and current techniques and best practises as established by the existing literature, beginning with general characterization approaches and gradually focusing in on specific techniques in energy data analysis.

2.1 General methods for household characterization

2.1.1 Traditional direct characterization methods

Traditional techniques for characterizing the energy-use context of individual households are the simplest and most direct, measuring exact characteristics of interest. Unfortunately, the price for such clarity is the expense and time commitment required to collect the relevant data, limiting the scalability of such approaches. While the indirect approaches identified in the next section are likely to have much lower marginal costs of data acquisition and be more applicable at larger scales, at their best they will generally serve to reconstruct information that could have been acquired directly through traditional means. As such, it remains important to understand the value and potential of these traditional direct, high-marginal-cost approaches, including novel extensions that may serve to leverage their results for greater impact moving forward.

Audits

An energy audit is a process in which a trained professional physically inspects a dwelling to determine how various subsystems, including heating and cooling, contribute to the building's energy consumption, and identifies opportunities for improvement. Performing an on-site audit is clearly the most direct and accurate means of assessing a dwelling's physical energy use characteristics. Unfortunately, the process is somewhat slow and expensive (a typical audit duration is on the order of hours and costs several hundred dollars), while also intruding on the homeowner's schedule and privacy. As a result, uptake among homeowners tends to be low, even when cost barriers are reduced through subsidized programs. For example, an early review of US utility-sponsored audit initiatives by Hirst et al. (1981) found that most programs had less than 5% participation rates, and regional studies suggest uptake has not increased considerably in the intervening 30 years (Song, 2008).

Rather than treating audit insights as a desired final output of deductive inference (the general premise of approaches in subsequent sections), some studies have used physical characteristics of specific households inductively to better understand average impacts on energy consumption. For example, Ndiaye and Gabriel (2011) used a sample of audit results in combination with surveys (see below) to infer the independent average impact of individual household characteristics on overall consumption.

One potentially promising opportunity that has generally not been investigated in the literature to date is the concept of using a relatively small sample of household audits to predict the physical characteristics of specific dwellings in a wider population, based on trends in more readily-accessible indicators such as the data sources for the low-marginal-cost characterization techniques outlined in the next section. This would enable a smaller quantity of costly-to-acquire audit data to be leveraged into specific yet widely-applicable insights on the characteristics of many more dwellings, in much the same way that automated property valuations are performed currently: Municipal Property Assessment Corporation (2015) provides one such example. While such a concept is beyond the scope of this research, it merits further investigation.

Surveys and Interviews

Ultimately, energy consumption is only a means to the specific end of provisioning desired energy services. While audits are effective tools for collecting precise data about physical building characteristics, taken alone they do not provide explicit information about the requirements, motivations and behavioural tendencies of household members and the desired outcomes of their energy use, key factors in developing a complete understanding of a household's energy context. Household surveys, or their sophisticated-but-costly counterparts, occupant interviews, can provide more formal, direct insights into the social and behavioural dynamics influencing a household's energy use.

A large body of work has developed relating a household's social characteristics (determined via surveys) with energy-use and conservation uptake behaviours. Guerin et al. (2000) provides a meta-analysis of 45 such studies and identifies multiple consistently-significant social factors influencing energy behaviours, including income, age, education, and home ownership. Shi (2011) and Sanquist et al. (2012) both investigate US Residential Energy Consumption Survey data to establish links between lifestyle factors, economic conditions (income, commodity prices, etc) and total energy consumption.

As referenced previously, Ndiaye and Gabriel (2011) combined survey data about social characteristics of households with audit data describing physical characteristics to understand the relative impact of various factors on total energy consumption. While the goal of that research was to identify the most significant contributors to aggregate household consumption at a population average level, such findings could also inform characterizations of households given sociodemographic survey results.

While surveys and interviews excel at collecting social data, they are perfectly capable of collecting physical data as well: Gaasch et al. (2014) investigate a physical characterization approach whereby physical properties of a building are collected through a survey or interview, then input into a building model to evaluate estimated energy performance. Natural Resources Canada's HOT2XP software is designed with similar principles, producing a sophisticated HOT2000 building energy model based on "only a small amount of critical information" (Natural Resources Canada, 2015) that can be provided by a homeowner directly.

While not usually as expensive or time-consuming to perform as physical audits, interviews are still a costly means of data collection. Surveys essentially automate the role of the interviewer (at the expense of some flexibility and interactivity), but remain intrusive for the interviewee. Since both rely on household self-reporting, they can also suffer from reduced accuracy or precision compared to other characterization methods, particularly in relation to specific quantitative or technical data. However, no other technique assessed here can provide the same accuracy and depth of social data at the individual household level, an important consideration given the highly heterogeneous nature of the residential energy sector.

2.1.2 Emerging indirect characterization methods

The availability of increasingly rich datasets and the ability to combine these in novel ways has created the potential for inferring household energy characteristics via highly-scalable, low-marginal-cost indirect analysis techniques. While the quality of resulting estimates will never be as reliable as a direct measurement, the ability to apply the resulting insights to deliver the kinds of benefits discussed in section 1.3 across large populations may often outweigh inferential uncertainties.

Furthermore, new kinds of data coupled with novel analytical techniques have the potential to enable richer characterizations then have previously been possible via traditional direct measurement methods. For example, Hutchinson et al. (2006) studied integrating known dwelling (physical) and household (social) characteristics with the goal of detecting potential underheated homes, but concluded that the approach lacked sufficient predictive power to provide meaningful targeting capabilities. It is reasonable to suspect that the availability of high-resolution consumption and temperature data would significantly enhance the ability to identify specific thermal situations - what's more, a smart thermostat could report this information directly, eliminating the need for probabilistic inference altogether. Four categories of indirect, data-driven characterization approaches are discussed here.

Household-level public data sources

While physical or social household data can be costly to collect directly, some characteristics are already collected by various institutions and may be publicly accessible. For example, Ontario's Municipal Property Assessment Corporation (MPAC) maintains detailed information on physical characteristics of residential dwellings (size, age, number of floors, dates of renovations, heating and cooling equipment, etc), which can be accessed by any member of the public, for a fee. Similarly, Hogör and Fischbeck (2015) used Florida voter registration and property tax records to determine significant household-level social and physical predictors of potential for energy-efficiency gains, allowing for the identification of specific homes with high savings potential.

Aggregate social and geographic profiling

As previously established, socioeconomic factors play an important role in understanding a household's energy context, but surveys (traditionally the primary vessel for learning about such factors) are not a particularly scalable means of information acquisition. While explicit, household-level information is clearly optimal for forming a household energy profile, it is not always available: in such cases, population averages of geographic areas (such as neighbourhoods) can often serve as reasonable approximations to the individual values.

Such a claim is not meant to dispute the ecological fallacy, which describes the perils of assuming that the characteristics of individuals in a population are uniformly identical to population averages. It only serves to note that, at sufficiently-small geographic scales, the effects of homophily (the tendency for similar individuals to be more closely linked) can cause relevant household-level social characteristics (occupant count, age, income, education, etc) to correlate with the characteristics of other nearby households. The same will also tend to be true for physical dwelling properties (age, size, heating systems, etc), although for more practical reasons (historical subdivision development, building codes and available technologies at the time of construction, etc).

Such local homogeneity makes it possible to perform useful analyses based on publicly-available aggregate social data (such as census records or polling station voting records) rather than information on individual households. For example, Morrison and Shortt (2008) combined census-derived aggregate sociodemographic data with sampled household-level physical characteristics to develop a fuel poverty risk indicator based on dwelling location in a community. Similarly, Song (2008) and Zhao (2013) both studied the influence of social factors (based on neighbourhood-level census statistics) on participation rates in local energy audit programs.

Remote sensing

Certain physical household characteristics can be observed remotely: for example, the integrity of a building's thermal envelope can be assessed according to the rate of heat loss through a dwelling's exterior surface area. Hay et al. (2011) combine aerial photography with thermal imaging to generate roof heat loss estimates for individual buildings on a city-wide scale. Similarly, several startup companies currently offer vehicle-based drive-by thermal imaging services in the same vein as Google StreetView (LaMonica, 2013). Coupled with appropriate image recognition software, such services have the potential to provide audit-style building envelope assessments without the need for an auditor to physically enter the premises.

Household-level private data sources

The rise of the "smart" or "connected" home has begun to result in an increase in volume and availability of data collected by various household systems. While it remains to be seen how specifically the smart home will develop, it is clear that the information measured and collected by such services has the potential to be incredibly valuable in the automated deduction of a household's behavioural and physical energy context.

At this early stage, one of the most commonly-adopted connected home technologies is the smart thermostat, an internet-connected device with the capability to perform more sophisticated temperature and energy use optimizations on behalf of the homeowner. Of particular interest in the context of developing a household's energy profile is the ability for units to log interior temperature, heating and cooling system states, and even occupant activity levels - indeed, it is through analysis of these data that such devices are able to gain a nuanced understanding of a household's energy needs and optimize heating and cooling loads accordingly.

Analysis of data from smart home appliances is unique relative to the other low-marginal-cost characterization approaches described above in that it requires the customer to purchase and install specific hardware in the household ("behind the meter", as described from an energy utility perspective). As such, unlike previously-considered approaches, such data are not readily available to an external analyst, nor are they even collected uniformly (not all households will be willing or able to adopt such technology, for instance). As such, while the study of information-rich "private" household-level data sources is no doubt a fertile approach to household characterization, it is particularly difficult to achieve uniform scaling and equitable distribution of the resulting benefits across all households.

Energy consumption analysis

While the previous characterization approaches have focused on observing upstream causes contributing to a household's energy context (building characteristics, behavioural patterns, geographic and socioeconomic contexts, etc), another obvious option is to study its downstream effect: that is, energy consumption itself. Unsurprisingly, studying energy use data can provide significant insights into a household's overall energy context. The introduction of advanced metering infrastructure quantifying household electricity use at hour, minute, or second timescales has made highly information-dense data on this topic readily and uniformly accessible, both to utilities and household occupants. As such, the vast majority of current energy characterization work focuses on the analysis of interval electricity data, and it is seen as a key to unlocking significant physical and behavioural energy savings in the future (Armel et al., 2013).

Companies like OPower, Bidgely, and PlotWatt leverage massive whole-house electricity datasets and advanced analysis techniques to provide deeper and more scalable engagement with households through the identification of directly-actionable opportunities for improved energy efficiencies (Wesoff, 2014). While this business model has been largely proven in the commercial and industrial sector, its viability in the diffuse and heterogeneous residential sector has yet to be established. The growth focus of venture-backed companies also tends to favour scalability and consistency of service across jurisdictions over a grassroots community-specific approach, which may result in optimizing economic returns at the expense of program customization, effectiveness, and integration with the local initiatives acting on characterization insights (energy efficiency programs, community social marketing initiatives, etc).

While the private sector has important contributions to make to the household characterization challenge, it should not be the sole solution provider. A significant body of academic literature also exists in this area, which will be reviewed in the following section.

2.2 Approaches to energy data analysis for household characterization

This section will outline multiple approaches to energy data analysis for household characterization as outlined in existing literature. In particular, intrusive monitoring, intrusive training with non-intrusive monitoring, and non-instrusive training and monitoring techniques will be compared.

2.2.1 Descriptive appliance-level analysis (intrusive monitoring)

Direct load monitoring of individual (labelled) electrical circuits provides a significantly richer description of behavioural and physical characteristics that constitute a household's overall energy use and context. Of course, such information is what is described above as a private data source: it requires specific hardware to be installed in individual dwellings (at significant cost) and so is not collected universally across all dwellings.

Nevertheless, the data extractable through such "intrusive" observation techniques can provide significant and often-generalizable insights into consumption trends, occupant behaviours, and resulting policy implications (Koksal et al., 2015), or allow for the automation of strategic shifts or reductions in consumption (Bozchalui et al., 2012). As such, while the analysis of intrusively collected data is generally not a scalable direct characterization technique, from a research perspective it has attracted significant attention, including for the purposes of validating many of the more scalable characterization processes described in subsequent sections. For example, Kolter and Johnson (2011) collected high-resolution appliance-level load data from multiple households for the purposes of training and testing non-intrusive load monitoring techniques - the dataset has subsequently been applied as a standard benchmark for evaluating high-resolution disaggregation techniques.

Disaggregated household load data can contribute more to characterization efforts than just a ground-truth reference for evaluating disaggregation algorithms, however. Rowlands et al. (2015) provide an overview of 13 different intrusive load studies with key findings including physical and social predictors of specific load uses (applying many of the previously described household characterization techniques at the disaggregated appliance level), seasonal and geographic trends in appliance use patterns, relative magnitude of various household loads, and variations in appliances' load cycle efficiencies.

2.2.2 Induction and inferential load analysis (intrusive training, non-intrusive monitoring)

One approach to striking a balance between the data-richness of intrusive load monitoring and its inherent lack of scalability is to intrusively collect data from a representative sample group of households or over a limited period of time, and then inductively generalize any observed trends to a broader population or observation period (the sample household approach is in fact the implicit aim of much of the intrusive load monitoring research reviewed by Rowlands et al. (2015) and described above).

Non-instrusive load monitoring

The field of "non-intrusive load monitoring" (NILM) provides the most actively-researched example of such a characterization approach. Traditionally, NILM has involved the collection of appliance power load profiles at high temporal resolutions (on the order of fractions of seconds or smaller intervals, and thus requiring specialized monitoring hardware), and the subsequent training of statistical models to disaggregate matching load signatures from whole-house aggregate consumption data (measured at equally high resolutions). Berges et al. (2010) and Zeifman and Roth (2011) provide reviews of specific statistical techniques for achieving this.

More recently, however, the rise of smart metering, making interval data widely available but at much coarser timescales than traditionally studied, has motivated the development of NILM techniques that can be applied at much lower sampling rates. Unforunately, while smart meters provide a widely-distributed non-intrusive aggregate monitoring platform, the process of collecting the required disaggregate training data remains costly and not easily scaled across households, for reasons identical to those discussed in the previous section. Multiple approaches have been proposed to address this challenge. Basu et al. (2015) are among those who propose a temporary intrusive approach whereby a household's disaggregate loads would be explicitly monitored via submetering for an initial training period, with the collected data informing future statistical disaggregations. However, given that the costs of disaggregated data collection are typically dominated by equipment installation, this approach does little to enhance scalability. Others, such as Rodríguez Fernández et al. (2015) make use of a centralized repository of submetered appliance data that can then be used to classify aggregate loads in other households. In this situation, however, variability in appliance profiles across households may reduce disaggregation performance. Finally, as a compromise between training data cost and quality, some approaches involve direct occupant participation in an initial training exercise, where known appliances are activated at specific times and the resulting profiles are extracted from the household's aggregate consumption data (Griffiths, 2015).

While recent advances in NILM have provided reasonable disaggregation performance at comparatively low sample rates, the fundamental fact remains that significant information losses occur when electricity consumption is aggregated to hourly intervals (compared to minute or second intervals), imposing fundamental limits on the ability of NILM techniques to perform reliably at the temporal resolution typical of currently deployed smart meter systems. While it has been suggested that existing systems could be adapted through remote software updates to record consumption at more frequent intervals (Armel et al., 2013), it not clear that utilities would be sufficiently motivated to do so, given the the effort required to both perform the transition and manage the significantly larger volume of data that would result. As such, harnessing NILM techniques for reliable characterization may ultimately require household technology investments (such as in-home display products to collect smart meter data at more frequent intervals) that would prevent scalability and accessiblity objectives from being achieved.

Instrusive training on non-energy characteristics while classifying on aggregate energy data

The idea of collecting some "training" dataset via intrusive means in order to calibrate a non-intrusive energy-use characterization model is by no means limited to electrical disaggregation work and the collection of specific appliance signatures. For example, after "instrusive" collection of the likelihood of households participating in an energy efficiency program (by offering the program to a subpopulation and recording household uptake), Zeifman (2014) developed a model to predict participation likelihood given a household's energy consumption profile. Generally speaking, in situations where costly-to-obtain but representative sample data already exists (survey or audit results, for example) and can be linked at a household level to more readily-available data with sufficient discriminatory power (such as hourly electrical loads), an "intrusive training, non-intrusive monitoring" approach presents a valuable potential tool for household characterization. In practise, however, these data availability conditions are not always easily satisfied.

2.2.3 Correlationary load analysis (non-intrusive training and monitoring)

Useful insights can often be gained from comparing or cross-referencing a household's electricity consumption with other accessible data sets, avoiding the need for intrusive information collection. Such analyses may use electricity data from other households to identify likely similarities in energy profiles (clustering) without necessarily needing to understand what sociotechnical characteristics a specific represents. Alternatively, physical domain knowledge of causal influences on energy use may be used to leverage non-energy observations to generate a contextualized characterization of a household.

Load Profile Clustering

Socially- and technically-similar households tend to reflect their likeness in their electricity use patterns. For example, dwellings that are vacant during the day (while occupants are at work or school, for example) will often have repeatedly lower consumption around midday and higher consumption in the mornings and evenings, while homes that are occupied all day will have a different daily load "shape". Similarly, heavily air-conditioned homes in the same region will experience contemporaneous demand spikes during summer afternoons, while electrically-heated homes will increase their loads in concert on winter nights. Other statistical metrics besides direct consumption levels can be used for grouping as well: for example, electricity users better able to take advantage of demand response programs may reveal more variability in their loads over the course of the day (Jang et al., 2016).

Given electricity consumption history for a group of households, these patterns can be exploited to estimate which houses have the most similar overall energy contexts - using similarities in load profiles as a proxies for similarities in deeper, harder to measure sociotechnical characteristics. A number of statistical techniques exist for performing this analysis: specific methods for extracting representative features from consumption history and grouping (clustering) the results are reviewed and compared by Chicco (2012). In each case, the focus remains on grouping likely-similar households, not necessarily understanding the nature of the defining characteristics of a group (although that may have been performed implicitly as part of the feature extraction process, or could still occur as a subsequent characterization step). In some situations, such as creating control groups for electricity intervention studies, such "context-free" clustering may be the only characterization required.

Domain-knowledge-derived load characterization

In many situations, the characterization question of interest concerns the sociotechnical context of an individual house, rather than which households are most likely to be similar to one another. Several means of achieving this task have been proposed that involve leveraging contemporaneous external factors known to exert a causal influence on household electricity consumption. In the most basic case, these involve a regression analysis of electricity consumption as a function of exterior temperature (Fels, 1986; Birt et al., 2012). More sophisticated approaches have also acknowledged and accounted for the sequential nature of electricity use readings, either through autoregressive time series approaches (Espinoza et al., 2005; Ardakanian et al., 2014) or probabilistic Markov methods (Huang et al., 2013; Albert and Rajagopal, 2015).

These approaches can compensate for a lack of detailed, intrusively-obtained information about a household's energy use by pairing easily accessible information (such as weather or the time of day) with electricity data through well-established relationships (such as the effects of exterior temperature on energy use or the cyclical nature of daily, weekly, and annual consumption patterns). The details of these methods are outlined in the following section.

2.3 Domain-knowledge-based approaches to electrical load characterization

This section outlines several methods for applying domain knowledge of the properties of household electricity consumption to analyze load data in combination with exogenous data sources. In particular, thermal regression models, time-series patterns, and conditional Markov model approaches are presented.

2.3.1 Thermal regression models

PRISM

The Princeton Scorekeeping Method, or PRISM, was introduced by Fels (1986) as a means for developing weather-normalized energy use profiles of buildings, in order to more accurately quantify fuel savings resulting from efficiency retrofits. The classic PRISM approach models monthly energy use as a linear relationship between heating fuel consumption and heating-degree-day temperature records, with the fuel consumption intercept α representing temperature-independent (e.g behavioural) consumption (such as use of gas appliances) and the slope β_h representing the sensitivity of the building's energy use to changes in external temperature (reflecting factors such as heat loss rates and furnace efficiency). Both parameters are determined by an ordinary-least-squares fit to the data. The heating-degree-day reference temperature τ_h is also determined from the data via iterative square-error minimization. Stram and Fels (1986) extend the basic model to electric heating and cooling by introducing a second regression parameter β_c and cooling-degree-day reference temperature τ_c to be estimated.

While developed primarily as a feature extraction tool for intervention analyses, PRISM involves the development of behavioural and physical household characterizations (α, β, τ) as intermediate products. Researchers including Hogör and Fischbeck (2015) have since applied the PRISM characterization methodology independently of any intervention study context.

Three-lines smart meter model

While the PRISM approach to household characterization is grounded in physical considerations, the precision of possible estimates is inherently limited by the low temporal resolution of its input data. The introduction of advanced metering infrastructure recording hourly (if not more frequent) electricity use provides a significant opportunity to enhance the analytic potential of PRISM in the age of the smart meter. Birt et al. (2012) developed

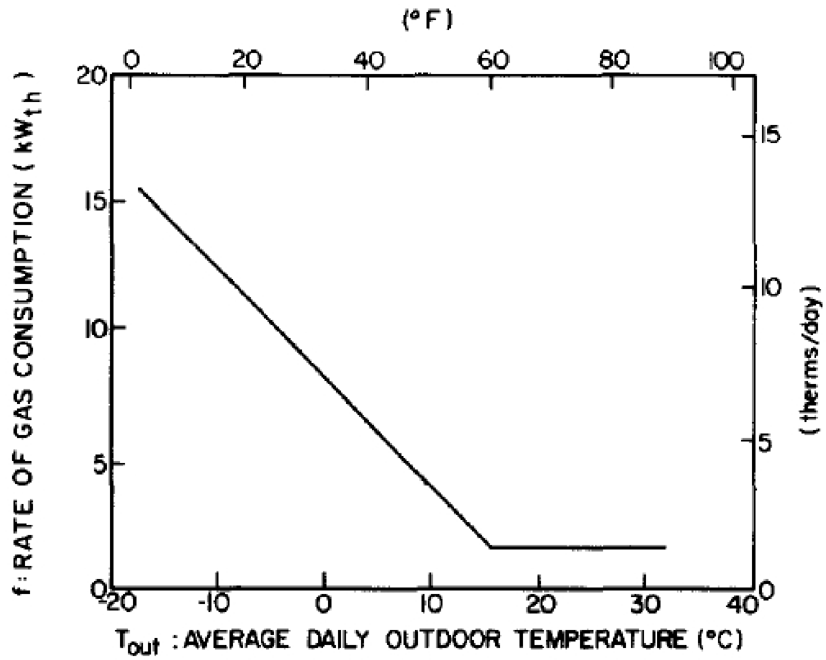

Figure 2.1: Classic PRISM approach to parametrizing a household's base fuel consumption, heating system temperature gradient, and internal reference temperature (Fels, 1986)

a three-regime (heating, passive, cooling) linear model regressing hourly household smart meter data on outdoor temperature as a means to characterize a household's key thermal properties with higher-resolution data.

In addition, rather than fitting a single trend line to hourly observations, this "three lines" approach further characterizes a household's energy profile by providing linear fits to the first, fifth (median) and ninth deciles of a household's electrical load after binning to uniform temperature increments. This approach also allows for a temperature gradient in the "passive" temperature regime (between heating and cooling).

2.3.2 Hybrid time series / regression models

PRISM and the three-lines model provide straightforward, direct insight into a dwelling's key physical characteristics (the state of its building envelope and heating systems), but the explicit behavioural insights offered are minimal - only thermal system activation thresholds and baseload or average temperature-independent consumption values are estimated. While more detailed behavioural patterns could be assessed by performing dummy variable regressions on times of day, week, or year, similar in method - although not in motivation - to the regression approach to intervention analysis taken by Newsham et al. (2011), a direct regression implicitly assumes independent consumption observations, thus ignoring the serially-correlated nature of the time series data, and fails to acknowledge the sequential behavioural patterns observable in hourly smart meter data. Accounting for the autocor-

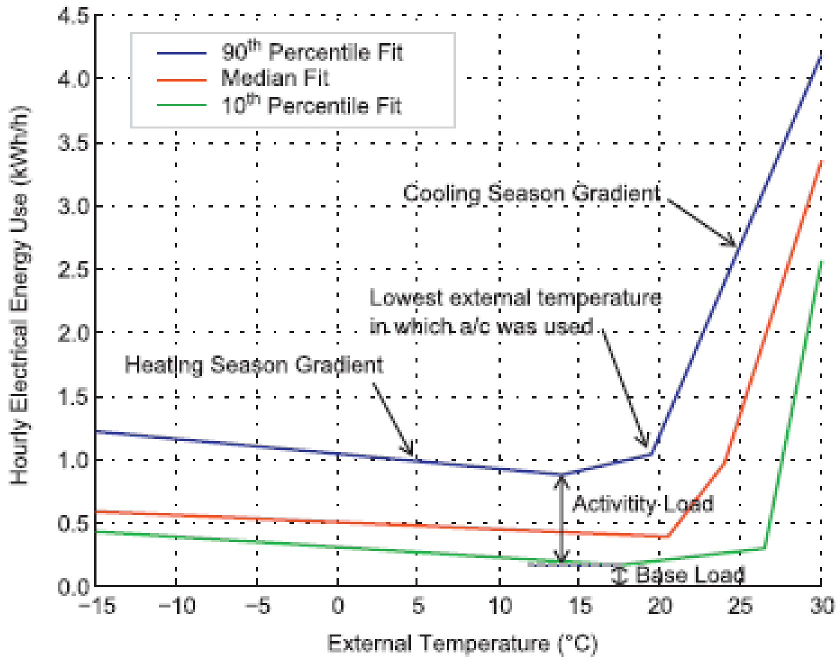

Figure 2.2: Three-regime, three-decile regression approach for parametrizing a household's heating, cooling, and base load characteristics (Birt et al., 2012)

related nature of such data becomes increasingly important at higher temporal resolutions.

Time series statistical techniques provide means for modelling this serial correlation. Such methods have historically been applied to aggregate electricity data for purposes of statistical load forecasting – Livera et al. (2011) provide one example – with less attention given to estimating model parameters that correspond to meaningful physical or behavioural indicators on a household scale. To date, those approaches that do pursue this goal of characterization (if only as one result among others) do so through a combination of autoregression (AR, where consumption values are regressed on other consumption values preceding them in a sequence) and standard regression techniques involving covariate time series (such as exterior temperature) and dummy variables (such as time of day or week) capturing cyclical or other sequentially-derived trends. Two such approaches, those of Espinoza et al. (2005) and Ardakanian et al. (2014), are elaborated here.

Periodic autoregressive time-series characterization

Espinoza et al. (2005) were one of the first to apply time series techniques to an explicit thermo-behavioural characterization effort (albeit at the electrical substation, not household, level) by estimating electrical heating load via a fixed-threshold four-regime (cooling, passive, and two levels of heating) regression on external temperatures while simultaneously modelling behavioural contributions via day-of-week and month-of-year regression dummy variables and an autoregression on the first through forty-eighth preceding observations -

what is known as an AR(48) model. As such, in addition to the effects of exogenous factors, each hourly observation is formulated as a linear combination of the hourly activity over the past two days.

Furthermore, a unique set of the 48 autoregression coefficients is fit for each of the 24 hours of the day, creating what is known as a *periodic* autoregressive model, or PAR(48). As such, the autoregressive component of the behaviour model results in $48 \times 24 = 1152$ autoregressive coefficients alone - while such numerous parameters can provide a good statistical fit to data, they contain little human-parseable insight into behavioural patterns or broader energy contexts. The implications of the estimated best-fit weight for electricity readings 35 hours previous to all readings taken at noon, for example, are at best unclear. Of course, the interpretability of model parameters is not a concern in the context of demand forecasting, the model's primary intended application, and the approach is frequently referenced in that literature (Kyriakides and Polycarpou (2007) provide one such example).

Hybrid seasonal-periodic autoregressive time-series characterization

Ardakanian et al. (2014) present several modifications to Espinoza's work. Most notably, the periodic autoregressive model is combined with a seasonal autoregressive model in which consumption in a given hour is defined not as a linear combination of values in preceding hours, but instead as a linear combination of values at the same hour in the preceding days of the same "type" (weekday or weekend/holiday). In addition, coefficient periodicity is extended from the autoregression to include the thermal regression coefficients, allowing for varying correlations between external temperature and household power consumption at different times of day (in PRISM's notation, 24 seperate β-values for each thermal regime). Two new "occupancy" exogenous indicator variables are also introduced to indicate when a household's consumption is in the highest or lowest population decile for a given hour of day and temperature (in theory suggesting that the household is abnormally full or empty). The regression coefficients for these variables are also allowed to vary according to time of day.

The choice to eschew autoregression on previous hours in favour of autoregression on previous days is interesting in that it implicitly suggests that electrical loads 24, 48, 72, etc hours ago have a direct influence on present demand, while loads one, two, or three hours ago have no influence whatsoever - a seemingly counterintuitive premise.

Finally, at first glance, the decision to allow for variable thermal regression coefficients according to time-of-day would seem to provide interesting opportunities to capture behavioural-driven influences on household heating requirements. However, from a physical perspective (as elaborated by PRISM), these coefficients correspond to the dwelling's building envelope properties and thermal system efficiency, characteristics unlikely to fluctuate over the course of a day (one exception could be the opening and closing of windows). Thermal factors likely to be both more significant and more variable, such as changes to thermostat setpoints, the influence of resistive heat dissipation from internal loads (cooking appliances, etc), or the number of occupants in the household and the nature of their activity would be reflected not in the heating or cooling gradient of a particular regime, but rather in the regime's temperature threshold (the point at which heating or cooling systems would activate) - PRISM's τ parameters.

In both of the time series models discussed above, these τ threshold parameters are not only fixed over the course of the day, but are in fact taken as predetermined and fixed across all households. While this assumption provides significant simplifications to parameter estimation procedures, it may not be justified given the dynamic influence of the factors outlined above. Inaccurate assumptions concerning these values can introduce significant bias into the models, as elaborated by Fels (1986). It seems reasonable to suggest that estimating these threshold values from available data rather than predetermining them according to "standard techniques within the energy industry" (Espinoza et al., 2005) could provide greater explanatory power and thermo-behavioural insight than is provided by the choice of Ardakanian et al. (2014) to allow for variable thermal regression coefficients.

It may also be the case that rigid temperature thresholds (predetermined or otherwise) may not sufficiently model the inherent variability in thermal system activation conditions. The next section outlines a more flexible probabilistic approach.

2.3.3 Conditional hidden Markov models

The regression and time-series approaches above all assume a primarily deterministic model of household energy use and thermal regime changes – heating and cooling systems activate predictably at specific temperatures with subsequent loads precisely related to external factors – with some residual statistical error to be minimized through the parameter estimation process. An alternate viewpoint is to consider a household's electricity consumption as fundamentally probabilistic, with the likelihood of behavioural, heating and cooling consumption values conditioned (rather than correlated) on the system state in the previous time period as well as on exogenous factors such as temperature. The conditional hidden Markov model provides an analytical framework for modelling from this perspective, and has emerged as an increasingly popular tool for electrical end-use disaggregation. The works of Albert and Rajagopal (2015) and Huang et al. (2013) are of particular relevance and are examined in more detail here.

Dynamic-regime thermal regression disaggregation

Albert and Rajagopal (2015) combine the multi-regime linear regression technique taken by the previous approaches in Sections 2.3.1 and 2.3.2 with a Markov model governing switches between thermal regimes probabilistically. At any point in time, a household in some unobserved ("hidden") thermal regime (such as heating state) has some probability of remaining in that state, and some other probability of switching to another regime (such as a passive or "thermally insensitive" state). These Markov transition probabilities are conditional on external temperature, so in cold weather a household may be very likely to enter a heating state and stay there, but as temperature rises the odds of switching to a passive or cooling state will increase as well. The combination of current state and temperature together specify a probabilistic distribution the household's immediate heating or cooling load. The parameters governing the conditional switching probabilities, somewhat analogous to the τ parameters of previous models, are fit according to the observed data. Additionally, rather than predetermining the number of regimes of each type (for example, two heating, one cooling, and one passive), an arbitrary number of total regimes can be created, with types subsequently specified according to what would best model the observed trends in a given household. This corresponds to allowing the β parameters of previous models to take on positive or negative values, with the sign of the coefficient indicating whether the regime is a heating, cooling or passive state - where passive states correspond to near-zero β values.

This approach provides a richer, less prescriptive means of describing a household's thermal characteristics that may better account for non-deterministic behavioural influences on otherwise-mechanistic thermal system operations, as suggested by the results of the authors' favourable performance assessment compared to Birt et al. (2012) using the REDD dataset of Kolter and Johnson (2011). However, it does so at the expense of parametric structure, reducing the direct interpretability of individual physical parameters. In addition, temperature-independent loads are described only by state-specific constant values, still ignoring cyclical behavioural factors of interest (such as weather-independent daily, weekly, or seasonal load profiles). While implicit insights could likely be derived from the nature of specific regimes and patterns in their temporal occurrences, no explicit consideration is given to these key behavioural (non-thermal) factors.

Thermodynamically-derived thermal load disaggregation

The models considered up to this point have all made the implicit assumption that heat loss (and thus heat delivery required of a thermal system) is directly proportional to external temperature. This is, of course, an oversimplification: conductive heat transfer theory stipulates that heat flow in or out of a building is proportional to the *difference* between internal and external temperatures. No models yet considered have acknowledged the fact that internal household temperatures also vary over time, and can have significant effects on energy requirements. A well-insulated household in the depths of winter may only require a small amount of supplementary active heating if it is already warmed to room temperature, while a drafty disused building next door may require enormous amounts of energy to restore and maintain habitable temperature levels.

Interior temperature is an unobserved system variable that, like a household's thermal state, is well suited to representation via a hidden Markov model framework. Huang et al. (2013) build on previous NILM work and apply a simplified thermodynamic model to govern interior temperature evolution and thermal regime transition probabilities. Unlike the previous dynamic-regime Markov approach, heating and cooling states are prespecified. In addition, and in contrast to all other models reviewed here, a given thermal state is assumed to deliver a fixed amount of heating or cooling power when active, irrespective of external temperature. While such an assumption is likely a more accurate representation of HVAC system dynamics at very high temporal resolutions (on the order of minutes or seconds), it may be less justifiable in the context of hourly readings. Finally, as in the previous case, no explicit consideration is made for the influence of cyclical occupant behavioural patterns.

One shortcoming of Markov modelling techniques (relative to regression or state-space time series approaches) is the lack of a direct means of dealing with missing sequential observations. This lack of robustness can prevent the application of these techniques when working with datasets involving significant data gaps, where imputation or other filling techniques may not be a reliable means of reconstructing missing values.

2.4 Summary

This chapter has reviewed existing methods for observing or inferring elements of a household's energy context, with a specific focus on scalable, non-intrusive methods. In particular, several specific energy data analysis techniques have been reviewed, providing an assessment of their relative strengths and shortcomings. In almost all cases studied, regression of thermal-related loads on exterior temperatures has figured prominently in disaggregation models. In the next chapter, insights from studying these models will be applied in combination with fundamental physical relationships and empirical observations in an attempt to formulate a modelling approach that improves on the status-quo.

Chapter 3

Model Development

The preceding review of existing characterization techniques in the literature identified specific strengths and weaknesses of current approaches, in particular for the case of the analysis of smart meter electricity data. This assessment of the state of the art suggests that there remains an opportunity to improve household energy characterization efforts by better integrating thermal models grounded in a physical understanding of building heat flow with behavioural models informed by established domain knowledge and statistical analyses of occupant patterns.

This chapter will seek to develop a residential electricity end-use model espousing this integrative approach by explicitly outlining the key elements of such a model. A set of foundational assumptions regarding the nature of household energy use will be posited and justified both theoretically and through the application of empirical data. Finally, thermal, behavioural, and joint thermo-behavioural models will be developed by applying thermodynamic theory and these established axioms as first principles. The following sections will develop these models from a qualitative and graphical viewpoint: the underlying mathematical derivations are provided separately in Appendices A and D.

3.1 Elements of a desirable, theoretically-grounded model

A residential energy use model making effective use of whole-house smart meter data should be able to parsimoniously contextualize the relevant physical and behavioural characteristics of a household. From a physical (thermal) perspective, this would include acknowledging the following general determinants of energy use:

- relationships between the size of heating and/or cooling loads (as appropriate to local climate) and both internal and external temperatures, reflecting the heat flow dynamics observable at the temporal resolution of the observation data
- thresholds past which heating and cooling loads activate and possible changes in those thresholds throughout the day
- endogenous and exogenous factors influencing interior temperatures

Behavioural factors are equally important in establishing a household's overall energy context. Specifically, a useful characterization would seek to determine the effects of:

- cyclical behavioural patterns and trends across daily, weekly, and annual time scales
- non-cyclical behavioural patterns depending on recent consumption values
- interactions between occupant activities and heating or cooling requirements

Finally, from a practical implementation and application standpoint, a strong statistical model should provide:

- tractable, interpretable, and information-rich model parameters that can be readily associated with specific household characteristics of interest so as to inform policy decisions and enable ad-hoc analysis independent of the model development process
- an efficient and stable mechanism for parameter fitting in order to support the rapid and reliable automatic characterization of large numbers of households

3.2 Empirical data as informing factors in model development

While it is important that domain knowledge and theoretical considerations inform any model development process, it can also be useful to draw insights directly from data representative of the systems being described. In the context of the residential energy modelling process described above, knowledge of independent heating, cooling, and behavioural loads in a household can be used to validate theoretical assumptions or determine unexpected trends that should be accounted for.

The Energy Hub Management System (EHMS) project is a data collection initiative that provides precisely such information. From 2011 through 2015, 25 households in Milton, Ontario provided circuit-level electricity use information to the project at five-minute intervals through intrusive load monitoring hardware. The resulting data set provides valuable disaggregated consumption information for electrical fans on natural gas furnaces, air conditioning units, and several other household appliances and plugs that can be applied in a wide variety of contexts: the research of both Bozchalui et al. (2012) and Koksal et al. (2015) referenced in the previous chapter made use of this dataset, for example. In the following sections, such disaggregated information from one participating household will be taken as a sample of residential heating, cooling, and behavioural (i.e. an aggregation of all non-HVAC) electrical loads.

3.3 Thermodynamic physical modelling

3.3.1 Empirically-informed considerations

A cursory examination of heating and cooling load time series (Figure 3.1) confirms a basic intuition concerning seasonal dependencies: the furnace tends to be most frequently active in winter months, while the air conditioner only runs in the peak of the summer (the furnace fan also runs in the summer, raising concerns about the validity of that time series as an indicator of heating load: this will be discussed further in later sections). Unfortunately, this simple explanation fails to account for significant intra-seasonal variations in electricity use. Is there a more sophisticated causal driver of thermal loads?

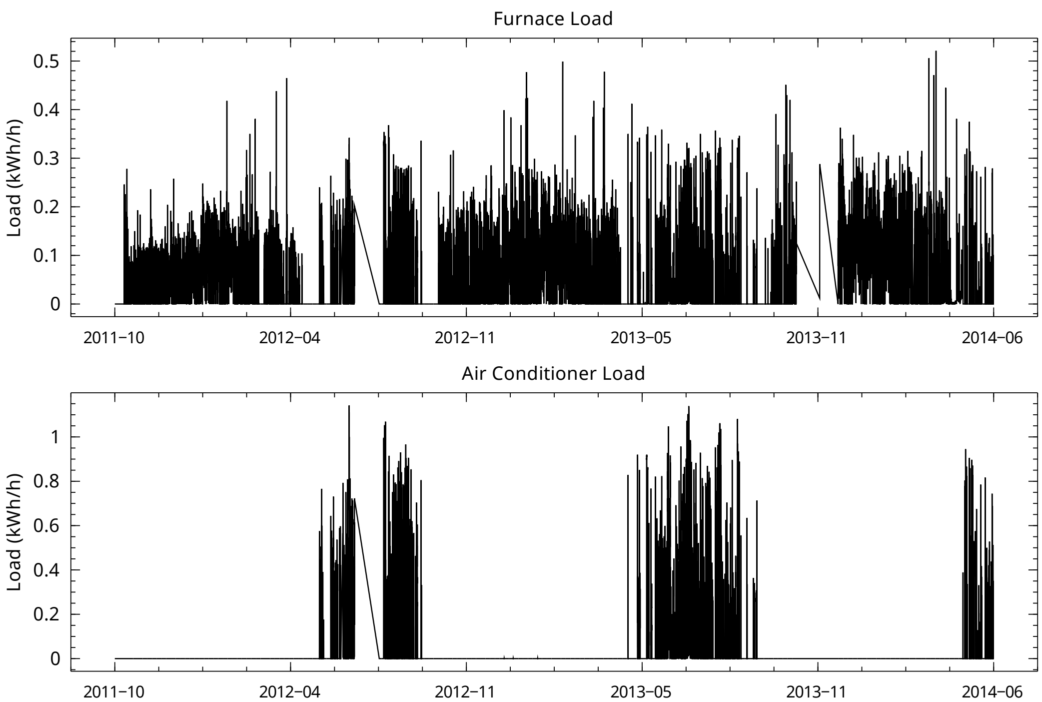

Figure 3.1: Heating and cooling electricity loads in a residential household with a forced-air natural gas furnace (with electric fan) and air conditioner.

The preceding chapter would suggest there is: every thermal model reviewed assumed some form of linear causality between external temperature and electrical heating or cooling loads. While such a claim would seem intuitively sound, it is nevertheless prudent to validate the assumption against available empirical data. A simple scatter plot visualization (Figure 3.2) suggests threshold-linear relationships - while furnace fan load during hot hours of summer months suggests that the circuit provides an imperfect measure of energy used for heating purposes, as discussed previously, air conditioner loads and furnace fan loads at cooler temperatures appear to confirm the linear relationship between heating energy use and exterior temperature.

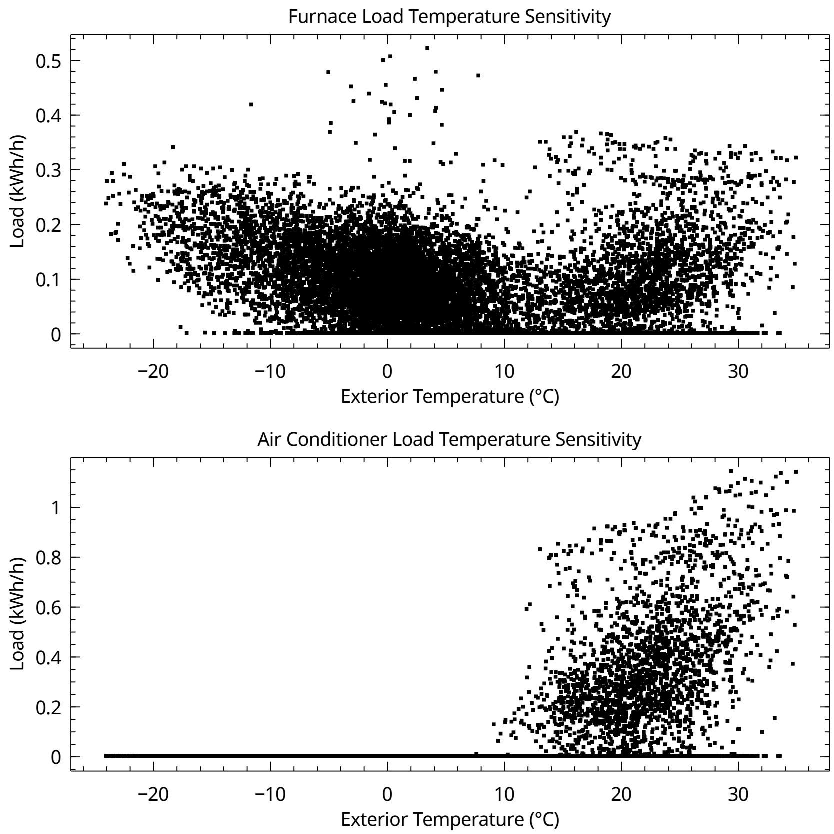

Figure 3.2: Heating and cooling electricity loads versus external temperature

3.3.2 Foundational assumptions

While encouraging, such correlationary analysis should not be taken as evidence of a causal link. Justifying such a link requires moving beyond a statistical characterization to a consideration of the underlying thermo-behavioural system. In particular, it would seem highly reasonable to adopt two core axioms:

1. Heat transfers (and thus energy flows) in and out of a dwelling conform to the physical laws of thermodynamics

2. At any given time, building occupants desire interior temperatures to remain constant or within some limited range, and apply heating and cooling systems to achieve this end

The first axiom hardly needs stating, while the second – although significantly less universal – is nonetheless highly defensible by referencing established modern behavioural norms and practises (ASHRAE (2013) define specific temperature ranges suitable for human occupancy, for example). It is worth noting that the actual value of the second axiom's desired interior temperature range is not prescribed, nor is it constrained to remain fixed throughout time - it may vary according to occupant behaviour and preferences.

3.3.3 Thermal models

The adoption of these axioms and their logical corollaries provides the basis for developing causal models of energy use in household heating and cooling. Two such models are derived from physical first principles in the Appendices, each adopting certain additional assumptions.

The first (Section A.1.2) assumes that heating and cooling systems maintain a strictly constant interior temperature, and can be shown to be a special two-regime (single-threshold) case of the threshold linear models common in the existing literature. In this model, a household heating system is assumed to be active at temperatures below some threshold point, with electrical load increasing linearly past that point, while cooling systems are assumed to be active at values above that same point and follow the same linear load increase pattern. The threshold value itself is fit to the observed data. Such a model satisfies most of the first two desirable physical model properties outlined in Section 3.1, and can be fit efficiently using modified linear regression techniques.

The second model developed (Section D.2.1) relaxes interior condition tolerances to allow temperatures to vary within some bounded range. Heating and cooling systems become active when internal temperature passes below or above the two defined temperature bounds, with internal temperature inferred according to previously-estimated heating or cooling system activity and observed outside temperatures. As in the previous case, the temperature threshold values are not predetermined but instead fit to observed data. This more sophisticated approach satisfies most elements of all three of the desirable physical model properties outlined in Section 3.1, but requires a state space modelling framework (see Section D.1) and the application of general nonlinear optimization techniques, significantly increasing computational requirements for parameter fitting and presenting a much higher risk of convergence issues (compromising the desirable parameter fitting properties also outlined in Section 3.1).

An alternate means to relax the constant-interior-temperature assumption is to adapt the first model described here to three heating regimes by arbitrarily imposing the existence of upper and lower temperature thresholds (much like the second model) in spite of a lack of theoretical physical justification. This relaxation results in a model that is essentially identical to that of Birt et al. (2012) as outlined in the previous chapter, which again satisfies most of the first two desirable physical model properties of Section 3.1 and can be fit by modified linear regression techniques.

3.3.4 Implicit secondary assumptions

Both theoretically-grounded models developed here make additional assumptions in modelling the automated control mechanisms governing heating and cooling systems (the third model discussed has already deviated from a theoretical grounding, so there is little point in elaborating its formal assumptions). In particular, the temporal resolution of input data (temperature and observed smart meter load) is taken to be sufficiently coarse-grained that a thermal system may only be active for some fraction of the time step, and it is assumed the relation between this duty cycle (fraction of time active) and heating or cooling output are linear. For example, in moderately warm weather, the air conditioner may be required to be active for 25% of the duration of the time step and provide an average of 250 watts of cooling in that time, while in hot weather, it may be active for 85% or 90% of the duration in order to provide an average of 850 or 900 watts over the course of the time step. Implicit in this assumption is the additional premise that heating and cooling systems never reach a saturation point where their required heat output would exceed that which is possible while operating at a 100% duty cycle.

This approach is motivated by the assumption of hourly observation data and is consistent with Albert and Rajagopal (2015), but contrasts with the assumptions made by Huang et al. (2013) who worked with 15-minute interval data and took observation timesteps to be shorter than HVAC system activity cycles. In that alternate formulation, heating or cooling loads are either consistently on (e.g. 1000 watt average output) or consistently off (0 watt average output) over the course of a single time period. While this alternate approach better represents the actual control dynamics of a typical heating or cooling system and would allow for an estimate of furnace or air conditioner sizing given sufficiently high-resolution data, in the context of the data available it is unreasonable to assume that an HVAC system would operate for a full hour at a time before reassessing its effects on interior temperature, thus necessitating the linear duty cycle approximation used here.

In the case of the bounded interior temperature model, it is also assumed (based on hour interval readings) that the temporal resolution of input data is sufficiently fine-grained to allow physical characteristics such as the dwelling's thermal mass to influence fluctuations in internal temperature, and that such fluctuations would be of sufficient magnitude to have an effect on energy consumption during the time period in question.

3.4 Statistical behavioural modelling

Unlike the case of exterior temperature in heating and cooling loads, there are no obvious and readily-available exogenous indicators of a household's behavioural (non-thermal) load at a given time. Instead, households tend to have individual endogenous cyclical patterns or short-term behavioural trends that characterize their non-thermal energy use habits.

Such trends are much less well-suited to causal, deterministic explanations and instead require a more rigorous statistical focus to capture. While behavioural load data can be visualized as a time series (Figure 3.3), this perspective may not readily illustrate underlying trends of interest. Instead, alternate visualizations of key statistical metrics can serve to identify characteristics explaining observed variance in known behavioural time series data. Three such metrics commonly employed in time series analyses (Hipel and McLeod, 1994) are:

1. Distributions of load values indicating the range of observed values and the relative

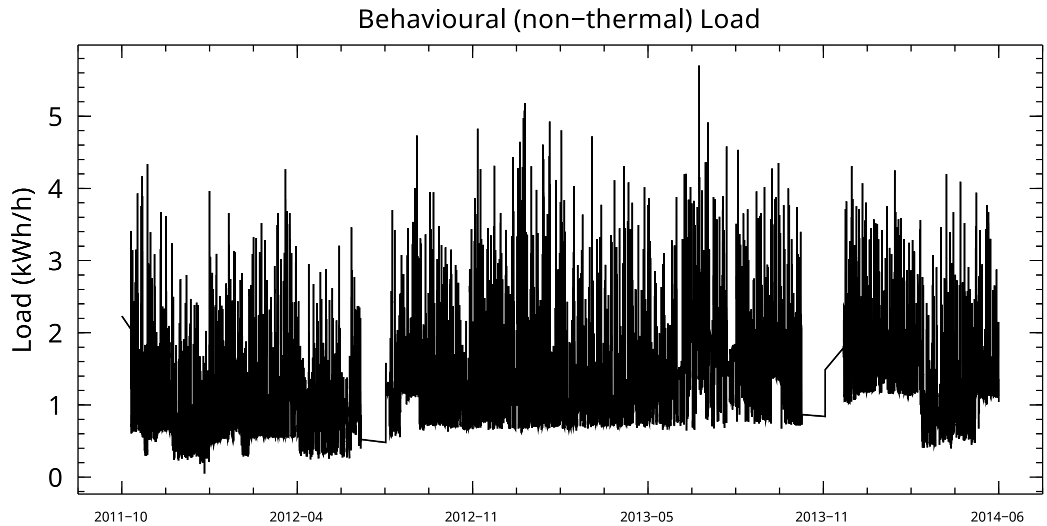

Figure 3.3: Behavioural (non-heating or cooling) loads in a residential household

frequency with which those values occur, as illustrated by histograms

2. Spectral frequency content indicating the occurrence and relative strength of various periodic cycles in behavioural data, visualised via periodograms with the x-axis adjusted to give output in the topically-relevant time domain (as opposed to the traditional frequency domain)
3. Autocorrelations describing the average correlation of an observation with the various observations preceding it, represented graphically with correlograms

The following sections will combine domain knowledge of typical behavioural load drivers with visualizations of these three statistical time series metrics to iteratively develop a model of behavioural household loads.

3.4.1 Model development

Figure 3.4 provides visualizations of the statistical metrics outlined above as applied to the sample behavioural data, adjusted to set the average value to zero. This adjustment can be interpreted as a trivial baseline model, where behavioural load is simply estimated as the average behavioural load across the entire observation period.

Each metric provides new insights into the statistical characteristics of the data: the histogram, subplot (a), indicates a log-normal distribution of values, the periodogram, subplot (b), shows several significant cyclical (periodic) trends, and the correlogram and partial correlogram, subplots (c) and (d), suggest relationships between sequentially-observed values. Applications of these various insights are discussed in the following sections.

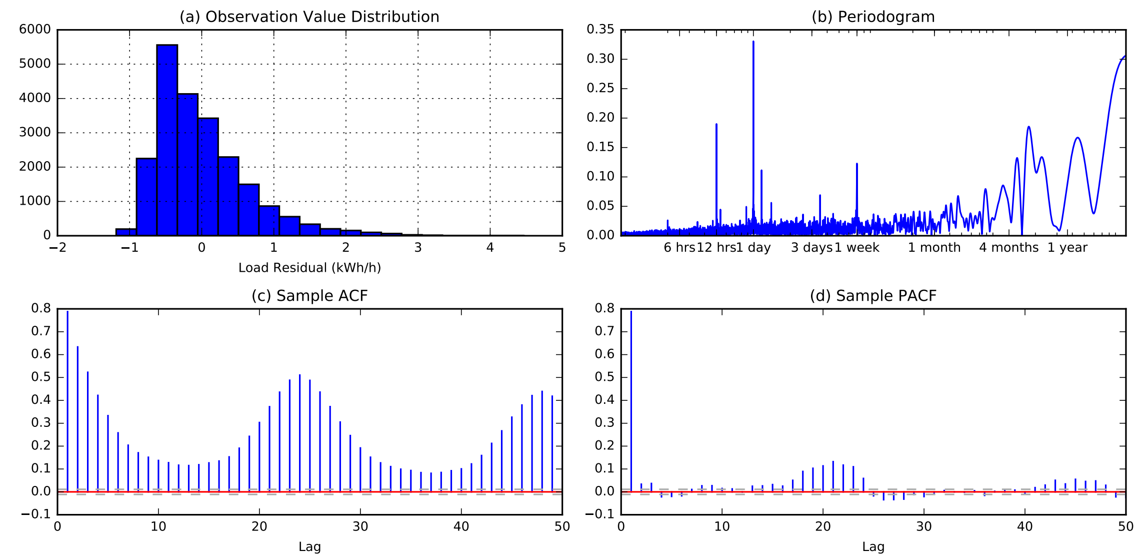

Figure 3.4: Statistical summary of sample behavioural electricity data residuals after mean adjustment

Distribution of values: log-normal behaviour

Given that electrical load values are constrained to always be positive, there is reason to believe that variations from a mean value may not be symmetric (for example, an average household load may be 1 kW, but 3 kW may be used occasionally to run a clothes dryer - clearly their is no counter-situation in which -2 kW may be occasionally used). Visualizing the distribution of sample behavioural data (Figure 3.4a) confirms this suspicion. One remedy to this situation would be to let error terms be drawn from an asymmetric probability distribution (for example, some form of gamma distribution). Alternatively, the behavioural load itself can be transformed: in this case, a log transform roughly centres the observations around a mean while also eliminating the possibility of negative values. (In the case where error terms are not just symmetric, but normally distributed, this becomes a log-normal distribution).

While the log transform improves the distribution of the data, the model overall remains based on poor implicit assumptions – behavioural energy use in one hour is clearly related to use in previous hours, and does not maintain some constant mean (as modelled here) throughout the day, week, or year. These assumptions will be revisited in the next sections.

Frequency content: periodic trends and low-frequency noise

The log-transformed constant-load behavioural model centres overall average load in a symmetric distribution, but ignores any repeating patterns or periodicity in behavioural loads. Examining the periodogram of the data shows multiple instances of well-defined higher-frequency periodicity as well as less-distinct low-frequency noise (noted here but

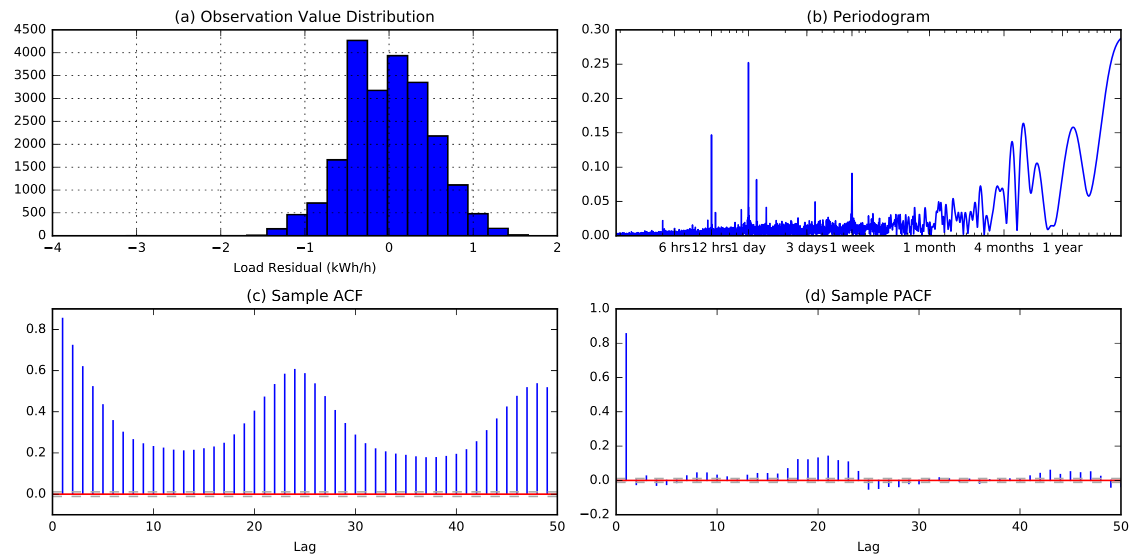

Figure 3.5: Statistical summary of sample behavioural electricity data residuals after log transformation and mean adjustment

addressed in the next section).

Domain knowledge of household behavioural loads suggests that occupant behaviour follows repeating daily and weekly consumption patterns. The periodogram strongly confirms this intuition, with distinct spectral density spikes visible for frequency components corresponding to daily and weekly periodic behaviour. There are multiple ways to model these patterns, including introducing seasonal autoregressive (SAR) terms (i.e. lag-24 and/or lag-168), or allowing such SAR terms to also vary according to the current hour or day of week (a periodic autoregressive model, or PAR).

A simpler approach that provides additional flexibility in adapting to aperiodicities in behavioural patterns (resulting from daylight savings time shifts, for example) is to introduce seasonal (cyclical) mean adjustment according to hour-of-day and/or day-of-week (as observed by a wall clock). This approach has the added benefit of providing easily interpretable model parameters (representing average behavioural consumption at a given time of day or week). Such a model follows directly from the basic mean-adjusted case, where now the average value simply takes on different values according to the "season" (i.e. time of day or week).

The exact definition of a season in relation to hour-of-day and day-of-week is still to be determined here. An ideal definition would reduce variance and fully remove seasonality from observation residuals while minimizing the number of model parameters required to do so. The highest degree-of-freedom case would be to simply define each hour-of-week as its own season, resulting in $24 \times 7 = 168$ unique mean adjustment terms. A more parsimonious option commonly seen in the literature would be to differentiate only between weekday and weekend days-of-week, resulting in $24 \times 2 = 48$ parameters. Additional parsimony could be achieved by adjusting for hour-of-day and day-of-week independently, leading to $24 + 7 = 31$

or $24+2=26$ parameters according to whether or not the weekday/weekend simplification described above is applied.

The properties of the model residuals for hour of weekday vs weekend deseasonalization are given in Figure 3.6. Mean deseasonalization effectively accounts for the weekly periodicities of the data (incorporating one of the desired elements of a behavioural model described in Section 3.1) while reducing the daily frequency content and overall variance of the residuals. In the sample behavioural data the 48-parameter deseasonalization approach was more effective than the 26- or 31 parameter alternatives at reducing cyclical trends in the model residuals, while giving similar results to the 168-parameter alternative, and so is preferred here as an expressive yet parsimonious option.

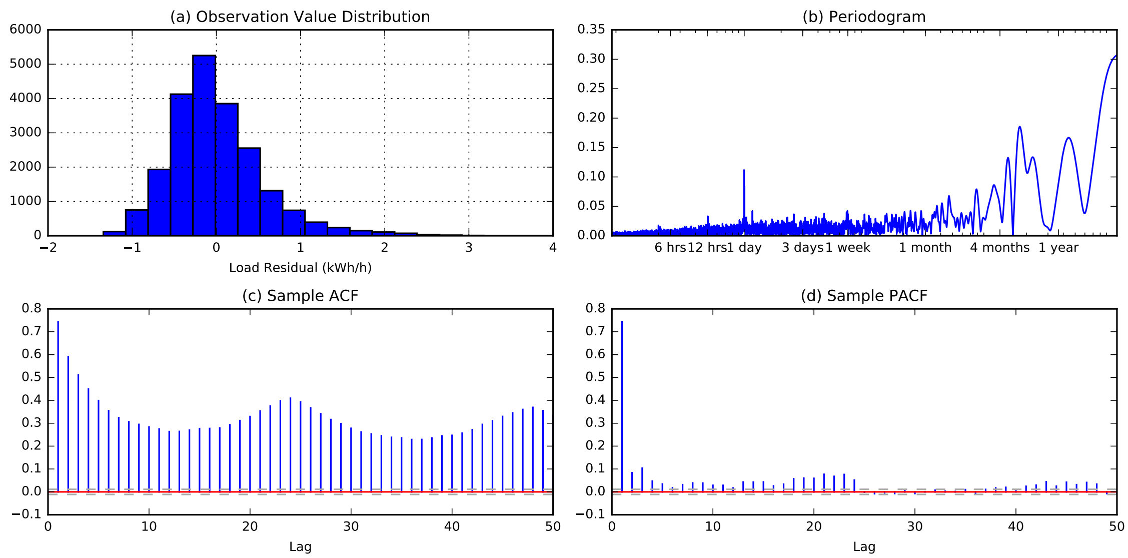

Figure 3.6: Statistical summary of sample behavioural electricity data residuals after hour-of-weekday/weekend mean adjustment

Autocorrelation: random walk behaviour

The periodograms visualized in the previous sections reveal significant low-frequency noise (deviations from the mean that play out over long time-scales) not attributable to behavioural seasonality. These disturbances can be better visualized in the time domain by applying a medium-term (such as one month) rolling average filter to original sample data, as shown in Figure 3.7 and revealing low-frequency, non-stationary fluctuations in behavioural load levels.

These findings suggest that modelling behavioural loads as independent deviations from stable means is generally insufficient to describe all behavioural trends, and highlight the danger of ignoring the sequential nature of such loads, as has been done up until this point of the analysis. Indeed, these results are consistent with an intuitive understanding that building occupant behaviour and resulting energy consumption depends heavily on a

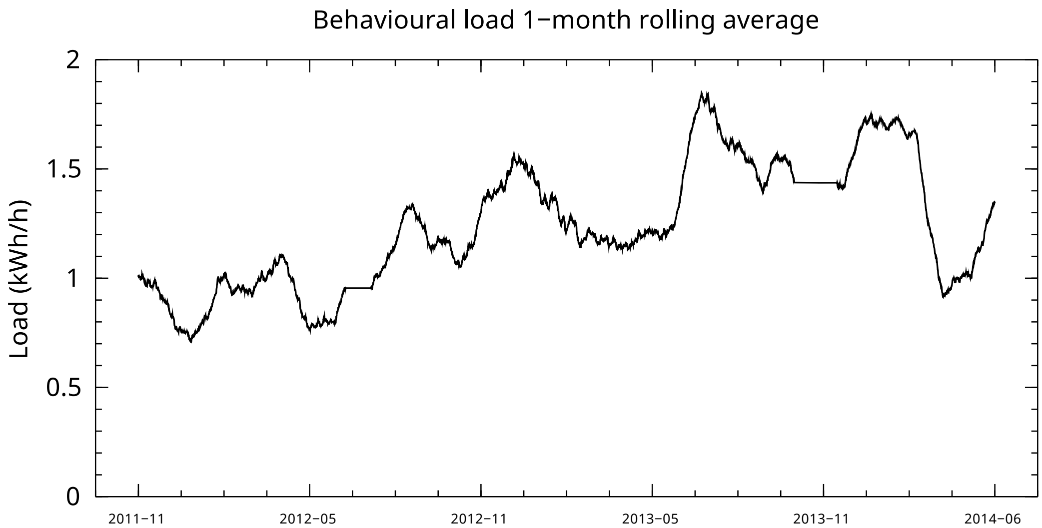

Figure 3.7: 1-month rolling average plot of sample behavioural electricity data

broader temporal context as well as past behaviour in recent hours. Autocorrelation and partial-autocorrelation functions can be used to examine this dependency more formally: doing so with sample behavioural data reveals slowly-decaying autocorrelations (Figure 3.4c) and a strong lag-1 partial-autocorrelation (Figure 3.4d). As elaborated in Hipel and McLeod (1994), this behaviour is evidence indicating that behavioural loads tend to depend strongly on their immediately preceding value.

The nature of the autocorrelations of the data suggests that an autoregressive (AR) time series model would be useful to model behavioural dynamics. However, AR models require that the process under consideration be stationary around some constant mean, and 3.7 suggests that this is not the case here. The traditional ARIMA approach to accommodating such non-stationarity is to apply differencing to the data. Indeed, differencing the sample behavioural data is highly effective at eliminating this noise: however, it also discards potentially useful information about a household's long-term consumption trends. Instead, it would be preferable to model this non-stationarity explicitly.

An intuitive approach to modelling the behavioural trends described above is to treat behavioural energy use as some form of random walk, with consumption in each hour deviating from the previous randomly, with the added possibility of ephemeral deviations that do not contribute to shaping long-term trends. Adding this random walk component to the previously developed seasonal-average model yields a structural time series model with seasonal and random walk components, incorporating the non-cyclical effects of recent behavioural values as suggested in Section 3.1.

Fitting sample data to a random walk model and checking the residual periodogram (Figure 3.8) confirms that low-frequency behavioural trends are fully captured by the random walk, eliminating low-frequency components from the series' spectral density while also significantly reducing residual variance and hence model fit error.

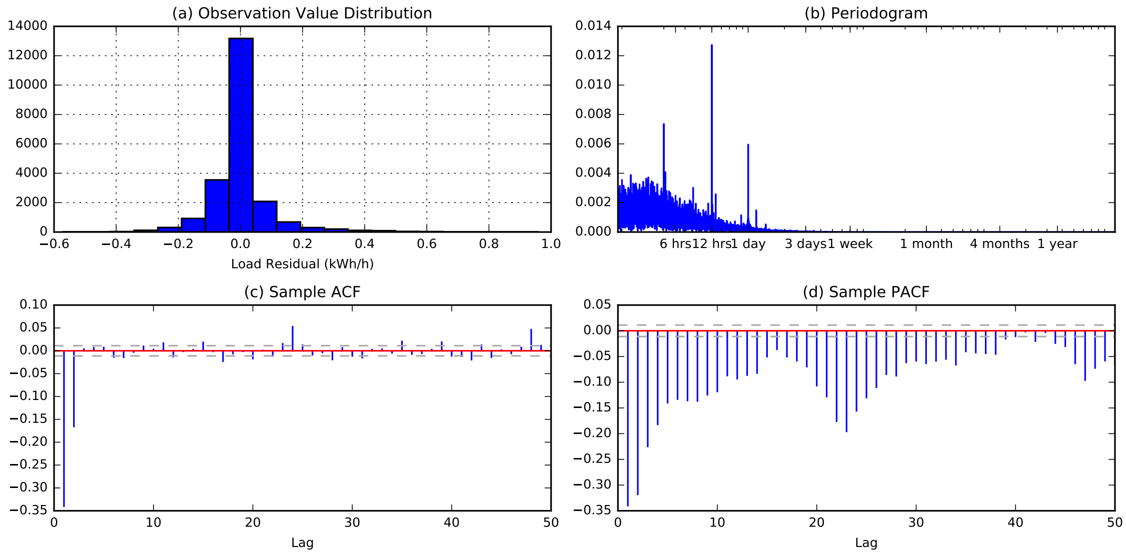

Figure 3.8: Statistical summary of sample behavioural electricity data residuals after random walk adjustment

Combining the elements

The model components developed independently in the preceding sections can of course be applied together as well. Figure 3.9 demonstrates that this combination provides a model that fits the sample behavioural data significantly better than any of the above models taken alone. Residual variance and spectral frequency content are near-zero while the autocorrelation structure of the data has been eliminated with the exception of autocorrelation and partial autocorrelations at daily lag intervals (24, 48, etc).

3.4.2 Model linearization

While the previously-developed model more closely represents established domain knowledge and observed statistical trends, its random walk component depends on computationally expensive nonlinear optimization and requires recursive regression techniques for estimating hourly-mean deseasonalization parameters. This additional complexity can reduce model fitting performance in the presence of large amounts of missing data - in such situations, a simpler approach may be desirable and more in line with the computational goals laid out previously in Section 3.1. Substituting the model's local-level component with a simple autoregression on the previous hour's value provides an (admittedly imperfect) approximation of the effects of correlated consumption data while reducing model fitting requirements to linear regression via ordinary-least-squares.

A second challenge arises in fitting log-transformed behavioural data simultaneously with untransformed thermal load data. While this is possible via a nonlinear state-space modelling approach and unscented Kalman filtering (Haykin, 2004), these techniques also sig-

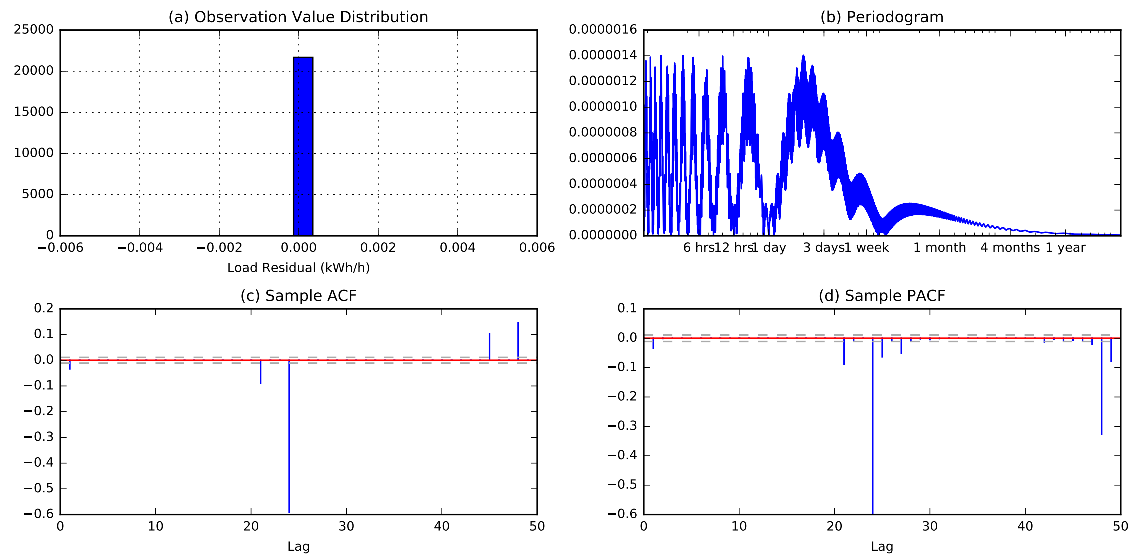

Figure 3.9: Statistical summary of sample behavioural electricity data after log transformation with hour-of-weekday/weekend mean and random walk adjustment

nificantly increase computational requirements and introduce nontrivial stability concerns into both the parameter fitting and state estimation processes. Reverting to fitting untransformed behavioural data simplifies the joint thermo-behavioural estimation process described in the following section, again advancing the computational goals outlined in Section 3.1, without sacrificing any of the behavioural considerations noted in that section and achieved by the random walk approach.

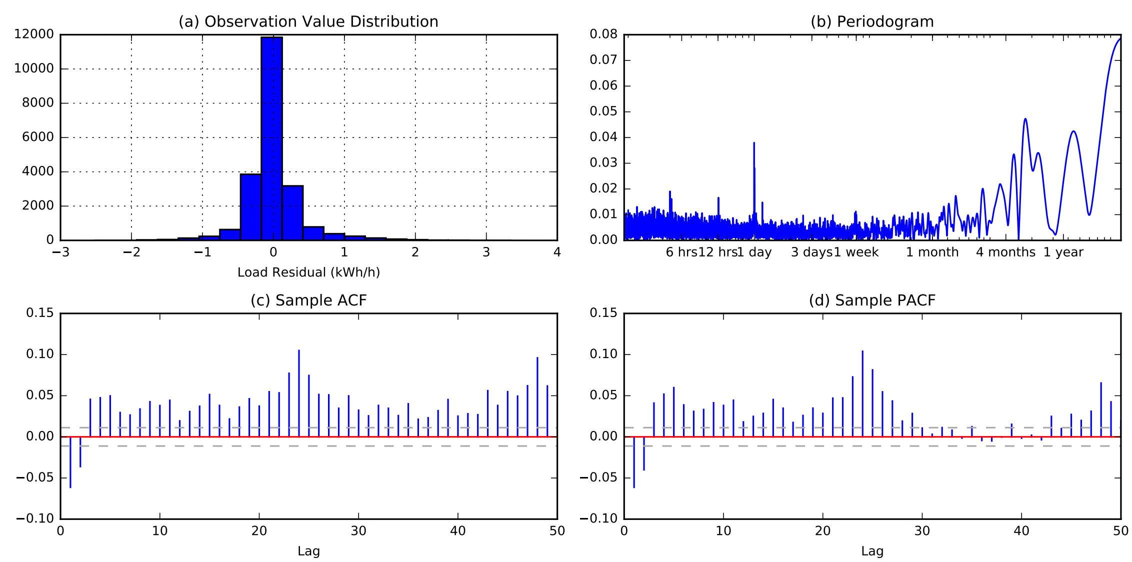

Figure 3.10: Statistical summary of sample behavioural electricity data after hour-of-weekday/weekend mean and hour-lag autoregression adjustment

Figure 3.10 visualizes the results of applying this simplified model to the sample data, and a simple comparison of goodness-of-fit between various possible combinations of the models considered here, as determined by the root-mean-squared values of the fit residuals, are provided in Table 3.1. As is to be expected, the ability to fit the data with the fully linearized model is significantly diminished relative to the combined model results shown in Figure 3.9. However, this linearized model does still outperform all of the non-random-walk models considered here.

Log	**DS**	**RW**	**AR**	**RMSE**
				0.643 66
•				0.496 87
	•			0.567 33
•	•			0.439 35
		•		0.097 09
•		•		0.022 60
	•	•		0.004 18
•	•	•		0.000 05
•	•		•	0.244 95
	•		•	0.378 56

Table 3.1: Model fit root mean square errors for combinations of log-tranformation (Log), hour-of-weekday/weekend mean adjustment/deseasonalization (DS), random walk adjustment (RW), and 1-hour lag autoregression (AR)

3.5 Combining physical and behavioural models

While electrical loads associated with heating or cooling and those related to occupant behaviour are clearly interconnected (a house left vacant may have different thermostat settings, for example), a reductionist approach that considers each component individually can serve as a useful starting point for eventually building a more complete residential energy model. Given the additive and independent nature of the thermal and behavioural models described above, creating a composite model is straightforward. In general, estimated heating and cooling load values can simply be added to behavioural load values to obtain whole-house aggregate load (smart meter) estimates. A more sophisticated approach allowing for influences between thermal and behavioural loads (incorporating the final behavioural element suggested in Section 3.1) is proposed in Appendix D.3, but is not considered further here for reasons of computational complexity.

In this chapter, insights from a review of the literature have been applied to develop a series of suggested or desirable elements of physical and behavioural models, as well as the process for fitting those models. Observed heating, cooling, and behavioural data have been applied in concert with domain knowledge concerning residential electricity use to develop several different separate thermal and behavioural modelling approaches.

When both the thermal and behavioural models to be combined can be represented as linear regression problems, the combined model can be fit reliably and efficiently with standard regression techniques as well. The next chapter will use such techniques to apply the linear models developed here to characterizing a set of validation households, and compare model performance to other approaches from the literature.

Chapter 4

Model Characterization Performance Comparison

In Section 3.2, disaggregated electrical end-use history from a single household in the Energy Hub Management System (EHMS) dataset was used to validate and motivate the development of novel household energy consumption models. In this chapter, homeowner-reported physical dwelling characteristics and the remainder of the disaggregated dataset are applied to assess the ability of a subset of those models to infer household energy-use characteristics given aggregate, whole-house electrical meter readings, and compare performance to existing techniques identified in Section 2.3.

An overview of the evaluation data is provided, followed by an outline of each of the models to be assessed. Finally, a comparison of the characterization performance of the various models according to multiple metrics is presented.

4.1 Energy Hub Management System data

The EHMS dataset consists of 25 homes, of which one was arbitrarily selected for use as a sample household in the disaggregated model development process (Chapter 3), and so was not included in the model disaggregation assessment process. Of the remaining 24 homes, three more were rejected as inappropriate for disaggregation assessment purposes as they did not include both labelled furnace fan (heating) and air conditioner (cooling) circuits.

4.1.1 Data preparation

Raw, circuit-level EHMS data were reported in five-minute intervals. From this starting point, any observation period was dropped when all circuits reported zero load, the sum of circuit loads differed from a separately-measured whole-house load reading by more than 80%, or any single circuit reported a load greater than the whole-house aggregate reading.

The remaining five-minute consumption totals within a given hour were aggregated and downsampled to hourly consumption values in order to better simulate typical smart meter readings. All circuit loads for a given hour that did not correspond to an air conditioner or furnace fan were then summed to generate a non-thermal (behavioural) load category. Finally, all of the thermal and non-thermal loads (air conditioner, furnace fan, and behavioural) were summed for each hour to generate household totals. To maintain consistency between the disaggregated and aggregated readings, this derived hourly household total was used in the place of the separately-measured whole-house data readings.

The external temperature data used in the assessment were obtained from the University of Toronto's Mississauga Campus weather station (`https://www.utm.utoronto.ca/geography/resources/meteorological-station/weather-data`). 0.92% of readings were missing and so were replaced with contemporaneous Environment Canada readings taken at the Guelph Turfgrass Institute (`http://climate.weather.gc.ca/climateData/hourlydata_e.html?StationID=45407`) where possible. The remaining 0.02% of values that were missing from both datasets were imputed by linear interpolation from the closest preceding and subsequent observed values.

4.1.2 Aggregate consumption data

Summary statistics for whole-house aggregate electrical load are reported in Table 4.1. Of particular note is the commonly large fraction of hours for which no household observation data are available, in some cases more than 30%. While most household observations span multiple years (and hence multiple heating and cooling seasons), others such as households P and U cover much more limited time periods.

Table 4.2 provides a breakdown of the total heating, cooling, and behavioural consumption observed for each household: while the proportion of electricity used by each category varies widely across the sample, the values are roughly consistent with estimated Ontario averages – 24% of residential electricity use for heating and ventilation, and 7% for air conditioning according to the Ontario Power Authority (2014). The following sections study the nature of heating and cooling use in more detail.

4.1.3 Heating load data

Summary statistics for the household furnace fan load readings are provided in Table 4.3. As in the aggregate consumption case, non-negligible fractions of data are commonly missing. Relative to the total number of observed hours, for most households there are a significant number of hours in which the heating system was active (drawing non-zero load): in many cases, the system is nearly always active, suggesting that furnace fan load may serve as a very poor proxy of heating system energy use.

Figures 3.2 and 3.1 demonstrated that it was common for the sample household's furnace fan to be running simultaneously with its air conditioner. In those cases, it would seem reasonable to assume that the heating system was not in fact active, but that the fan was simply distributing cooled air through the central heating system. Under that assumption,

ID	Start	End	Total Hours	Missing Hours	% Missing	Hourly Mean (kWh)	Hourly StDev (kWh)	Daily Mean (kWh)	Daily StDev (kWh)
A	2011-10-18	2014-05-19	22652	2777	12.3	0.8	0.6	19.8	8.4
B	2011-06-14	2014-07-08	26897	1857	6.9	1.5	0.7	34.9	8.1
C	2011-12-09	2014-07-08	22625	3531	15.6	1.2	0.7	29.5	8.9
D	2011-12-29	2014-03-23	19581	870	4.4	1.1	0.6	24.9	9.0
E	2011-12-09	2014-05-22	21499	386	1.8	0.9	0.6	21.5	7.4
F	2011-12-17	2014-07-08	22433	2478	11.0	0.8	0.6	19.5	7.6
G	2011-12-29	2014-07-08	22145	1383	6.2	0.6	0.5	13.3	5.3
H	2011-12-17	2013-12-29	17842	654	3.7	2.5	0.8	59.0	11.1
I	2011-12-17	2013-08-20	14701	2813	19.1	0.6	0.5	14.0	6.5
J	2011-12-09	2013-10-01	15904	314	2.0	1.3	0.7	31.0	12.8
K	2012-01-07	2013-04-18	11227	2291	20.4	1.1	0.7	26.1	9.4
L	2012-01-07	2014-02-19	18586	6498	35.0	0.7	0.5	16.1	5.7
M	2012-03-20	2014-07-08	20174	2023	10.0	0.3	0.3	8.2	3.5
N	2012-03-20	2014-01-29	16319	4149	25.4	0.9	1.0	21.5	14.3
O	2012-01-07	2013-12-28	17319	3182	18.4	0.8	0.6	18.5	7.5
P	2012-03-20	2013-01-09	7089	338	4.8	0.7	0.8	16.9	10.2
Q	2012-03-22	2014-07-08	20128	979	4.9	0.8	0.9	19.7	10.3
R	2012-04-18	2014-04-20	17580	5468	31.1	0.6	0.5	13.4	6.2
S	2012-03-22	2014-07-08	20129	5987	29.7	0.9	1.0	19.5	11.3
T	2012-04-17	2014-05-30	18549	4389	23.7	0.4	0.3	10.4	3.3
U	2012-03-22	2012-11-09	5577	277	5.0	0.7	0.5	16.3	4.9

Table 4.1: Total aggregate consumption profile summary of each validation household. Missing hours correspond to both time periods in which no data was reported as well as where data was reported but removed in the pre-processing steps of Section 4.1.1.

when the air conditioner is known to be active the furnace fan's consumption should in fact be attributed as a cooling load, not a heating load. Modifying the ground truth observations to account for this fact essentially eliminates supposed warm-weather heating system activity from the sample household observations, as seen in Figure 4.1.

Re-evaluating the heating system data after this adjustment indicates a more probable representation of hours in which the heating system was active (i.e. not 100%). Table 4.4 reflects this adjusted summary.

4.1.4 Cooling load data

Summary statistics for the household air conditioner load (after the adjustment described in the previous section) are provided in Table 4.5. As noted previously, a significant number of hourly observations are missing. Unlike the furnace fan circuit readings, home air conditioner circuits tend not to be constantly active, and those that are (such as households R, T, and U) have much smaller average active loads compared to homes with much rarer circuit activity, suggesting that while there may be a load on such circuits throughout the year, it is generally not significant.

ID	Total Heating (kWh)	Total Cooling (kWh)	Total Behavioural (kWh)	% Heating	% Cooling	% Behavioural
A	1 587.7	490.0	14 516.6	9.6	3.0	87.5
B	7 963.4	3 088.2	26 042.9	21.5	8.3	70.2
C	3 285.4	1 623.7	18 868.6	13.8	6.8	79.4
D	6 897.5	2 026.2	10 749.7	35.1	10.3	54.6
E	4 011.3	562.0	14 440.6	21.1	3.0	75.9
F	4 104.4	1 320.3	11 444.2	24.3	7.8	67.8
G	4 114.5	1 091.4	6 801.0	34.3	9.1	56.6
H	8 068.9	2 027.5	32 444.6	19.0	4.8	76.3
I	1 562.1	1 222.3	4 688.4	20.9	16.4	62.7
J	2 155.2	453.1	17 643.1	10.6	2.2	87.1
K	1 896.9	695.2	7 199.1	19.4	7.1	73.5
L	836.1	237.6	7 223.1	10.1	2.9	87.1
M	974.3	150.7	5 176.4	15.5	2.4	82.1
N	2 750.5	3 017.9	5 393.5	24.6	27.0	48.3
O	1 448.2	781.8	8 748.9	13.2	7.1	79.7
P	1 879.1	1 027.5	2 095.3	37.6	20.5	41.9
Q	2 706.2	3 389.3	9 992.7	16.8	21.1	62.1
R	1 841.9	894.5	4 423.3	25.7	12.5	61.8
S	2 999.2	2 927.7	6 249.0	24.6	24.0	51.3
T	1 477.9	14.0	4 824.1	23.4	0.2	76.4
U	2 156.6	449.1	1 192.8	56.8	11.8	31.4

Table 4.2: Total and proportion of heating, cooling, and behavioural electricity consumption over the entire observation period for each household

4.2 Models to be evaluated

To ensure parameter stability during the model fitting process, evaluation was limited to only the regression-compatible model components outlined in Chapter 3, specifically:

- the constant-interior-temperature model derived in Section A.1.2, consisting of two thermal regimes (active heating and active cooling) with a temperature threshold point fit according to the data, hereafter referred to as **Fit2**
- The three-regime (active heating, passive, active cooling) model outlined in Birt et al. (2012) or alternatively obtained by relaxing the physical constraints of the Fit2 model (as discussed in Section 3.3.3), hereafter referred to as **Fit3**.
- the linearized behavioural model outlined in Section 3.4.2, consisting of deseasonalizing loads with weekday or weekend hour-of-day consumption averages in combination with an autoregression on total consumption in the previous hour, hereafter labelled as **DAR**.

The joint thermo-behavioural models resulting from adding these thermal and behavioural

ID	Total Hours	Missing Hours	% Missing	Active Hours	% Active	Active Hourly Mean (kWh)	Active Hourly StDev (kWh)
A	22652	2777	12.3	9813	49.4	0.2	0.1
B	26897	1857	6.9	24977	99.7	0.3	0.1
C	22625	3531	15.6	10055	52.7	0.3	0.2
D	19581	870	4.4	17125	91.5	0.4	0.3
E	21499	386	1.8	18607	88.1	0.2	0.2
F	22433	2478	11.0	13078	65.5	0.3	0.1
G	22145	1383	6.2	13356	64.3	0.3	0.2
H	17842	654	3.7	16937	98.5	0.5	0.2
I	14701	2822	19.2	5429	45.7	0.3	0.2
J	15904	314	2.0	13758	88.2	0.2	0.1
K	11227	2291	20.4	8931	99.9	0.2	0.2
L	18586	6498	35.0	5743	47.5	0.1	0.1
M	20174	2023	10.0	16164	89.1	0.1	0.1
N	16319	4150	25.4	11916	97.9	0.2	0.3
O	17319	3182	18.4	5214	36.9	0.3	0.2
P	7089	338	4.8	6751	100.0	0.3	0.1
Q	20128	979	4.9	19149	100.0	0.1	0.2
R	17580	5470	31.1	12042	99.4	0.2	0.2
S	20129	5987	29.7	14137	100.0	0.2	0.2
T	18549	4389	23.7	14158	100.0	0.1	0.1
U	5577	278	5.0	5299	100.0	0.4	0.0

Table 4.3: Heating consumption profile summary for each validation household. % active hours are reported relative to all non-missing hours. Missing hours correspond to both time periods in which no data were reported as well as where data were reported but removed in the pre-processing steps of Section 4.1.1.

model components together were ultimately used in the assessment and are referred to here as **DAR-Fit2** and **DAR-Fit3**. Julia code implementing the model and fitting procedure for each are provided in Sections B.4 and B.5, respectively.

Several model components presented in the literature were also evaluated, namely:

- The three-regime model (Fit3) exactly as outlined by Birt et al. (2012), without any behavioural component
- The periodic autoregressive behavioural model (including outlier detection) outlined by Ardakanian et al. (2014), referred to here as **PARX**.
- The four-regime model with fixed temperature thresholds employed by both Espinoza et al. (2005) and Ardakanian et al. (2014), and referred to here as **Fixed4**

The **Fit3** (as used by Birt, and implemented as given in Section B.1) and joint **PARX-Fixed4** model (as used by Ardakanian et al., and implemented as given in Section B.2)

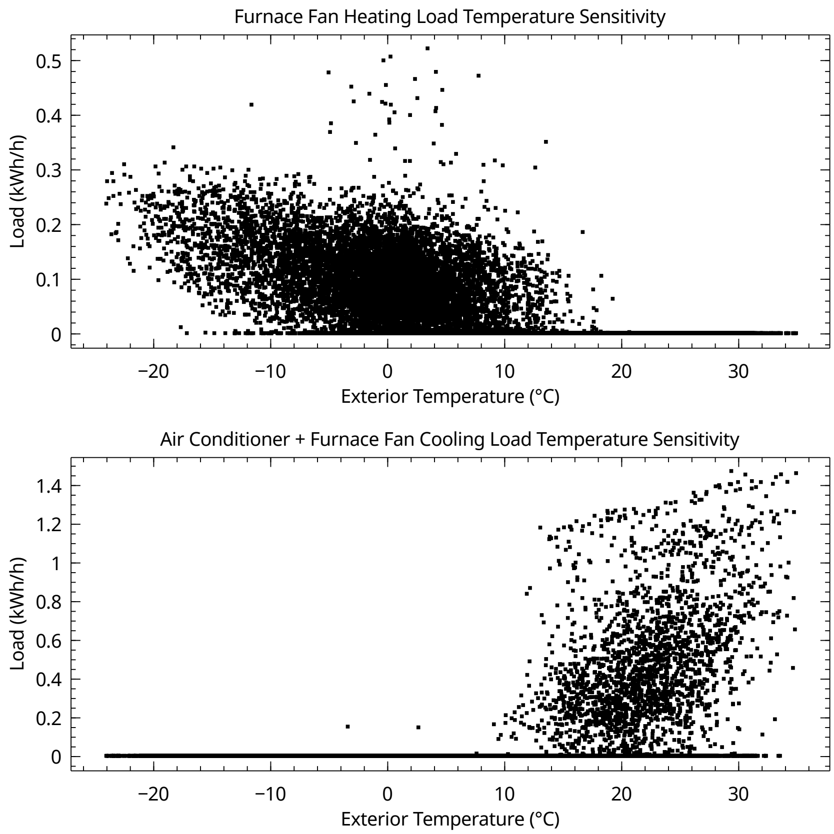

Figure 4.1: Temperature sensitivity of heating and cooling loads in the sample household, after reclassifying furnace fan loads while the home's air conditioner is active as cooling consumption

were evaluated. In addition, to better compare the relative contributions of the PARX and Fixed4 model components, a **DAR-Fixed4** model combining the novel behavioural DAR model with the established Fixed4 model was also tested (an implementation is provided in Section B.3).

As noted in Section 2.3.3, Markov-modelling approaches such as that of Albert and Rajagopal (2015) have no direct means of handling data gaps that cannot be filled artificially by imputation. Given the significant number and size of gaps in the validation dataset, a Markov-modelling approach was not assessed here.

4.3 Disaggregation performance assessment

The whole-house hourly electricity data for each of the 21 validation homes (obtained by summing the observed heating, cooling, and behavioural loads in each hour, and corre-

ID	Total Hours	Missing Hours	% Missing	Active Hours	% Active	Active Hourly Mean (kWh)	Active Hourly StDev (kWh)
A	22652	2777	12.3	8799	44.3	0.1	0.1
B	26897	1857	6.9	17937	71.6	0.3	0.1
C	22625	3531	15.6	7647	40.0	0.3	0.2
D	19581	870	4.4	14352	76.7	0.4	0.3
E	21499	386	1.8	17021	80.6	0.2	0.2
F	22433	2478	11.0	10274	51.5	0.3	0.1
G	22145	1383	6.2	11074	53.3	0.3	0.2
H	17842	654	3.7	13107	76.3	0.5	0.2
I	14701	2822	19.2	3798	32.0	0.2	0.1
J	15904	314	2.0	12393	79.5	0.1	0.1
K	11227	2291	20.4	7993	89.4	0.2	0.1
L	18586	6498	35.0	5317	44.0	0.1	0.1
M	20174	2023	10.0	15986	88.1	0.1	0.1
N	16319	4150	25.4	9905	81.4	0.2	0.2
O	17319	3182	18.4	2601	18.4	0.2	0.2
P	7089	338	4.8	6146	91.0	0.3	0.1
Q	20128	979	4.9	16861	88.1	0.1	0.1
R	17580	5470	31.1	1140	9.4	0.1	0.1
S	20129	5987	29.7	3230	22.8	0.2	0.2
T	18549	4389	23.7	207	1.5	0.3	0.2
U	5577	278	5.0	115	2.2	0.4	0.0

Table 4.4: Heating consumption profile summary for each validation household, after reclassifying furnace fan loads while the home's air conditioner is active as cooling consumption

sponding to smart meter readings that would be available to a utility or homeowner) were taken with matching hourly exterior temperature values and provided as inputs to each of the five models outlined above. The resulting heating, cooling, and behavioural load estimates produced by each model were then compared with the original observed load components (with the furnace fan adjustment described in Section 4.1.3).

Figures 4.2 to 4.7 visualize the disaggregation performance results across all households and models for a number of evaluation metrics (accuracy, performance, recall, and F1 statistic for thermal system activity detection, and root mean square error and mean absolute percent error for component load estimation), with full numeric results available in Appendix C. In each case, the range of values of each metric of interest are reported in standard box-and-whisker plots (Tukey, 1977), where the centre line denotes the metric's value for the median (50th percentile) household, the upper and lower limits of the box denote the values of the 25th and 75th percentile households, and the outer bars (whiskers) represent the values for the furthest outlying households still within 1.5 interquartile ranges (the distance between the 25th and 75th percentiles) of the box limits. Outlying household values beyond the whiskers are plotted directly.

ID	Total Hours	Missing Hours	% Missing	Active Hours	% Active	Active Hourly Mean (kWh)	Active Hourly StDev (kWh)
A	22652	2777	12.3	1014	5.1	0.8	0.5
B	26897	1858	6.9	7064	28.2	0.8	0.4
C	22625	3531	15.6	2410	12.6	1.2	0.6
D	19581	870	4.4	2774	14.8	1.2	0.7
E	21499	386	1.8	1586	7.5	0.6	0.3
F	22433	2478	11.0	2808	14.1	0.7	0.5
G	22145	1383	6.2	2342	11.3	0.7	0.5
H	17842	654	3.7	3865	22.5	1.0	0.5
I	14701	2816	19.2	1636	13.8	1.1	0.5
J	15904	314	2.0	1365	8.8	0.7	0.3
K	11227	2291	20.4	938	10.5	1.2	0.6
L	18586	6498	35.0	426	3.5	0.9	0.4
M	20174	2023	10.0	178	1.0	1.2	0.6
N	16319	4150	25.4	2011	16.5	2.0	1.3
O	17319	3182	18.4	2638	18.7	0.6	0.4
P	7089	338	4.8	605	9.0	2.1	1.0
Q	20128	979	4.9	2288	11.9	1.9	0.9
R	17580	5471	31.1	10913	90.1	0.2	0.4
S	20129	5987	29.7	10907	77.1	0.5	0.8
T	18549	4389	23.7	13951	98.5	0.1	0.1
U	5577	278	5.0	5184	97.8	0.5	0.3

Table 4.5: Cooling consumption profile summary for each validation household, after reclassifying furnace fan loads while the home's air conditioner is active as cooling consumption.

4.3.1 Heating and cooling activity detection: Accuracy

While the magnitude of heating or cooling loads throughout the year provides clear insight into a home's energy context, in many situations simply knowing when a thermal system was active or not can be equally enlightening. A rudimentary metric for assessing the relative ability of the various models to infer heating or cooling system activity is the accuracy with which the model can predict an "On" or "Off" state, where "On" corresponds to a nonzero system load (estimated or observed), and "Off" indicates the circuit is not currently drawing any electrical power. With this definition, accuracy indicates the ratio of the number of hours the system state was estimated correctly, according to ground-truth disaggregation data, relative to the total number of hours sampled. Figure 4.2 provides a box-plot representation of the accuracy results for each model across all validation households.

For heating activity detection, the DAR-Fixed4 model provided the best accuracy on average out of all of the models studied (approximately 68%), with most other models following behind closely, and DAR-Fit3 lagging furthest behind with an average accuracy of 51%.

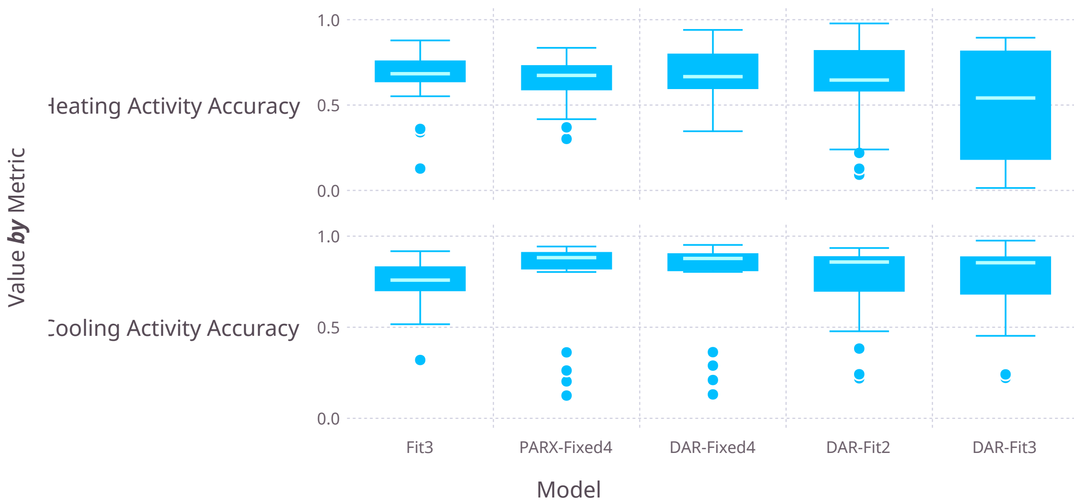

Figure 4.2: Accuracy distribution for heating and cooling system activity detection across the 21 validation homes for each of the compared models.

However, no model consistently outperformed any other across all households. Cooling activity detection had similar results, with DAR-Fixed4 again performing best at a 76% average, and all other models trailing closely (the lowest, Fit3, still averaged 72%). These averages are somewhat lower than the medians shown in Figure 4.2 due to the effects of outliers: in particular, the top-scoring models (DAR-Fixed4 and PARX-Fixed4) also demonstrated the worst results overall in outlying cases.

While accuracy serves as an intuitive measure of model performance, it can be flawed, as hinted at by the somewhat contradictory results above. Specifically, cases where one state (i.e. "On" or "Off") dominates the observations can mask a poor-performing model. (For example, consider a fundamentally flawed model that assumes no households have air conditioners: a validation household using its air conditioning system two weeks a year would still result in the model being considered at least 96% accurate!) Precision and recall provide more robust (but less intuitive) alternate metrics for assessing state detection performance and are discussed below.

4.3.2 Heating and cooling activity detection: Precision, recall, and F1 statistic

In a system activity detection context, precision (Van Rijsbergen, 1979) is defined as the ratio of number of hours in which a system was *correctly* estimated to be active out of the total number of hours in which the system was estimated to be active (whether correctly or not). Conversely, recall (Van Rijsbergen, 1979) is the ratio of number of hours in which a system was *correctly* guessed to be active of of the total number of hours the system was *actually* active. These metrics can provide a much more nuanced assessment of model performance: for example, the hypothetical fundamentally flawed model discussed in the preceding section would be assessed with a recall of 0% and an undefined precision (since there were zero hours in which it guessed that the system was active).

There are still limitations to these metrics taken individually: for example, a model estimating that an air conditioner was continuously active year-round would likely score a very low precision but still have recall of 100%. Alternately, a model only inferring AC activity on the hottest, most humid of days may have a precision of 100% but a very low recall. The F1 statistic balances these tradeoffs by taking the geometric mean of the two metrics to generate a single value. Precision, recall, and F1 values for heating and cooling activity detection are visualized in Figures 4.3 and 4.4.

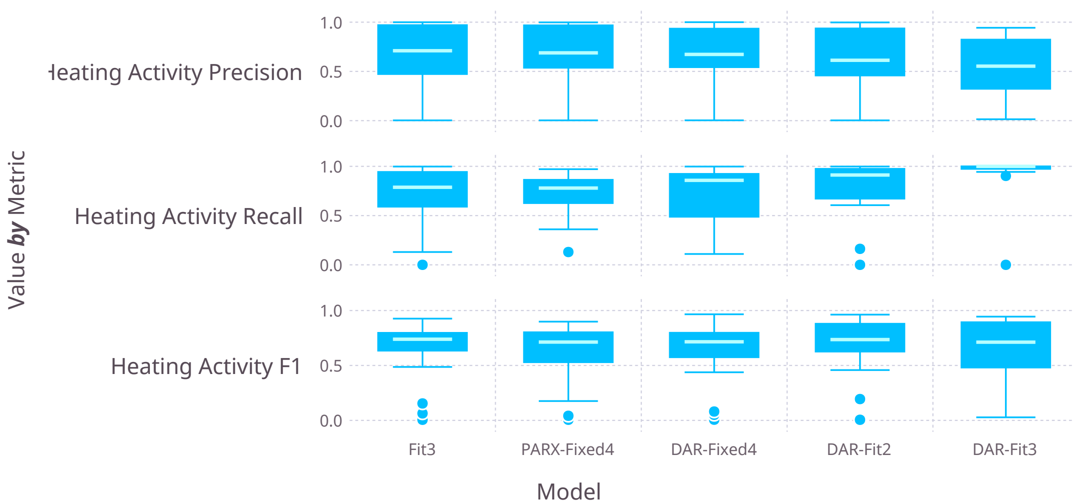

Figure 4.3: Precision, recall, and F1 statistic distribution for heating system activity detection across the 21 validation homes for each of the compared models.

The models scored mostly similarly on heating activity detection precision, with DAR-Fit3 again scoring the lowest. That same model scored by far the highest on recall, however, resulting in generally similar average F1 statistics across all models, with DAR-Fit2 providing slightly best performance compared to the rest.

The Fixed4 thermal models (PARX-Fixed4 and DAR-Fixed4) yielded the highest cooling activity precision but the lowest recall, again resulting in roughly equivalent, near-50% average F1 statistics across the board.

4.3.3 Heating, cooling, and behavioural load inference: Disaggregation error

The root mean square error (RMSE) of disaggregation estimates relative to observed circuit loads provides a richer understanding of the ability of the various models to infer load

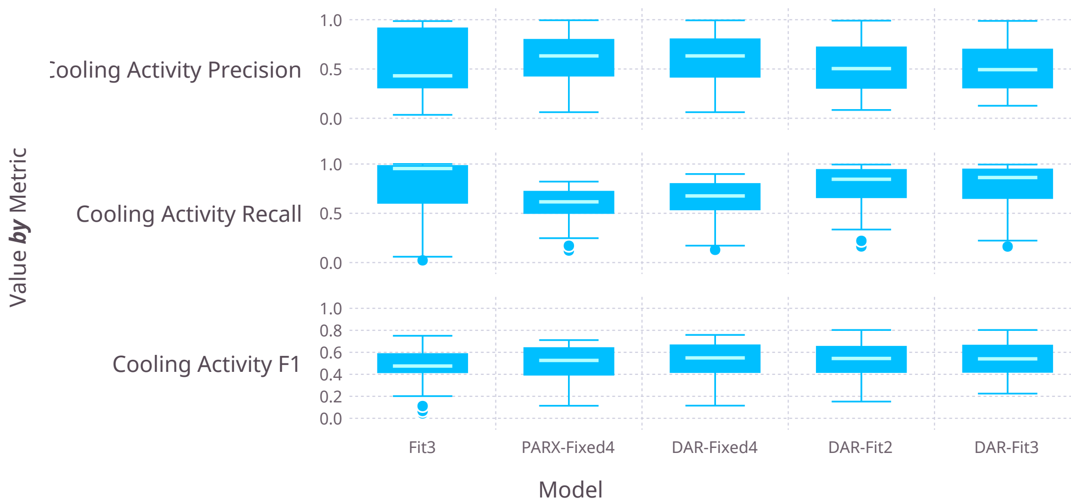

Figure 4.4: Precision, recall, and F1 statistic distribution for cooling system activity detection across the 21 validation homes for each of the compared models.

types from whole-house aggregate smart meter data. RMSE ranges as determined from the various validation homes are visualized in Figure 4.5.

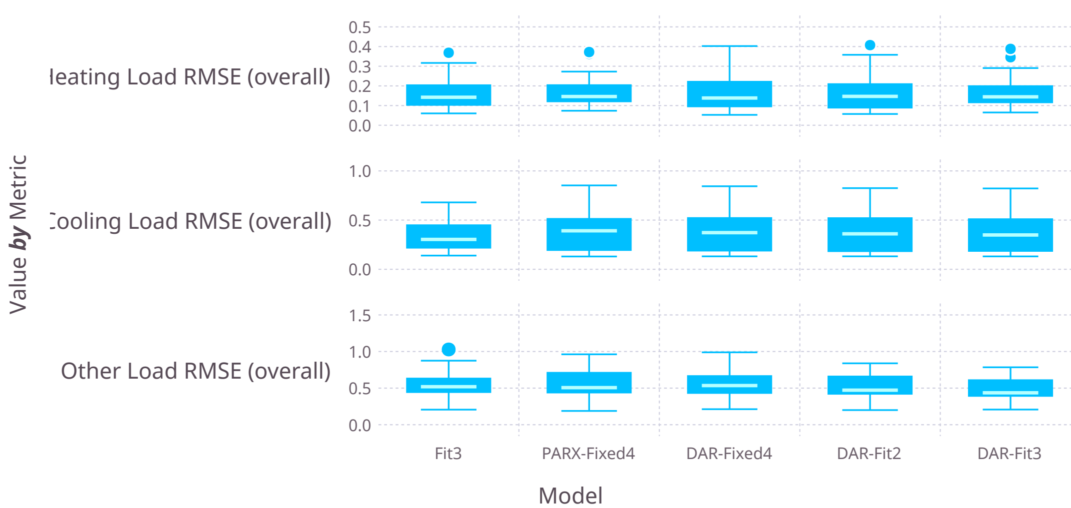

Figure 4.5: Root mean square error distribution for disaggregated heating, cooling, and behavioural loads across the 21 validation homes for each of the compared models.

All of the assessed models achieved very similar average heating RMSE performance at around 0.17 kWh/h error. Cooling loads were also similar, with Fit3 demonstrating a slight advantage, averaging at 0.34 kWh/h error compared to the 0.35-0.4 kWh/h range for the other models. Behavioural loads were also similar, with DAR-Fit3 demonstrating slightly better average performance at 0.49 kWh/h, while the other models gave errors closer to 0.55 kWh/h.

While these results provide a good overview of average model performance, they are derived from data that may include trivial thermal load estimates (for example, it would not require a particularly sophisticated model to correctly estimate that a given air conditioner load remains zero throughout all of December and January). Repeating the RMSE calculations including only samples where either the estimated or true circuit loads were nonzero (i.e. an active system) can better discriminate between model abilities and provide a clearer performance comparison. Results for this metric are visualized in Figure 4.6.

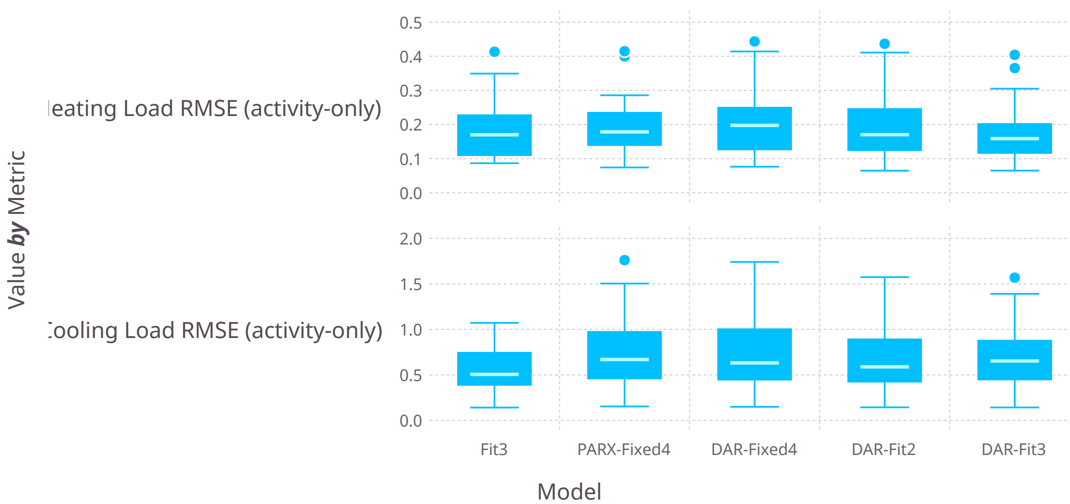

Figure 4.6: Root mean square error distribution for active (either estimated or true) disaggregated heating and cooling loads across the 21 validation homes for each of the compared models.

Relative results for heating disaggregation under metric remained similarly close, while cooling load errors increased by different amounts, with Fit3 retaining the best performance (averaging under 0.6 kWh/h error) and PARX-Fit4 having the worst (averaging almost 0.8 kWh/h error).

Finally, while RMSE provides an indication of the absolute magnitude of average error in model estimates, its values need to be considered in context: an RMSE of 0.2 kWh/h is much more impressive when considered with a true load at 20 kWh/h than with 0.3 kWh/h. To consider the relative size of error, the mean absolute percent error (MAPE) can be used instead. Results for this metric are given in Figure 4.7.

Assessing the relative disaggregation estimate error indicates that errors are large relative to ground truth across all models, with some households as outliers that are even more severe. For heating estimates, DAR-Fixed4 performed the best (averaging at approximately 111% error), while DAR-Fit3 performed the worst (averaging 230% error). Relative cooling

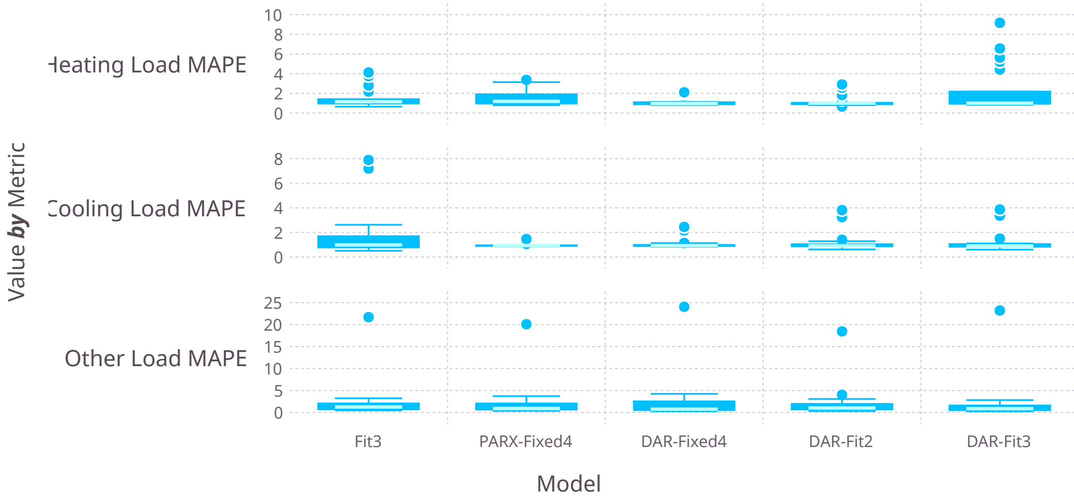

Figure 4.7: Mean absolute percent error distribution for disaggregated heating, cooling, and behavioural loads across the 21 validation homes for each of the compared models.

estimate errors were similarly large with average values ranging from 97% error for PARX-Fixed4 to almost 172% error for Fit3.

MAPE values when disaggregating other behavioural loads were larger than either of the the heating or cooling case: DAR-Fit3 performed the best (averaging approximately 210% error) while the remainder were higher but similar (up to 243% for DAR-Fixed4).

4.3.4 Disaggregation performance summary

Table 4.6 provides an overview of performance metric values for each model, averaged over all households in the dataset. Table 4.7 provides rankings for each of these performance averages on each metric. In general, the DAR-Fit models (particularily DAR-Fit3) provided the best behavioural disaggregation while the best thermal disaggregation was split between Fit3 in the absolute (RMSE) case and the Fixed4 models in the relative (MAPE) case. These results are discussed further in Section 5.1.

4.4 Physical characterization performance assessment

While the ability to infer disaggregated electrical loads from whole-house smart meter data provides useful insight into a household's energy use patterns, it does not in itself provide a description of a household's physical characteristics. Knowledge of factors such as the age and size of a home and the nature of its mechanical and electrical systems can be highly desirable in particular contexts. It is therefore useful to assess the ability of the various models under study to distinguish between homes with different such characteristics.

	Fit3	PARX Fixed4	DAR Fixed4	DAR Fit2	DAR Fit3
Heating Activity Accuracy	0.65	0.65	0.68	0.63	0.51
Heating Activity Precision	0.66	0.66	0.66	0.64	0.54
Heating Activity Recall	0.71	0.72	0.70	0.74	0.89
Heating Activity F1	0.65	0.63	0.63	0.69	0.64
Cooling Activity Accuracy	0.72	0.76	0.76	0.74	0.74
Cooling Activity Precision	0.52	0.61	0.61	0.53	0.53
Cooling Activity Recall	0.77	0.56	0.61	0.75	0.74
Cooling Activity F1	0.46	0.49	0.51	0.51	0.52
Heating Load RMSE (overall)	0.16	0.17	0.17	0.17	0.17
Cooling Load RMSE (overall)	0.34	0.40	0.39	0.38	0.37
Other Load RMSE (overall)	0.54	0.56	0.55	0.52	0.49
Heating Load RMSE (activity-only)	0.19	0.20	0.20	0.20	0.18
Cooling Load RMSE (activity-only)	0.56	0.79	0.76	0.67	0.68
Heating Load MAPE	1.47	1.51	1.11	1.17	2.30
Cooling Load MAPE	1.72	0.97	1.07	1.19	1.20
Other Load MAPE	2.30	2.20	2.43	2.15	2.10

Table 4.6: Average metric values for each model over all households

	Fit3	PARX Fixed4	DAR Fixed4	DAR Fit2	DAR Fit3
Heating Activity Accuracy	2	3	1	4	5
Heating Activity Precision	1	2	3	4	5
Heating Activity Recall	4	3	5	2	1
Heating Activity F1	2	5	4	1	3
Cooling Activity Accuracy	5	2	1	4	3
Cooling Activity Precision	5	2	1	3	4
Cooling Activity Recall	1	5	4	2	3
Cooling Activity F1	5	4	3	2	1
Heating Load RMSE (overall)	1	5	4	2	3
Cooling Load RMSE (overall)	1	5	4	3	2
Other Load RMSE (overall)	3	5	4	2	1
Heating Load RMSE (activity-only)	2	3	4	5	1
Cooling Load RMSE (activity-only)	1	5	4	2	3
Heating Load MAPE	3	4	1	2	5
Cooling Load MAPE	5	1	2	3	4
Other Load MAPE	4	3	5	2	1

Table 4.7: Ranking of average metric values (according to whether higher or lower values are more desirable) for each model over all households

The EHMS dataset provides some limited insights into the physical characteristics of the participant houses. In particular, it provides approximate age and square footage for each home. Plotting this data (Figure 4.8) reveals two distinct "reference" groups of houses:

one older, generally smaller set of homes, and a second set with a wider range of sizes.

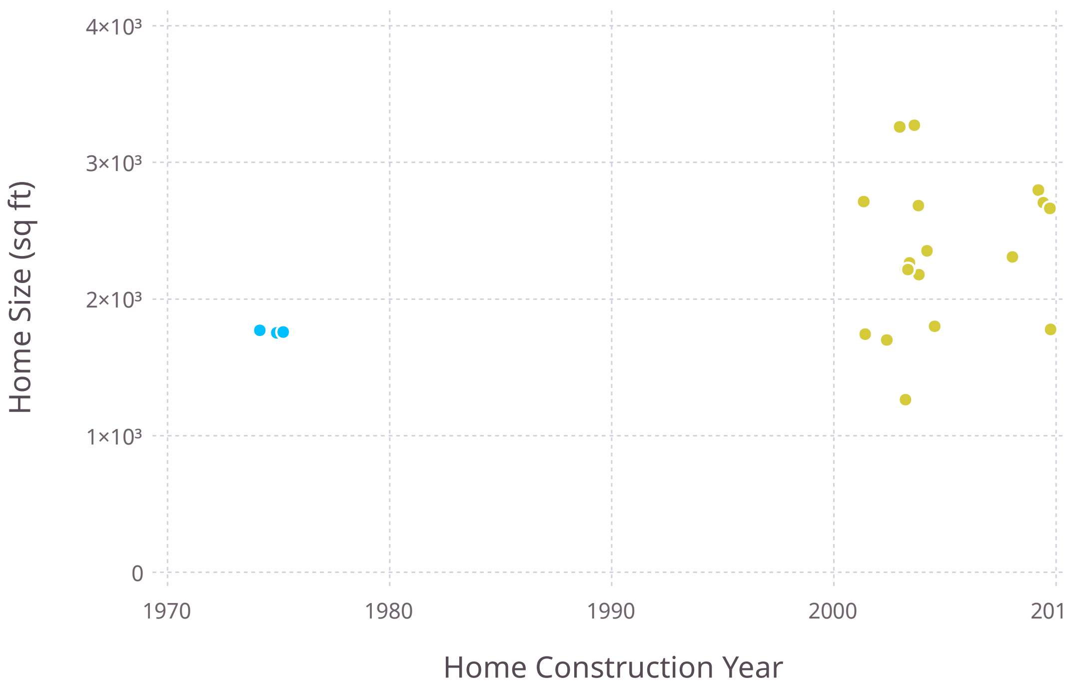

Figure 4.8: Ground-truth cluster assignment of EHMS homes. Two distinct groups (coloured blue and yellow for reference) are apparent in the dataset.

In theory, a well-performing model capable of discriminating between different "kinds" of physical houses would fit model parameters that reflected the differences in home types. To test the ability of the various models under study to perform this task, estimated model parameters corresponding to physical traits (e.g. heating and cooling temperature sensitivities) for each household were used as input features in a k-means clustering operation to split the validation household set into two groups. The resulting clusters (as determined by the parameters from a particular model) were then compared to the aforementioned reference grouping generated from the known household characteristics, with similarity assessed by the Adjusted Rand Index (ARI).

An ARI score of 1.0 corresponds to perfectly identical grouping, while a score of 0.0 denotes that any similarities in cluster assignments are no more than would be expected to happen by chance. Clustering for each model was repeated and ARI scores generated 100 times per model to account for fluctuations introduced by random initialization of the k-means clustering process. The distribution of results of this assessment are plotted in Figure 4.9. For each model, the cluster assignment resulting in the highest ARI value achieved is visualized in Appendix C.

Of the models studied, only DAR-Fixed4 was able to frequently score above zero (averaging at approximately 0.04): the remaining models all averaged below zero, or worse than

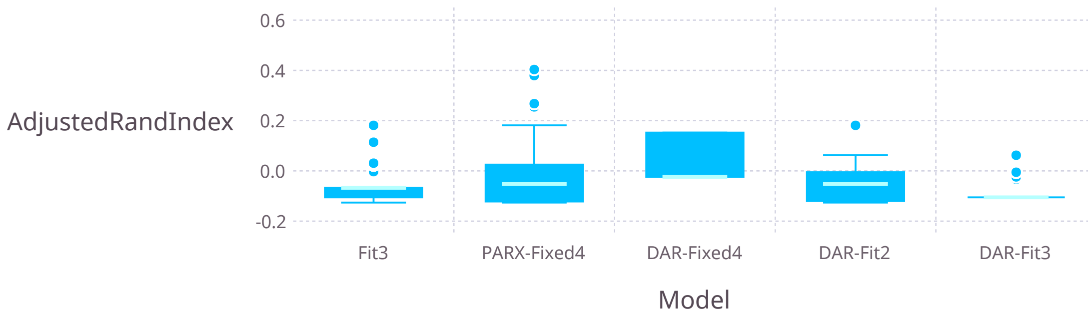

Figure 4.9: Adjusted Rand Index value ranges resulting from 100 independent k-means clustering operations on each model's fit physical parameters

random guessing could be expected to score.

4.5 Summary

The relative strengths and weaknesses of five residential electricity consumption models in characterizing a set of households have been compared according to multiple measures of accuracy in heating and cooling system load detection and estimation, behavioural load estimation, and dwelling characterization and grouping. The following chapter will discuss the implications of the relative and absolute performances of the various models, the dataset used to derive the results, and explore possible extensions and applications of these models moving forward.

Chapter 5

Discussion

The assessment results presented in the previous chapter should be understood and interpreted in context relative to the validation dataset and practical requirements of potential application scenarios: this chapter will provide that contextualization and interpretation. The resulting insights will then be used to inform a discussion concerning the required next steps in model development and considerations of how such models could contribute to sustainable energy transitions moving forward.

5.1 Performance assessment analysis

5.1.1 Heating loads

Heating activity detection was generally consistent across all models, with the fixed-temperature-threshold thermal models suffering no notable performance penalty relative to the fit-threshold alternatives. All models outperformed random guessing, but the approximate 65% reliability level consistently achieved leaves considerable room for improvement.

The models were equally evenly matched when considering absolute disaggregation error, with the Fit3 and DAR-Fit3 models showing slight advantages, perhaps due to their higher parameter count yielding a greater ability to adapt to provided data. Counterintuitively, those same models were among the worst performers when considering relative disaggregation error, in part due to extreme outliers, although with average errors consistently exceeding 100%, no model can claim particularly impressive performance in that area. Of course, this may not matter for gas-heated homes, where fan load is at best a rough proxy for a furnace's fuel consumption rate. In this context, knowledge of when and for how long the furnace is active (activity detection) may in fact be the more useful metric in assessing a home's heating requirements.

There is no clear best-performing model for heating disaggregation. Of the seven heating metrics evaluated, almost every model (with the exception of PARX-Fixed4) placed first in some category: Fit3, DAR-Fixed4, and DAR-Fit3 each placed first in two. This is not unsurprising given the similar nature of the various thermal models (Fit2, Fit3, Fixed4) applied here: future research developing and studying a broader range of thermal modelling approaches would be well-warranted.

5.1.2 Cooling loads

As in the heating case, no model was able to establish itself as dominant in cooling load detection or disaggregation. While the adaptability of the fit-threshold models (Fit3, DAR-Fit2, DAR-Fit3) provided slightly better RMSE performance, as in heating it came at the cost of decreased resilience to the effects of outlying data, as illustrated by the extreme MAPE values resulting from fit-threshold models in certain households. The fit-threshold models tended to more readily infer air conditioner activity, resulting in strong performance in detecting cooling action when it was happening, but also leading to numerous false positives. Conversely, the fixed-threshold models (PARX-Fixed4 and DAR-Fixed4, with cooling thresholds set to 20 °C) were more conservative in their estimates, somewhat reducing false positives but also increasing true negatives as a result: at some points in time the predetermined cooling threshold was too low, while in others it was too high.

The seeming inability for any model to excel at precision and recall simultaneously (yielding low F1 statistics) could indicate a fundamental information deficit in predicting cooling activity based solely on some external temperature threshold (whether fit from the data or predetermined). This may best be illustrated by the performance of the Fit3 model, which generated estimates based exclusively on outside temperature, and yielded the overall highest recall but lowest precision and F1 statistics, and the most extreme MAPE values. The DAR-Fit2 and DAR-Fit3 models' cooling estimates were generated from a similar process but were moderated by time-of-day and -week considerations, resulting in less extreme metrics and slightly improved overall detection performance. New exogenous information sources may be required to generate more accurate estimates of cooling activity (see Section 5.4.1 below).

5.1.3 Behavioural loads

Each of the models demonstrates a roughly similar aptitude for estimating the non-thermal component of whole-house electricity use. DAR-Fit3 provides the best performance, averaging slightly under 0.5 kWh/h RMSE, while the other models have average RMSE values slightly greater above 0.5 kWh/h. DAR-Fit3 also provides the best MAPE performance (relatively speaking), despite averaging 210%. While that model provides marginally better performance across both metrics, it should still be noted that no single model consistently outperforms any other.

5.1.4 Physical characteristics

In general, all of the models assessed are essentially ineffective in providing better-than chance household grouping based on fit parameters corresponding to physical properties of a dwelling. Perhaps unsurprisingly, the best performance is achieved by the PARX-Fixed4 and DAR-Fixed4 models, those with three parameters corresponding to physical characteristics as opposed to two (temperature thresholds are taken here as behavioural parameters relating to thermostat setpoints, eliminating any potential advantages of the more adaptable fit-threshold models).

This poor performance may be in part due to fundamental differences between criteria for the ground-truth grouping and the groupings inferred from parameters: while the base groups were assigned based on home age (primarily) and size, the clustering features related to the home's building envelope. While it may be reasonable to assume some correlation between these different characteristics, they by no means represent a direct mapping: older homes can easily have differing levels of weatherproofing, just as newer homes can as well. Data permitting, a fairer assessment would have provided a ground-truth grouping of the homes according to some measure of their insulation levels or draftproofing, rather than age or size.

Of course, there still remains much potential to improve the discriminatory power of the models themselves. A more sophisticated physically-derived model has the potential to much better capture specific technical properties of a home: for example, a model built on thermodynamic principles could capture the effects of potentially-distinguishing building properties such as thermal mass and the relative rates of internal and external temperature changes. Such a possibility is explored more fully in Section 5.4.2 below. Incorporating new exogenous sources of environmental data may provide another opportunity to capture novel physical building characteristics - some possibilities are outlined further in Section 5.4.1.

5.1.5 Overall performance

DAR-Fit3 emerged as the best choice for disaggregating non-thermal loads and a strong choice (among a generally equivalent field) for both heating and cooling disaggregation. While DAR-Fixed4 provided the best performance for clustering homes according to physical characteristics, no model consistently performed better than random chance at grouping, indicating that none should be relied upon for that purpose. Overall, DAR-Fit3 would appear to be the best performer, although not by a particularily notable margin.

The significantly higher complexity and parameter count of the behavioural component of the PARX-Fixed4 model did not appear to provide it with any performance benefit over the simpler and more interpretable DAR approach to non-thermal modelling (it is plausible that the PARX approach would perform better in a load forecasting context, however such an application is beyond the scope of household characterization tasks studied here). In fact, in many cases the PARX approach performed no better than the non-behavioural Fit3 model, illustrating the power of a well-designed parsimonious modelling method.

5.2 Validation data limitations

5.2.1 Sample size and homogeneity

The performance assessment results reported here are derived from averaging household-level metrics across all of the homes in a validation dataset. That dataset was comprised of only 21 homes located in a single town, each with similar thermal system characteristics (an air conditioning unit and a natural gas furnace). As such, the resulting evaluation fails to assess model performance in other situations, such as jurisdictions with different sociodemographic contexts, climates, and housing bylaws and building codes, or dwellings with different thermal systems such as electric resistive heating or ground-source heat pumps. A more authoritative assessment of the relative merits of these models would include validation data from households representing a wider, more representative range of possible configurations.

5.2.2 Missing observations

The validation dataset contained extended blocks of missing or invalid electrical load readings – totalling up to more than 30% of the study period in some cases – possibly due to intermittent failures in the energy management system's monitoring hardware or logging software. Such observation gaps have clear negative consequences when fitting statistical models, which stand to be exacerbated by the serially-correlated nature of the data in question. For example, some information describing features of interest, such as the sensitivity of a home's electrical demand to hot weather, will be inherently unevenly distributed across time: a block of missing data occurring in July or August has the potential to disproportionately and severely impair accurate cooling gradient parameter fitting, for example, or unfairly disadvantage certain models over others (an otherwise strong cooling disaggregator may yield similar performance to a weaker model if the necessary cooling season observations are disproportionately missing).

Missing data also prevented the assessment of specific model types - while Markov modelling approaches are growing in popularity in the electrical load disaggregation literature, they were purposefully excluded from this analysis given their inability to explicitly handle missing values. As a result of this issue, only regression-based approaches were able to be studied in detail here.

5.2.3 Possible unlabelled or mislabelled heating and cooling loads

Observed furnace fan data was corrected in this analysis to represent a cooling load when an air conditioning unit was also active, while remaining a heating load in all other cases. It is reasonable to expect, however, that in some cases a furnace fan was running in spite of neither the furnace nor the air conditioner being active (simply circulating unheated air through the home). This would still result in erroneously-observed activity in ground-truth heating load data, skewing analysis results in favour of more liberal activity detection methods.

Additionally, while the validation dataset provided labelled circuit-level electrical load data for major appliances in each household, it was in general not able to account for particular uses of wall outlets, which may have contributed to additional thermal electrical loads (for example, floor fans for effective cooling purposes or space heaters for auxiliary heating). As a result, the supposed ground truth loads used to evaluate model disaggregation performance may not have accurately reflected the full thermal load of a home at any given time.

Conversely, there is no guarantee that the labelled air conditioner and furnace fan circuits in each home were used exclusively for their eponymous purpose: it is possible that additional loads could have been present in some cases, again reducing the accuracy of the reported "true" thermal loads. One instance of this issue observed in the data was the case of air conditioner circuits reporting low levels of activity through winter months.

5.3 Possible analysis extensions

As noted above in Section 5.2.1, the overall model performances reported here are based on aggregating results from a relatively small set of households. Several metrics, however, yield extreme outlier values that could distort reported averages. The numeric averages reported in Chapter 4.3 are supported by box-plot visualizations, providing sample median and distribution data in order to increase resilience under the effects of outliers. However, the possibility remains that certain homes in the validation dataset may be particularly well- or poorly-suited to fitting to a given model, whether due to inherent characteristics of the home or data quality issues such as those outlined in the previous section. Such outlying households may disproportionately punish or favour specific approaches when compared to a "typical" household of interest. What's more, as previously discussed, some characteristics of the overall group of validation households may not be representative of actual homes in a particular region of interest for model application.

To address these potential challenges, the metric aggregation process could be performed differently, weighting houses differently according to some criteria. For example, the issue of outlier households could be addressed by determining which homes are most similar to each other and giving less weight to the assessment results of those that do not conform to overall trends. The group-level misrepresentation issue could be addressed by weighting individual results by the degree to which validation households are representative of some known "ideal" or "typical" household or match a set of auxiliary variables with known population distributions, much like the practise of survey response weighting (Biemer and Christ, 2008).

5.4 Possible model extensions

5.4.1 Additional covariates

Disaggregation models to date have tended to focus exclusively on the relationship between thermal system electrical load and exterior temperature. However, there are likely other environmental factors that contribute to electrical demand. Given that air conditioners provide both thermal regulation and dehumidification, it would stand to reason that ambient humidity could contribute to cooling system loads. Passive heat gains due to incident solar irradiation could affect not only cooling season electrical use, but cold-weather heating requirements as well, and possibly also behavioural loads (through artificial lighting requirements). Poorly-sealed and draft-prone building envelopes could also introduce relations between external wind speeds and heating or cooling requirements.

5.4.2 Deeper physical modelling approaches

The Fit2 physical model is predicated on a somewhat unrealistic assumption that a dwelling's interior temperature is strictly maintained at a fixed value throughout the entire year. As noted previously, it is more likely that ambient internal temperatures would be maintained between some range of comfortable values: this relaxed assumption was the driver behind the adoption of the Fit3 physical model. The generally superior performance of the Fit3 model over Fit2 in the model assessment would seem to support this line of reasoning.

Unfortunately, the Fit3 model represents a regression from the more rigourous, physically derived nature of the Fit2 model. A more physically-informed approach would model interior temperature changes as a function of both thermal system activity and heat flows through conduction (as considered in the Fit2 model) or other mechanisms (such as the exfiltration and passive solar effects discussed in the previous section). Changes in this *interior* temperature above or below particular thresholds could be assumed to activate the relevant heating systems - more closely modelling the actions of a thermostat. As noted in previous sections, this approach would also result in fitting a larger number of physically-oriented model parameters, potentially improving the discriminatory power of a model to distinguish between various kinds of dwellings. A preliminary approach to developing such a model is provided in Appendix D.2.1.

5.5 Possible applications

The parameters estimated by current models - and any improved characterizations that future extended models might provide - can be operationalized to improve residential energy efficiency goals in a number of different ways. The most obvious application may be targeted engagement and outreach for utility and government energy efficiency programs - characterizations could be used to identify homes with the highest sensitivity of heating or cooling loads to changes in external temperatures, and efforts to improving the weatherproofing of the dwelling's building envelope or the efficiency of the home's heating or cooling systems could be prioritized through offers to participate in energy audit programs or retrofit incentives and subsidies.

Characterization data may also suggest opportunities for customized behavioural interventions or feedback: abnormally high heating thresholds or low cooling thresholds may indicate a household that activates thermal systems in mild weather instead of opening or closing windows - the occupants may be able to achieve their desired temperature regulation goals in a more energy-efficient manner, and may be promising targets for an educational outreach campaign run by a utility or environmental organization. Similarly, average hourly consumption trends could be studied to identify opportunities for shifting heavy discretionary electrical loads away from peak consumption hours, reducing total grid demand and saving energy costs if the household is in a jurisdiction using time-of-use electricity pricing - personalized savings estimates could even be generated based on historical consumption patterns and provided to a homeowner as part of a utility's demand response initiative.

Rather than providing actionable insights to specific households, the behavioural and thermal trends derived from characterization models could also be used to group a large sample of households into smaller subgroups according to similarities in a feature of interest. For example, homes with a very strong relationship between low external temperatures and high electrical loads may be clustered together as likely resistive-electrically-heated, while a more moderately sensitive group may be considered as likely users of heat pumps, and a homes with a weak cold-weather vs electrical load relation may be inferred to use natural gas furnaces.

These separate groups could then be employed as enhanced control groups in intervention analyses. As an example, such clustering might be desirable when assessing the impact of an informational weatherproofing campaign - the campaign may in fact be highly successful in inspiring building envelope enhancements among gas furnace users, but the more subtle ex-post electricity effects on the intervention group may be missed if the study's control group (who did not receive the informational campaign) was dominated by electrically-heated homes with much higher and more variable load profiles - ideally an intervention household would be compared to a control group with similar household characteristics. Likewise, an intervention study adopting a synthetic control approach (Abadie et al., 2010) would be much better served by comparing a participating household to a large group of homes with very similar inferred energy contexts than a control group chosen at random.

Chapter 6

Conclusions and Recommendations

This research has attempted to determine the extent to which existing methodological tools and techniques may be used to infer the "energy context" of a given household based on its energy consumption data and other readily-available information sources, and the degree to which the state-of-the art in this area could be improved. To this end, a review of existing techniques in the literature was performed and applied in combination with physical first principles and empirically-observed statistical trends to develop new residential electricity modelling approaches aiming to improve upon existing efforts.

The performance of a subset of these models (chosen for their speed and stability of parameter estimation) were then compared to existing techniques described in the literature. The results of the analysis are summarized in Section 6.1. While one of the novel approaches yielded overall improved behavioural disaggregation performance and a simpler formulation compared to alternatives in the literature, there would seem to remain considerable opportunity for further improvement. Several potentially-promising areas for continued research are outlined in Section 6.2. Finally, several recommendations for energy practitioners and policy-makers arising from this work are outlined in Sections 6.3 and 6.4.

6.1 Research results

6.1.1 Model performance assessment

Currently available tools seem best-equipped to detect heating loads, with the existing and novel techniques studied yielding roughly comparable performance, detecting heating system activity in an average of 65% of cases, and estimating the corresponding load with a root mean square error of less than 0.2 kWh/h. Cooling system load detection and estimation was also approximately equivalent across the various models studied, but with poorer average performance (averaging just under 0.5 kWh/h for the best-case model). Introducing new external inputs to models may help to improve estimate accuracy.

Absolute error in behavioural (non-thermal) load detection was generally larger than that for heating or cooling load components. A novel technique combining deseasonalized autoregression and three-regime threshold regression (referred to in this document as DAR-Fit3) performed the best relative to the alternatives studied, in terms of minimizing both absolute and relative error.

Attempts to partition the test dataset households according to age and size based on physical model parameters yielded results no better than could be expected from random assignment. This may be a result of a combination of a lack of discriminatory power of the models studied and the inability of home size and age to correlate to the characteristics represented by the model parameters (representing the sensitivity of home energy requirements to external temperature).

Overall, the DAR-Fit3 model introduced here provided the best overall performance for the given test data, although it did not significantly surpass any other model studied. Based on these results, there remains significant room for improvement in the state-of-the-art of published strategies for thermal load disaggregation and dwelling characterization when working with datasets with significant numbers of missing observations (techniques requiring complete datasets were not able to be assessed - see Section 6.2.1 below).

6.1.2 Emergent insights

In many cases over the course of the assessment, simpler models were able to outperform or at least match the performance of more complicated approaches, while also providing more tractable, interpretable parameter fits. In particular, the PARX behavioural modelling approach involved fitting several hundred model variables, but was often bested by the simpler DAR method which fit an order of magnitude fewer parameters, resulting in quicker computations and easier interpretation into policy insights.

6.2 Open research questions

6.2.1 Performance of less resilient models

Certain notable thermal disaggregation approaches were excluded from the assessment given their lack of robustness in situations with large amounts of sequential missing observations (as was the case in the validation dataset available for this study). It is also plausible that the relative performance of the models studied here may also change under more favourable data conditions. Further consideration and testing of model performance using a higher-quality dataset is warranted.

6.2.2 Novel model inputs

The lack of benefit of fitting temperature thresholds in predicting air conditioner use and the failure of any model to provide improved inference of known dwelling characteristics highlights the fact that hourly electrical loads and external temperatures alone do not provide the information content required for a complete contextualization of a household's energy use. There are other variables that are widely available but not being used in an integrated manner: additional weather times series including external sunlight, humidity, and wind speed provide basic examples. While smart metering of natural gas services has lagged behind electricity, such data are becoming more commonly available and would seem to have to potential to provide significant simplifications and performance improvements to modelling household heating use. Further study is warranted into the performance impacts of expanding models to incorporate these different inputs, and the magnitude of their relative contributions to improving estimation accuracy.

6.2.3 Deeper physical-statistical model integration

Many of the models considered here eschew physically-derived foundations in favour of statistical expedience and standard modelling techniques. While this approach simplifies parameter fitting and improves model parsimony, there may be additional insights to be gleaned from a more rigorous model construction process informed by the underlying thermodynamic processes driving household heating and cooling requirements, and less strict assumptions about the nature of occupant behaviour and preferences. Additional work exploring the trade-offs and performance limits of more detailed modelling techniques given the non-deterministic nature and inherent variability of household energy use would be justified.

6.3 Recommendations for practitioners

While there is no doubt more work to be done in improving the quality of residential energy model estimates and characterizations, existing approaches already provide a usable starting point for practitioners (such as utility analysts, policy researchers, or individual homeowners) seeking to make better use of available energy data. None of the models studied are likely to be the best possible option for all applications: when selecting a model to apply, one should consider the nature of the problem at hand and which kinds of characterization are most important, striving to balance model intricacy with parsimony and interpretability. The most intricate or complicated model available may not be the best option.

Different models have been shown here to excel in different areas: for grouping homes according to physical characteristics, a model with more physical parameters may be of use. Heating load estimations may benefit more from allowing for flexibility in temperature threshold points. Behavioural characterizations may be easier to interpret when dealing with a limited number of highly relevant parameters instead of spreading potential insights across a large number of parameters, even if they may help to fit the data better in some situations.

6.4 Recommendations for policy-makers

The capabilities of models are fundamentally limited by the information available to them: richer, higher-quality data yield richer, higher-quality insights, and possibilities for combinatoric innovation increase exponentially relative to the number of different types of datasets available. With that in mind, policy-makers should continue to support initiatives that provide and incentivize the simplified exchange of datasets, striving for (or better yet, mandating) high data quality and completeness in order to broaden the range of available analysis techniques and make full use of any capital investment that was required to obtain the information.

In the case of residential energy use specifically, that could involve making it simpler for authorized agents to access timely and precise information about a household that the the homeowner or occupants wish to share. The Green Button standard is an example of an emerging means of facilitating this data access, and its continued adoption by utilities should be encouraged. Real estate records are another example of rich datasets with important implications for residential efficiency initiatives that have historically been closed and costly to access: more open access mechanisms could yield tremendous benefits for dwelling energy modelling, for example. (Of course, such dwelling records provide no guarantee that their contents are fully up-to-date: energy data would remain a valuable resource for assessing the current state of a household.)

Navigating the transition to a sustainable energy future will require a deep reassessment of the efficiency with which society procures energy services to achieve its desired ends. This reassessment involves both widespread and substantial behavioural shifts as well as physical enhancements to existing infrastructure, including residential dwellings. These are non-trivial undertakings: achieving them will require society to make full use of all of the social- and knowledge-based resources at its disposal. The digitization of key informational and technical systems presents a massive opportunity to leverage data to reduce the socioeconomic costs of such a transition, and the models studied here provide means to carry that data towards urgently-required operationalization. None of the modelling options assessed here are perfect, or even close to it: there is much work yet to be done. But there is no doubt that it is work worth doing.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of californias tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.

Albert, A. and Rajagopal, R. (2015). Thermal Profiling of Residential Energy Use. *Power Systems, IEEE Transactions on*, 30(2):602–611.

Ardakanian, O., Koochakzadeh, N., Singh, R. P., Golab, L., and Keshav, S. (2014). Computing Electricity Consumption Profiles from Household Smart Meter Data. In *EDBT/ICDT Workshops*, pages 140–147.

Armel, K. C., Gupta, A., Shrimali, G., and Albert, A. (2013). Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52(0):213 – 234. Special Section: Transition Pathways to a Low Carbon Economy.

ASHRAE (2013). Thermal environmental conditions for human occupancy. Technical report, ASHRAE.

Basu, K., Debusschere, V., Bacha, S., Maulik, U., and Bondyopadhyay, S. (2015). Nonintrusive load monitoring: A temporal multilabel classification approach. *Industrial Informatics, IEEE Transactions on*, 11(1):262–270.

Belanger, N. and Rowlands, I. H. (2013). Smart Energy Networks: Progress and Prospects. Technical report, Waterloo Institute for Sustainable Energy.

Berges, M. E., Goldman, E., Matthews, H. S., and Soibelman, L. (2010). Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring. *Journal of Industrial Ecology*, 14(5):844–858.

Biemer, P. P. and Christ, S. L. (2008). *Weighting survey data*, chapter 17, pages 317–341. Taylor Francis Group, New York.

Birt, B. J., Newsham, G. R., Beausoleil-Morrison, I., Armstrong, M. M., Saldanha, N., and Rowlands, I. H. (2012). Disaggregating categories of electrical energy end-use from whole-house hourly data. *Energy and Buildings*, 50(0):93 – 102.

Bishop, J., Amaratunga, G., and Rodriguez, C. (2010). Quantifying the limits of HANPP and carbon emissions which prolong total species well-being. *Environment, Development and Sustainability*, 12(2):213–231.

Bozchalui, M. C., Hashmi, S. A., Hassen, H., Canizares, C. A., and Bhattacharya, K. (2012). Optimal operation of residential energy hubs in smart grids. *IEEE Transactions on Smart Grid*, 3(4):1755–1766.

Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42(1):68 – 80. 8th World Energy System Conference, {WESC} 2010.

Depuru, S. S. S. R., Wang, L., and Devabhaktuni, V. (2011). Smart meters for power grid: Challenges, issues, advantages and status. *Renewable and Sustainable Energy Reviews*, 15(6):2736 – 2742.

Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Oxford Statistical Science Series. Oxford University Press.

Espinoza, M., Joye, C., Belmans, R., and De Moor, B. (2005). Short-Term Load Forecasting, Profile Identification, and Customer Segmentation: A Methodology Based on Periodic Time Series. *Power Systems, IEEE Transactions on*, 20(3):1622–1630.

Farhangi, H. (2010). The path of the smart grid. *IEEE Power and Energy Magazine*, 8(1):18–28.

Fels, M. F. (1986). PRISM: an introduction. *Energy and Buildings*, 9(1):5–18.

Gaasch, W., Travis, J., Zhao, F., Kaplan, M., and Muth, C. (2014). A comparison of methods for early-stage retrofit analyses. In *2014 ACEEE Summer Study on Energy Efficiency in Buildings*, pages 11–149 – 11–161.

Griffiths, S. (2015). The device that makes any home smart: 250$ Neurio keeps tabs on household gadgets to tell you if your boiler's broken - or you've left your lights on. *Daily Mail*.

Guerin, D. A., Yust, B. L., and Coopet, J. G. (2000). Occupant predictors of household energy behavior and consumption change as found in energy studies since 1975. *Family and Consumer Sciences Research Journal*, 29(1):48–80.

Hay, G. J., Kyle, C., Hemachandran, B., Chen, G., Rahman, M. M., Fung, T. S., and Arvai, J. L. (2011). Geospatial Technologies to Improve Urban Energy Efficiency. *Remote Sensing*, 3(7):1380–1405.

Haykin, S. (2004). *Kalman Filtering and Neural Networks*. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley.

Hipel, K. and McLeod, A. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Developments in Water Science. Elsevier Science.

Hirst, E., Berry, L., and Soderstrom, J. (1981). Review of utility home energy audit programs. *Energy*, 6(7):621 – 630.

Hirst, E. and Brown, M. (1990). Closing the efficiency gap: barriers to the efficient use of energy. *Resources, Conservation and Recycling*, 3(4):267 – 281.

Hogör, E. and Fischbeck, P. S. (2015). Virtual home energy auditing at scale: Predicting residential energy efficiency using publicly available data. *Energy and Buildings*, 92(Complete):67–80.

Huang, D., Thottan, M., and Feather, F. (2013). Designing customized energy services based on disaggregation of heating usage. In *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*, pages 1–6.

Hutchinson, E. J., Wilkinson, P., Hong, S. H., and Oreszczyn, T. (2006). Can we improve the identification of cold homes for targeted home energy-efficiency improvements? *Applied Energy*, 83(11):1198 – 1209.

ICF Marbek (2014). Achievable Potential: Estimated Range of Electricity Savings from Energy Efficiency and Energy Management. Technical report, Ontario Power Authority, Ottawa, Ontario.

International Energy Agency (2013a). Energy flow charts.

International Energy Agency (2013b). Redrawing the Energy-Climate Map. Technical report, International Energy Agency.

International Energy Agency (2014a). CO2 Emissions from Fuel Combustion Highlights 2014. Technical report, IEA / OECD.

International Energy Agency (2014b). World Energy Outlook 2014. Technical report, OECD, Paris.

International Institute for Applied Systems Analysis (2012). *Global Energy Assessment - Toward a Sustainable Future*. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Jackson, T. (2009). The Transition to a Sustainable Economy. In *Prosperity Without Growth: Economics for a Finite Planet*, pages 171–185. Earthscan.

Jang, D., Eom, J., Park, M. J., and Rho, J. J. (2016). Variability of electricity load patterns and its effect on demand response: A critical peak pricing experiment on korean commercial and industrial customers. *Energy Policy*, 88:11 – 26.

Koksal, M. A., Rowlands, I. H., and Parker, P. (2015). Energy, cost, and emission end-use profiles of homes: An ontario (canada) case study. *Applied Energy*, 142:303–316.

Kolter, J. Z. and Johnson, M. J. (2011). Redd: A public data set for energy disaggregation research. In *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, volume 25, pages 59–62. Citeseer.

Kyriakides, E. and Polycarpou, M. (2007). *Trends in Neural Computation*, chapter Short Term Electric Load Forecasting: A Tutorial, pages 391–418. Springer Berlin Heidelberg, Berlin, Heidelberg.

LaMonica, M. (2013). How to create thermal images for millions of homes. *MIT Technology Review*.

Leach, M., Rockstrm, J., Raskin, P., Scoones, I., Stirling, A. C., Smith, A., Thompson, J., Millstone, E., Ely, A., Arond, E., Folke, C., and Olsson, P. (2012). Transforming Innovation for Sustainability. *Ecology and Society*, 17(2).

Livera, A. M. D., Hyndman, R. J., and Snyder, R. D. (2011). Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527.

Morrison, C. and Shortt, N. (2008). Fuel poverty in Scotland: Refining spatial resolution in the Scottish Fuel Poverty Indicator using a GIS-based multiple risk index. *Health & Place*, 14(4):702 – 717.

Municipal Property Assessment Corporation (2015). Automated Valuation Model. Technical report, Municipal Property Assessment Corporation.

Natural Resources Canada (2015). Hot2xp.

Ndiaye, D. and Gabriel, K. (2011). Principal component analysis of the electricity consumption in residential dwellings. *Energy and Buildings*, 43(23):446 – 453.

Newsham, G. R., Birt, B. J., and Rowlands, I. H. (2011). A comparison of four methods to evaluate the effect of a utility residential air-conditioner load control program on peak electricity use. *Energy Policy*, 39(10):6376 – 6389. Sustainability of biofuels.

Ontario Power Authority (2014). 2013 ltep demand forecast. Technical report, Ontario Power Authority.

Parker, P., Rowlands, I. H., and Scott, D. (2003). Innovations to reduce residential energy use and carbon emissions: an integrated approach. *Canadian Geographer / Le Gographe canadien*, 47(2):169–184.

Rockstrom, J., Steffen, W., Noone, K., Persson, A., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sorlin, S., Snyder, P. K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R. W., Fabry, V. J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., and Foley, J. A. (2009). A safe operating space for humanity. *Nature*, 461(7263):472–475.

Rodríguez Fernández, M., Cortés García, A., González Alonso, I., and Zalama Casanova, E. (2015). Using the big data generated by the smart home to improve energy efficiency management. *Energy Efficiency*, 9(1):249–260.

Rowlands, I. H., Reid, T., and Parker, P. (2015). Research with disaggregated electricity end-use data in households: review and recommendations. *Wiley Interdisciplinary Reviews: Energy and Environment*, 4(5):383–396.

Sanquist, T. F., Orr, H., Shui, B., and Bittner, A. C. (2012). Lifestyle factors in u.s. residential electricity consumption. *Energy Policy*, 42:354 – 364.

Shi, W. (2011). Econometric analysis of household energy consumption in the united states, 2006 and 2008. Master's thesis, Auburn University.

Song, G. (2008). Spatial Analysis of Participation in the Waterloo Residential Energy Efficiency Project. Master's thesis, University of Waterloo.

Stram, D. O. and Fels, M. F. (1986). The applicability of prism to electric heating and cooling. *Energy and Buildings*, 9(1-2):101–110.

Tukey, J. (1977). *Exploratory Data Analysis*. Addison-Wesley series in behavioral science. Addison-Wesley Publishing Company.

Van Rijsbergen, C. (1979). *Information retrieval*. University of Glasgow.

Walker, B., Salt, D., and Reid, W. (2012). In the Loop. In *Resilience Thinking: Sustaining Ecosystems and People in a Changing World*, pages 74–95. Island Press.

Wesoff, E. (2014). Can Bidgely or PlotWatt Compete With Opower in Home Energy Engagement?

Wickson, F., Carew, A. L., and Russell, A. W. (2006). Transdisciplinary research: characteristics, quandaries and quality. *Futures*, 38(9):1046 – 1059.

Wilson, C. and Dowlatabadi, H. (2007). Models of Decision Making and Residential Energy Use. *Annual Review of Environment and Resources*, 32(1):169–203.

Zeifman, M. (2014). Smart meter data analytics: Prediction of enrollment in residential energy efficiency programs. In *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*, pages 413–416.

Zeifman, M. and Roth, K. (2011). Nonintrusive appliance load monitoring: Review and outlook. *Consumer Electronics, IEEE Transactions on*, 57(1):76–84.

Zhao, J. (2013). Is Local Policy Effective in Changing Energy Efficiency Decision Making: Effectiveness of House Retrofitting Programs in Waterloo Region. Master's thesis, University of Waterloo.

Appendices

Appendix A

Thermal Model Derivations

A.1 Thermodynamic modelling

A.1.1 Passive thermal model

Interior heating and cooling loads are fundamentally related to interior temperature: as such, obtaining a better understanding of the dynamics influencing interior temperature is important. A simple model for understanding changes in indoor temperature as they relate to heat flow (i.e. power) is given as:

$$C_{th}\frac{\delta T_{int}(t)}{\delta t} = q_{net}(t)$$

where q_{net} is the overall net heat flow in or out of the building (whether due to heating/cooling systems, outdoor temperatures, solar gains, etc) as a function of time, and C_{th} is the aggregate heat capacity or thermal mass of the building.

At a minimum, $q_{net}(t)$ is influenced passively by ambient environmental conditions, primarily conduction of heat across exterior walls. Conductive heat transfer is modelled as:

$$q_{conductive}(t) = \tilde{U}\big(T_{ext}(t) - T_{int}(t)\big)$$

That is, magnitude and direction of heat flow is proportional to the difference between current indoor and outdoor temperatures with some proportionality constant $\tilde{U}$ (proportional to the surface area of exterior walls and inversely proportional to their thermal resistivity). Other means of passive heat loss and gain can include air exfiltration (drafts) and gains from absorbed solar radiation. For now, however, we represent the passive thermal dynamic case as simply:

$$C_{th}\frac{\delta T_{int}(t)}{\delta t} = q_{net}(t) = q_{passive}(t) = q_{conductive}(t) = \tilde{U}\big(T_{ext}(t) - T_{int}(t)\big)$$

A.1.2 Active constant-interior-temperature model

A dwelling's interior temperature is clearly not just affected by passive thermal heat transfer - if it were, it would be of little interest from the perspective of human energy consumption. Instead, active mechanical heating and cooling systems are usually in place to control indoor temperatures and maintain occupant comfort. The effect of such systems can be quantified as:

$$C_{th}\frac{\delta T_{int}(t)}{\delta t} = q_{net}(t) = q_{passive}(t) + q_{active}(t) = q_{passive}(t) + q_h(t) + q_c(t)$$

where we treat active heating and cooling systems as separate entities. The thermal heat output of the systems can be related to their electrical power load via effective electrical efficiencies e_c, e_h:

$$q_h(t) = e_h P_h(t),\ q_c(t) = -e_c P_c(t)$$

A question that now arises is how best to model the means and ends of active heating and cooling systems. A simple approach is to assume that such systems work to maintain a constant interior temperature at all times, such that $\frac{\delta T_{int}}{\delta t} = 0$ (T_{int} has no time-dependence). In this case our model simplifies to:

$$C_{th}\frac{\delta T_{int}}{\delta t} = 0 = q_{net}(t) = q_{passive}(t) + q_h(t) + q_c(t)$$
$$q_{passive} + q_h + q_c = 0$$

Applying our conductive-only model of passive heat flow, we can expand this as:

$$\tilde{U}\big(T_{ext}(t) - T_{int}\big) + q_h(t) + q_c(t) = 0$$

From this representation it becomes clear that when $T_{ext}(t) = T_{int}$, no active intervention is required to achieve the system's objective. When $T_{ext}(t) < T_{int}$, $q_h(t) > 0$ is required to maintain balance, and conversely, $q_c(t) < 0$ is required when $T_{ext}(t) > T_{int}$. More precisely, the following is necessary:

$$q_h(t) = \begin{cases} -\tilde{U}\big(T_{ext}(t) - T_{int}\big), & T_{ext}(t) < T_{int} \\ 0, & T_{ext}(t) \geq T_{int} \end{cases}$$

$$q_c(t) = \begin{cases} 0, & T_{ext}(t) \leq T_{int} \\ -\tilde{U}\big(T_{ext}(t) - T_{int}\big), & T_{ext}(t) > T_{int} \end{cases}$$

In terms of electrical power, the relations are

$$P_h(t) = \begin{cases} -\frac{\tilde{U}}{e_h}\big(T_{ext}(t) - T_{int}\big), & T_{ext}(t) < T_{int} \\ 0, & T_{ext}(t) \geq T_{int} \end{cases}$$

$$P_c(t) = \begin{cases} 0, & T_{ext}(t) \leq T_{int} \\ \frac{\tilde{U}}{e_c}\big(T_{ext}(t) - T_{int}\big), & T_{ext}(t) > T_{int} \end{cases}$$

These relations relate very closely to the piecewise linear regression approaches taken by most existing residential energy models in the literature, although in most cases the constant T_{int} is assigned different values in the heating and cooling power equations to respresent the fact that HVAC systems generally do not attempt to maintain a single temperature year-round, but instead keep interior temperatures within an acceptable range of values. A more rigorous (thermodyamically-derived) approach to this constraint relaxation is given in Section D.2.1.

A.2 Statistical behavioural modelling

A.2.1 Log-transformed average model

$$\ln(P_{o_t}) = \mu + \epsilon_t,\ \epsilon_t \sim (0, \sigma_\epsilon^2)$$
$$P_{o_t} = e^{\mu + \epsilon_t},\ \epsilon_t \sim (0, \sigma_\epsilon^2)$$

A.2.2 Seasonal-average model

$$P_{o_t} = \mu_{s_t} + \epsilon_t,\ \epsilon_t \sim (0, \sigma_\epsilon^2)$$

A.2.3 Random walk model

$$P_{o_t} = \mu_{w_t} + \epsilon_t,\ \epsilon_t \sim (0, \sigma_\epsilon^2)$$
$$\mu_{w_t} = \mu_{w_{t-1}} + \xi_t,\ \xi_t \sim (0, \sigma_\xi^2)$$

A.2.4 Log-transformed random walk model with seasonal average

$$\ln(P_{o_t}) = \mu_{w_t} + \mu_{s_t} + \epsilon_t,\ \epsilon_t \sim (0, \sigma_\epsilon^2)$$
$$\mu_{w_t} = \mu_{w_{t-1}} + \xi_t,\ \xi_t \sim (0, \sigma_\xi^2)$$

A.2.5 Linearized (autoregression with seasonal average, DAR) model

$$P_{o_t} = \phi_1 P_{o_{t-1}} + \mu_{s_t} + \epsilon_t,\ \epsilon_t \sim (0, \sigma_\epsilon^2)$$

Appendix B

Model Implementation Julia Code

The following code blocks provide the Julia code used to implement the various models assessed in this analysis. All assume the existence in scope of `readings` and `temperature` data structures of type `TimeArray`, which contain relevant whole-house electrical load and exterior temperature observations.

B.1 Three-Regime Thermal Regression Model (Fit3)

Implements Birt et al. (2012):

```
using TimeSeries
using DataFrames
using MultivariateStats

immutable ThreeLinesModelSingleLine
  g_h::Float64
  T_ch::Float64
  P_ch::Float64
  T_cc::Float64
  P_cc::Float64
  g_c::Float64

  function ThreeLinesModelSingleLine(g_h, T_ch, P_ch, T_cc, P_cc, g_c)
    @assert T_ch < T_cc
    new(g_h, T_ch, P_ch, T_cc, P_cc, g_c)
  end #ThreeLinesModelSingleLine

end #ThreeLinesModelSingleLine


immutable ThreeLinesModel

    average::ThreeLinesModelSingleLine
```

```
    low::ThreeLinesModelSingleLine
    median::ThreeLinesModelSingleLine
    high::ThreeLinesModelSingleLine

    function ThreeLinesModel(average::ThreeLinesModelSingleLine, low::ThreeLinesModelSingleLine,
                median::ThreeLinesModelSingleLine, high::ThreeLinesModelSingleLine)
      new(average, low, median, high)
    end #ThreeLinesModel

end #ThreeLinesModel

function threelines_prep_data{S,T}(P_sm::TimeArray{S,1}, T_ext::TimeArray{T,1})

  data = merge(T_ext, P_sm)
  data.colnames[:] = ["T_ext", "P_sm"]
  data = data[find(!any(isnan(data.values), 2))]
  T_ext = data["T_ext"].values
  P_sm = data["P_sm"].values

  # Bin data to nearest degree celsius
  obs = DataFrame(T_ext=T_ext, P_sm=P_sm)
  obs[:T_ext] = round(obs[:T_ext])

  # Calculate low, median, and high values
  bins = DataFrames.by(obs, :T_ext) do df
    DataFrame(
      count=size(df,1),
      low=percentile(df[:P_sm],10),
      median=median(df[:P_sm]),
      high=percentile(df[:P_sm], 90)
    )
  end #do

  # Eliminate bins with < 20 data points
  bins[bins[:count] .>= 20, [:T_ext,:low,:median,:high]]

  return T_ext, P_sm, collect(bins[:T_ext]),
          collect(bins[:low]), collect(bins[:median]), collect(bins[:high])

end #function

function fit_threelines_singleline(T_ext::Vector, P_sm::Vector; min_T_ch = 10, min_regime_size=5)

  to_matrix{T<:Any}(v::AbstractArray{T,1}) = reshape(collect(v),(size(v,1),1))

  gridrange(eps::Float64) = -4eps:eps:4eps

  predict{T<:Float64}(a::Tuple{T,T}, b::Tuple{T,T}, values::AbstractArray{T}) =
    (b[2]-a[2])/(b[1]-a[1]).*(values .- a[1]) .+ a[2]

  min_srmse = Inf
  best_breakpoints = fill(NaN, 6)
```

```
T_ext_max = maximum(T_ext)

for T_ch in min_T_ch:T_ext_max-2min_regime_size
  for T_cc in T_ch+min_regime_size:T_ext_max-min_regime_size

    heating_regime = T_ext .<= T_ch
    passive_regime = T_ch .<= T_ext .<= T_cc
    cooling_regime = T_cc .<= T_ext

    regimes = (heating_regime, passive_regime, cooling_regime)

    m = fill(NaN, 3)
    b = fill(NaN, 3)
    srmse = 0

    for r in 1:size(regimes,1)
      regime_T = T_ext[regimes[r]]
      regime_P = P_sm[regimes[r]]
      fit = llsq(to_matrix(regime_T), to_matrix(regime_P))
      m[r] = fit[1]
      b[r] = fit[2]
      srmse += rmse(regime_P .- (regime_T*m[r] .+ b[r]))
    end #for

    if srmse < min_srmse
      min_srmse = srmse
      best_breakpoints = [T_ch, m[1]*T_ch + b[1], m[2]*T_ch + b[2],
                          T_cc, m[2]*T_cc + b[2], m[3]*T_cc + b[3]]
    end #if

  end #for
end #for

# 4 - Generate 9x9 grids around optimal T_ch, T_cc and pick best breakpoints to minimize SRMSE
final_parameters = fill(NaN, 6)
min_srmse = Inf

T_chs = best_breakpoints[1] + gridrange(.5)
P_chs = mean(best_breakpoints[2:3]) +
  gridrange((best_breakpoints[3]-best_breakpoints[2])/4)
T_ccs = best_breakpoints[4] + gridrange(.5)
P_ccs = mean(best_breakpoints[5:6]) +
  gridrange((best_breakpoints[6]-best_breakpoints[5])/4)

for T_ch in T_chs

  heating_T = T_ext[T_ext .<= T_ch] - T_ch

  for P_ch in P_chs

    heating_P = P_sm[T_ext .<= T_ch] - P_ch
```

```
    heating_fit = llsq(to_matrix(heating_T), to_matrix(heating_P), bias=false)
    residuals_heating = heating_P - heating_fit[1]*heating_T

    for T_cc in T_ccs

      cooling_T = T_ext[T_cc .<= T_ext] - T_cc

      for P_cc in P_ccs

        cooling_P = P_sm[T_cc .<= T_ext] - P_cc
        cooling_fit = llsq(to_matrix(cooling_T), to_matrix(cooling_P), bias=false)
        residuals_cooling = cooling_P - cooling_fit[1]*cooling_T

        passive_T = T_ext[T_ch .<= T_ext .<= T_cc]
        passive_P = P_sm[T_ch .<= T_ext .<= T_cc]
        residuals_passive = passive_P - predict((T_ch,P_ch),(T_cc,P_cc), passive_T)

        srmse = rmse(residuals_heating) + rmse(residuals_passive) + rmse(residuals_cooling)

        if srmse < min_srmse
            min_srmse = srmse
            final_parameters = [heating_fit[1], T_ch, P_ch, T_cc, P_cc, cooling_fit[1]]
        end #if

      end #for
    end #for
  end #for
end #for

return ThreeLinesModelSingleLine(final_parameters...)

end #fit_threelines_singleline

function fit_threelines{S,T}(P_sm::TimeArray{S,1}, T_ext::TimeArray{T,1};
            min_T_ch = 10, min_regime_size=5)

  T_ext, P_sm, T_binned, low_vals, med_vals, high_vals = threelines_prep_data(P_sm, T_ext)

  averagefit = fit_threelines_singleline(T_ext, P_sm,
                           min_T_ch=min_T_ch, min_regime_size=min_regime_size)
  lowfit = fit_threelines_singleline(T_binned, low_vals,
                           min_T_ch=min_T_ch, min_regime_size=min_regime_size)
  medianfit = fit_threelines_singleline(T_binned, med_vals,
                           min_T_ch=min_T_ch, min_regime_size=min_regime_size)
  highfit = fit_threelines_singleline(T_binned, high_vals,
                           min_T_ch=min_T_ch, min_regime_size=min_regime_size)

  return ThreeLinesModel(averagefit, lowfit, medianfit, highfit)

end #fit_threelines
```

```
function disaggregate{S,T}(P_sm::TimeArray{S,1}, m::ThreeLinesModel, T_ext::TimeArray{T,1})

  g_h, T_ch, T_cc, g_c = min(0, m.average.g_h), m.average.T_ch,
                         m.average.T_cc, max(0, m.average.g_c)

  is_heating = T_ext .<  T_ch
  is_cooling = T_ext .> T_cc

  P_h = is_heating*g_h .* (T_ext.-T_ch)
  P_c = is_cooling*g_c .* (T_ext.-T_cc)
  P_r = 0P_sm + (m.average.P_ch + m.average.P_cc)/2
  # P_r = P_sm .- P_h .- P_c
  # P_r = (P_r .> 0) .* 1. .* P_r

  estimates = merge(P_h, merge(P_c, P_r))
  estimates.colnames[:] = ["P_h", "P_c", "P_r"]

  return estimates

end #disaggregate ThreeLinesModel

function assess{S,T}(m::ThreeLinesModel, P_sm::TimeArray{S,1},
                              T_ext::TimeArray{T,1}, groundtruth::TimeArray)
  data = data[find(!any(isnan(merge(P_sm, merge(T_ext, groundtruth)).values), 2))]

  estimates  = disaggregate(P_sm, m, T_ext)

  return assess(estimates, groundtruth)
end #assess_threelines

function plot_threelines(fp::Matrix, data::TimeArray)
  _, _, T_binned, low_vals, med_vals, high_vals = threelines_prep_data(data)


  function pale(cname::String, op::Float64=0.3)
    col = parse(Colorant, cname)
    RGBA(red(col), green(col), blue(col), op)
  end #pale

  heating_line(i::Int, col::String) =
    layer(x -> fp[1,i]*x + (fp[3,i] - fp[1,i]*fp[2,i]), -25, fp[2,i],
    Theme(default_color=parse(Colorant, col)))
  passive_line(i::Int, col::String) =
    layer(x -> fp[3,i] + (fp[5,i]-fp[3,i])/(fp[4,i]-fp[2,i])*(x-fp[2,i]), fp[2,i], fp[4,i],
    Theme(default_color=parse(Colorant, col)))
  cooling_line(i::Int, col::String) =
    layer(x -> fp[6,i]*x + (fp[5,i] - fp[6,i]*fp[4,i]), fp[4,i], 35,
    Theme(default_color=parse(Colorant, col)))

  plot(heating_line(1, "black"),passive_line(1, "black"),cooling_line(1, "black"),
    layer(x=T_binned, y=low_vals, Geom.point, Theme(default_color=pale("blue"))),
    heating_line(2, "blue"), passive_line(2, "blue"), cooling_line(2, "blue"),
```

```
layer(x=T_binned, y=med_vals, Geom.point, Theme(default_color=pale("orange"))),
heating_line(3, "orange"), passive_line(3, "orange"), cooling_line(3, "orange"),
layer(x=T_binned, y=high_vals, Geom.point, Theme(default_color=pale("red"))),
heating_line(4, "red"), passive_line(4, "red"), cooling_line(4, "red"))

threelinesfit = fit_threelines(readings, temperatures)
estimates = disaggregate(readings, threelinesfit, temperatures)
estimates.colnames[:] = ["Heating", "Cooling", "Other"]

physicalparams = [threelinesfit.high.g_h, threelinesfit.high.g_c]

baseload = min(threelinesfit.low.P_ch, threelinesfit.low.P_cc)
activity = min(threelinesfit.high.P_ch, threelinesfit.high.P_cc) - baseload
behaviouralparams = [threelinesfit.high.T_ch, threelinesfit.high.T_cc, baseload, activity]
```

B.2 Periodic Autoregressive Fixed Four-Regime Model (PARX-Fixed4)

Implements Ardakanian et al. (2014):

```
using TimeZones
using TimeSeries
using MultivariateStats

# Generate exogenous factors
m = 5

T_h1 = (Tch1 - temperatures) .* (temperatures .< Tch1)
T_h1.colnames[:] = ["T_h1"]

T_h2 = (Tch2 - temperatures) .* (temperatures .< Tch2)
T_h2.colnames[:] = ["T_h2"]

T_c  = (temperatures - Tcc) .* (temperatures .> Tcc)
T_c.colnames[:] = ["T_c"]

OL = readings .< quantile(readings.values[!isnan(readings.values)], 0.1)
OL.colnames[:] = ["OL"]

OH = readings .> quantile(readings.values[!isnan(readings.values)], 0.9)
OH.colnames[:] = ["OH"]

regressors = merge(merge(merge(merge(T_h1, T_h2), T_c), 1.0*OL), 1.0*OH)
regressors.colnames[:] = ["T_h1", "T_h2", "T_c", "OL", "OH"]

# Split data into periods

readingseasons = TimeArray(readings.timestamp,
                           map(season_hourofdaytype,
                           readings.timestamp), ["Season"])
```

```
S = length(unique(readingseasons.values))
readingsperiods = map(s -> readings[readingseasons .== s], 1:S)
regressorsperiods = map(s -> regressors[readingseasons .== s], 1:S)

# Add in lag-3 AR terms
p = 3
readingsperiods = map(ta -> merge(merge(merge(ta, lag(ta, 1, padding=true)),
                                        lag(ta, 2, padding=true)), lag(ta, 3, padding=true),
                                        colnames=["P_t", "P_t-1",  "P_t-2", "P_t-3"]),
                                        readingsperiods)

dataperiods = map((readings,regressors) -> merge(readings, regressors),
                  readingsperiods,
                  regressorsperiods)

# Fit parameters via OLS
cleandataperiods = map(dropnan, dataperiods)
params = Array{Float64}(m+p+1, S)
cleandataperiods = map(ta -> dropnan(ta, :any), dataperiods)
for s in eachindex(cleandataperiods)
    timestamp = cleandataperiods[s].timestamp
    y, X = cleandataperiods[s].values[:,1], cleandataperiods[s].values[:,2:end]
    nodata = sum(abs(X), 1) .== 0
    nodata_idx, data_idx = find(nodata), find(!nodata)
    params[nodata_idx, s] = 0
    params[[data_idx; 9], s] = llsq(X[:, data_idx], y)
end #for

physicalparams = vec(params[5:7, :])
behaviouralparams = vec(params[[1:4;8:9], :])

# Disaggregate readings
disaggs = TimeArray(ZonedDateTime[], Array{Float64}(0,3), ["Heating", "Cooling", "Other"])
for s in eachindex(dataperiods)
    timestamp = dataperiods[s].timestamp
    heating = max(0, params[4, s]) * dataperiods[s]["T_h1"] .+
              max(0, params[5, s]) * dataperiods[s]["T_h2"]
    cooling = max(0, params[6, s]) * dataperiods[s]["T_c"]
    other = TimeArray(timestamp, dataperiods[s].values[:, 2:end] * params[1:end-1, s] +
                params[end, s], ["Other"])
    other = other .* (other .> 0)
    disaggs = vcat(disaggs, merge(merge(heating, cooling), other,
                                  colnames=["Heating", "Cooling", "Other"]))
end #for
```

B.3 Deseasonalized Autoregressive Fixed Four-Regime Model (DAR-Fixed4)

```
using TimeZones
```

```
using TimeSeries
using MultivariateStats

readings = dropnan(merge(readings, lag(readings, padding=true),
                         colnames=["P_sm", "P_sm_t-1"]), :any)
temperatures = temperatures[readings.timestamp]
Tch1, Tch2, Tcc = 16, 5, 20

T_h1 = (Tch1 - temperatures) .* (temperatures .< Tch1)
T_h1.colnames[:] = ["T_h1"]

T_h2 = (Tch2 - temperatures) .* (temperatures .< Tch2)
T_h2.colnames[:] = ["T_h2"]

T_c  = (temperatures - Tcc) .* (temperatures .> Tcc)
T_c.colnames[:] = ["T_c"]

seasonalfactors = hoursofdaytype(readings)

regressors = [T_h1.values T_h2.values T_c.values
                readings["P_sm_t-1"].values seasonalfactors.values]

params = llsq(regressors, readings["P_sm"].values, bias=false)

physicalparams = params[1:3]
behaviouralparams = params[4:end]

heating = max(0, physicalparams[1]) * T_h1 .+ max(0, physicalparams[2]) * T_h2
cooling = max(0, physicalparams[3]) * T_c
# other = readings["P_sm"] .- heating .- cooling
other = TimeArray(readings.timestamp,
                        behaviouralparams[end] +
                        seasonalfactors.values * behaviouralparams[1:end-1], ["Other"])
other = other .* (other .> 0)
disaggs = merge(merge(heating, cooling), other, colnames=["Heating", "Cooling", "Other"])
```

B.4 Deseasonalized Autoregressive Two-Regime Model (DAR-Fit2)

```
using TimeZones
using TimeSeries
using MultivariateStats

function fit_thresholdregression{T}(y::Vector{T}, z::Vector{T},
                                   ::Vector{T}, X::Matrix=Array{T}(length(y), 0);
bias::Bool=true, niter::Int=100, tol::Float64=1e-9, verbose::Bool=false)

@assert length(y) == length(z)
@assert length(y) == size(X, 1)
```

```
    K = length()
    done, timeout = false, false
    , , , , _prev = [], [], [], [], Inf*ones()
    i = 0

    while !done

        i += 1
        verbose && println("\nIteration $i:")
        verbose && println(" = $")

        V = (z .> ')
        U = (z .- ') .* V
        V *= -1

        rank([z U V X]) < size([z U V X], 2) && warn("Rank deficient...")
        params = llsq([z U V X], y, bias=bias)
        ,  = params[1], params[2:K+1]
        ,  = params[K+2:2K+1], params[2K+2:end]

         =  ./  .+

        verbose && println("Max  = $(maximum(abs()))")
        converged = all(abs() .< tol)
        timeout = i >= niter || any(abs() .> abs(_prev))

        done = converged || timeout
        _prev =

    end #while

    timeout && warn("Fitting timed out after $i iterations: may have failed to converge.
    Convergence threshold was $tol; largest absolute value at final iteration
    was $(maximum(abs()))")

    return , , ,

end #fit_thresholdregression



    readings = dropnan(merge(readings, lag(readings, padding=true),
                             colnames=["P_sm", "P_sm_t-1"]), :any)
    temperatures = temperatures[readings.timestamp]
    seasons = hoursofdaytype(readings)

    partial_regressors = [readings["P_sm_t-1"].values seasons.values]
    y = readings["P_sm"].values

    , , ,  = fit_thresholdregression(y, temperatures.values, [20.],
                                          partial_regressors, bias=false)
```

```
Ts, g_h, g_c = [1], ,  + [1]

T_h = (Ts .- temperatures) .* (temperatures .< Ts)
T_c = (temperatures .- Ts) .* (temperatures .> Ts)

physicalparameters = [g_h, g_c]
behaviouralparameters = [Ts; ]
heating = T_h .* max(0, -g_h)
cooling = T_c .* max(0, g_c)
# other = readings["P_sm"] .- heating .- cooling
other = TimeArray(readings.timestamp, partial_regressors * , ["Other"])
other = other .* (other .> 0)
disaggs = merge(merge(heating, cooling), other, colnames=["Heating", "Cooling", "Other"])
```

B.5 Deseasonalized Autoregressive Three-Regime Model (DAR-Fit3)

```
using TimeZones
using TimeSeries
using MultivariateStats
using Optim

readings = dropnan(merge(readings, lag(readings, padding=true),
                         colnames=["P_sm", "P_sm_t-1"]), :any)
temperatures = temperatures[readings.timestamp]
seasons = hoursofdaytype(readings)

partial_regressors = [readings["P_sm_t-1"].values seasons.values]
y = readings["P_sm"].values

function rmse(p::Vector{Float64})
    Tch, Tcc = p[1], p[2]
    T_h = (Tch .- temperatures) .* (temperatures .< Tch)
    T_c = (temperatures .- Tcc) .* (temperatures .> Tcc)
    regressors = [T_h.values T_c.values partial_regressors]
    params = llsq(regressors, y, bias=false)
    return sqrt(mean(abs2(y - regressors * params)))
end #rmse

p = optimize(rmse, [Tch, Tcc]).minimum

Tch, Tcc = p[1], p[2]
T_h = (Tch .- temperatures) .* (temperatures .< Tch)
T_c = (temperatures .- Tcc) .* (temperatures .> Tcc)
regressors = [T_h.values T_c.values partial_regressors]
params = llsq(regressors, y, bias=false)

physicalparameters = params[1:2]
```

```
behaviouralparameters = [Tch; Tcc; params[3:end]]
heating = T_h .* max(0, params[1])
cooling = T_c .* max(0, params[2])
# other = readings["P_sm"] .- heating .- cooling
other = TimeArray(readings.timestamp, partial_regressors * params[3:end], ["Other"])
other = other .* (other .> 0)
disaggs = merge(merge(heating, cooling), other, colnames=["Heating", "Cooling", "Other"])
```

Appendix C

Model Comparison and Validation Results

C.1 Three-Regime Thermal Regression Model (Fit3)

ID	Heating Activity Accuracy	Heating Activity Precision	Heating Activity Recall	Heating Activity F1	Heating Load RMSE (overall)	Heating Load RMSE (active)	Heating Load MAPE
A	0.64	0.56	1.00	0.72	0.23	0.26	4.12
B	0.88	0.96	0.87	0.91	0.20	0.23	0.65
C	0.77	0.66	0.96	0.78	0.13	0.16	1.27
D	0.76	0.91	0.75	0.83	0.32	0.35	1.15
E	0.68	0.86	0.72	0.79	0.18	0.19	1.00
F	0.75	0.67	0.96	0.79	0.19	0.22	1.07
G	0.81	0.75	0.94	0.84	0.17	0.21	0.87
H	0.81	0.95	0.79	0.86	0.37	0.41	0.89
I	0.69	0.51	0.98	0.67	0.13	0.17	1.18
J	0.66	0.99	0.59	0.74	0.14	0.16	0.84
K	0.88	1.00	0.86	0.93	0.12	0.13	0.85
L	0.70	0.61	0.84	0.71	0.08	0.10	1.18
M	0.13	NA	0.00	NA	0.09	0.09	1.00
N	0.67	1.00	0.60	0.75	0.21	0.23	1.41
O	0.69	0.37	0.95	0.53	0.11	0.16	2.80
P	0.55	1.00	0.51	0.67	0.28	0.30	0.97
Q	0.64	1.00	0.59	0.74	0.10	0.11	1.24
R	0.36	0.09	0.62	0.15	0.09	0.11	1.41
S	0.60	0.35	0.82	0.49	0.14	0.18	2.18
T	0.34	0.00	0.13	0.01	0.07	0.09	3.75
U	0.67	0.03	0.52	0.06	0.06	0.10	0.99

Table C.1: Heating characterization assessment metrics for each household under the Fit3 model

ID	Cooling Activity Accuracy	Cooling Activity Precision	Cooling Activity Recall	Cooling Activity F1	Cooling Load RMSE (overall)	Cooling Load RMSE (active)	Cooling Load MAPE	Behavioural Load RMSE	Behavioural Load MAPE
A	0.88	0.29	0.95	0.45	0.17	0.41	1.15	0.64	1.00
B	0.74	0.99	0.02	0.04	0.43	0.84	1.00	1.03	2.06
C	0.78	0.34	0.98	0.50	0.32	0.56	0.92	0.64	0.48
D	0.82	0.45	0.96	0.61	0.38	0.67	0.70	0.48	1.46
E	0.83	0.30	0.98	0.46	0.16	0.31	0.64	0.52	0.83
F	0.85	0.48	0.95	0.64	0.24	0.46	1.83	0.47	1.25
G	0.78	0.31	0.94	0.47	0.22	0.39	1.93	0.47	3.24
H	0.84	0.59	0.86	0.70	0.45	0.75	0.80	0.87	0.50
I	0.78	0.38	0.99	0.55	0.30	0.51	1.69	0.37	2.13
J	0.92	0.95	0.06	0.11	0.22	0.74	0.99	0.81	1.15
K	0.91	0.55	0.98	0.70	0.24	0.55	0.51	0.60	0.93
L	0.72	0.11	0.99	0.20	0.17	0.31	0.92	0.45	0.74
M	0.70	0.03	1.00	0.07	0.14	0.25	0.79	0.21	0.74
N	0.76	0.40	0.97	0.57	0.68	1.07	1.02	0.33	0.47
O	0.76	0.43	0.91	0.58	0.28	0.44	0.89	0.54	1.40
P	0.62	0.19	0.99	0.32	0.60	0.88	0.67	0.56	21.71
Q	0.74	0.31	0.97	0.47	0.59	0.98	1.21	0.59	1.26
R	0.32	0.91	0.27	0.42	0.34	0.36	7.19	0.38	1.36
S	0.52	0.94	0.39	0.56	0.59	0.67	7.89	0.63	2.07
T	0.31	0.96	0.32	0.48	0.14	0.14	2.63	0.27	0.69
U	0.61	0.98	0.61	0.75	0.48	0.48	0.74	0.47	2.88

Table C.2: Cooling and behavioural characterization assessment metrics for each household under the Fit3 model

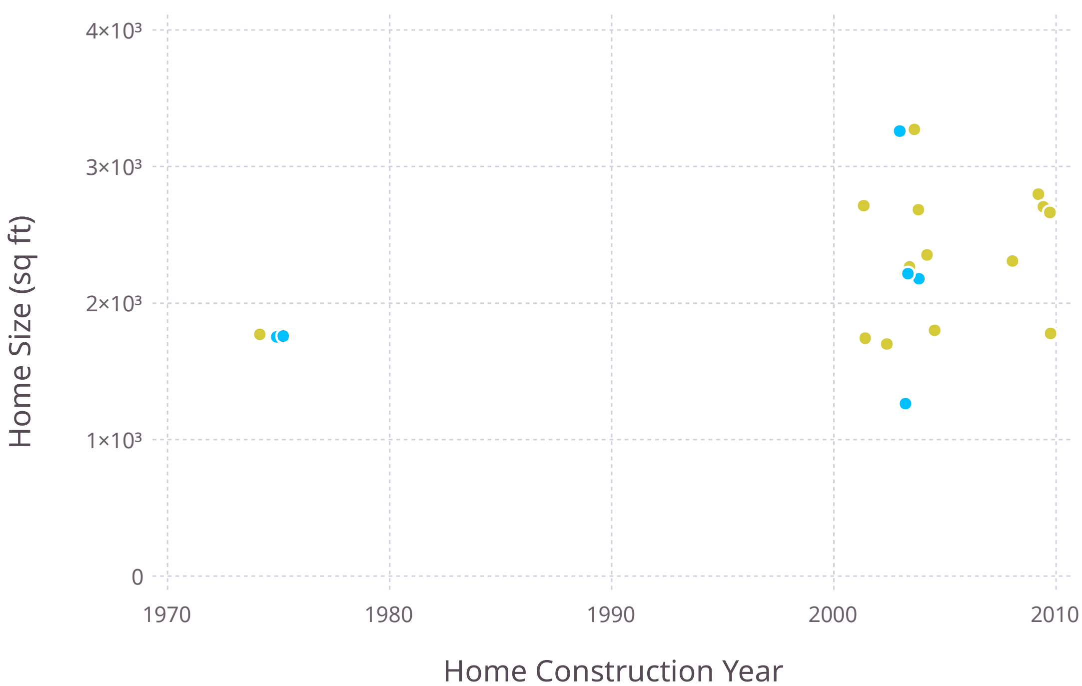

Figure C.1: Best clustering assignment of EHMS homes using physical parameters from the Fit3 model

C.2 Periodic Autoregressive Fixed Four-Regime Model (PARX-Fixed4)

ID	Heating Activity Accuracy	Heating Activity Precision	Heating Activity Recall	Heating Activity F1	Heating Load RMSE (overall)	Heating Load RMSE (active)	Heating Load MAPE
A	0.67	0.58	0.93	0.71	0.12	0.13	1.32
B	0.79	0.97	0.74	0.84	0.24	0.28	0.80
C	0.70	0.59	0.97	0.73	0.15	0.17	1.42
D	0.73	0.91	0.72	0.80	0.36	0.40	0.96
E	0.72	0.86	0.78	0.82	0.19	0.20	1.03
F	0.73	0.67	0.86	0.76	0.20	0.24	1.13
G	0.75	0.69	0.94	0.80	0.20	0.23	0.87
H	0.78	0.93	0.77	0.84	0.37	0.41	0.97
I	0.72	0.54	0.87	0.67	0.13	0.17	1.30
J	0.59	1.00	0.49	0.66	0.15	0.16	0.93
K	0.84	1.00	0.82	0.90	0.17	0.18	0.89
L	0.67	0.59	0.84	0.69	0.09	0.11	1.19
M	0.56	0.83	0.63	0.71	0.07	0.07	1.01
N	0.74	0.99	0.68	0.81	0.21	0.23	1.90
O	0.67	0.35	0.90	0.50	0.13	0.18	3.37
P	0.42	1.00	0.36	0.53	0.27	0.29	1.96
Q	0.66	1.00	0.61	0.76	0.12	0.13	1.37
R	0.30	0.10	0.79	0.18	0.12	0.14	1.97
S	0.56	0.32	0.82	0.46	0.18	0.22	3.14
T	0.37	0.00	0.13	0.01	0.08	0.10	3.13
U	0.62	0.02	0.38	0.04	0.09	0.15	0.98

Table C.3: Heating characterization assessment metrics for each household under the PARX-Fixed4 model

ID	Cooling Activity Accuracy	Cooling Activity Precision	Cooling Activity Recall	Cooling Activity F1	Cooling Load RMSE (overall)	Cooling Load RMSE (active)	Cooling Load MAPE	Behavioural Load RMSE	Behavioural Load MAPE
A	0.92	0.38	0.82	0.52	0.19	0.53	0.96	0.44	0.88
B	0.85	0.88	0.51	0.65	0.40	0.74	0.92	0.65	0.89
C	0.91	0.57	0.72	0.64	0.39	0.94	0.90	0.61	0.54
D	0.90	0.67	0.66	0.67	0.51	1.15	0.91	0.71	2.05
E	0.91	0.43	0.68	0.53	0.17	0.46	0.86	0.41	0.66
F	0.91	0.73	0.55	0.63	0.27	0.67	0.89	0.48	1.25
G	0.89	0.48	0.59	0.53	0.25	0.62	0.96	0.44	3.73
H	0.85	0.73	0.50	0.59	0.50	0.97	0.93	0.86	0.45
I	0.91	0.65	0.79	0.71	0.40	0.91	0.93	0.51	2.40
J	0.88	0.41	0.80	0.55	0.20	0.45	0.85	0.45	0.36
K	0.94	0.80	0.62	0.69	0.38	1.08	0.88	0.59	0.79
L	0.88	0.18	0.68	0.29	0.17	0.46	0.90	0.35	0.53
M	0.87	0.06	0.76	0.11	0.13	0.35	0.91	0.19	0.56
N	0.88	0.63	0.72	0.67	0.85	1.76	0.87	0.96	1.27
O	0.80	0.47	0.41	0.44	0.30	0.58	0.96	0.45	0.91
P	0.82	0.29	0.69	0.41	0.66	1.36	0.93	0.77	20.10
Q	0.87	0.46	0.56	0.51	0.67	1.50	0.96	0.79	1.42
R	0.20	0.98	0.12	0.21	0.39	0.41	1.47	0.49	1.77
S	0.36	0.98	0.17	0.29	0.73	0.83	1.44	0.80	2.07
T	0.12	0.92	0.12	0.22	0.15	0.15	1.07	0.25	0.67
U	0.26	0.99	0.25	0.40	0.56	0.57	0.96	0.55	2.83

Table C.4: Cooling and behavioural characterization assessment metrics for each household under the PARX-Fixed4 model

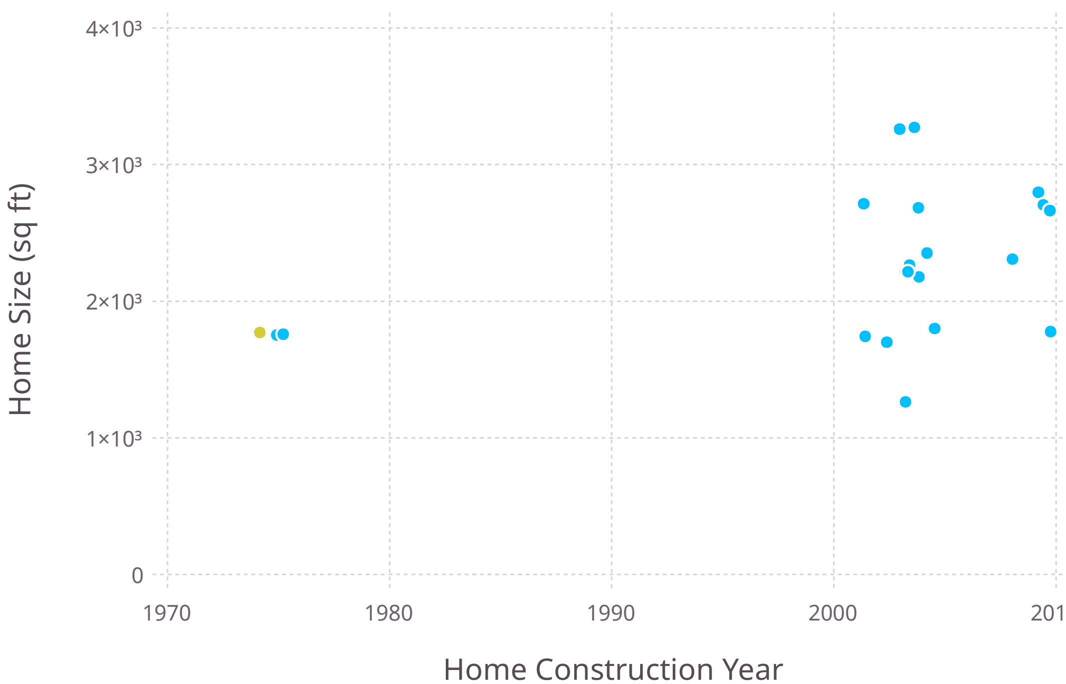

Figure C.2: Best clustering assignment of EHMS homes using physical parameters from the PARX-Fixed4 model

C.3 Deseasonalized Autoregressive Fixed Four-Regime Model (DAR-Fixed4)

ID	Heating Activity Accuracy	Heating Activity Precision	Heating Activity Recall	Heating Activity F1	Heating Load RMSE (overall)	Heating Load RMSE (active)	Heating Load MAPE
A	0.64	0.56	1.00	0.72	0.08	0.09	0.94
B	0.89	0.93	0.92	0.93	0.27	0.31	0.96
C	0.67	0.56	0.98	0.72	0.13	0.15	1.05
D	0.81	0.89	0.86	0.87	0.38	0.41	0.92
E	0.74	0.86	0.81	0.83	0.20	0.21	0.93
F	0.67	0.60	0.99	0.75	0.22	0.25	0.95
G	0.74	0.67	0.98	0.80	0.22	0.25	0.85
H	0.84	0.92	0.87	0.89	0.40	0.44	0.89
I	0.80	0.63	0.88	0.74	0.14	0.20	1.09
J	0.59	1.00	0.49	0.66	0.13	0.15	0.81
K	0.94	0.99	0.94	0.97	0.19	0.20	0.81
L	0.62	0.54	0.92	0.68	0.10	0.11	0.80
M	0.64	0.84	0.72	0.78	0.08	0.08	0.83
N	0.60	1.00	0.51	0.67	0.22	0.24	1.10
O	0.76	0.42	0.89	0.58	0.11	0.17	1.78
P	0.35	1.00	0.28	0.44	0.28	0.30	1.13
Q	0.54	1.00	0.48	0.65	0.12	0.13	1.42
R	0.40	0.05	0.28	0.08	0.07	0.08	1.07
S	0.63	0.36	0.71	0.47	0.15	0.21	1.99
T	0.52	0.00	0.11	0.01	0.05	0.08	2.11
U	0.87	0.03	0.16	0.05	0.06	0.17	0.99

Table C.5: Heating characterization assessment metrics for each household under the DAR-Fixed4 model

ID	Cooling Activity Accuracy	Cooling Activity Precision	Cooling Activity Recall	Cooling Activity F1	Cooling Load RMSE (overall)	Cooling Load RMSE (active)	Cooling Load MAPE	Behavioural Load RMSE	Behavioural Load MAPE
A	0.92	0.37	0.84	0.51	0.19	0.53	0.98	0.72	0.58
B	0.86	0.89	0.55	0.68	0.40	0.76	0.94	0.60	0.73
C	0.91	0.57	0.80	0.67	0.37	0.88	0.85	0.66	0.35
D	0.90	0.66	0.67	0.66	0.49	1.11	0.88	0.43	0.79
E	0.90	0.42	0.80	0.55	0.18	0.45	0.90	0.53	0.40
F	0.91	0.74	0.58	0.65	0.26	0.63	0.90	0.45	0.70
G	0.90	0.51	0.68	0.58	0.24	0.59	1.13	0.45	3.54
H	0.85	0.73	0.54	0.62	0.52	1.00	0.96	0.73	0.27
I	0.91	0.65	0.81	0.72	0.39	0.86	0.90	0.33	1.78
J	0.88	0.42	0.87	0.57	0.19	0.43	0.83	0.74	0.48
K	0.95	0.80	0.72	0.76	0.38	1.07	0.89	0.62	0.55
L	0.88	0.20	0.81	0.32	0.17	0.45	0.90	0.42	0.87
M	0.85	0.06	0.90	0.12	0.13	0.33	0.92	0.21	0.72
N	0.89	0.63	0.73	0.68	0.84	1.74	0.86	0.33	0.32
O	0.80	0.47	0.45	0.46	0.32	0.60	0.99	0.53	0.75
P	0.81	0.29	0.74	0.42	0.65	1.29	0.89	0.60	24.06
Q	0.85	0.42	0.59	0.49	0.67	1.44	1.00	0.99	3.48
R	0.21	0.98	0.13	0.23	0.36	0.38	2.45	0.52	2.97
S	0.36	0.98	0.17	0.29	0.71	0.81	2.17	0.67	2.54
T	0.13	0.92	0.13	0.23	0.15	0.15	1.14	0.29	0.98
U	0.29	0.99	0.28	0.43	0.54	0.54	0.94	0.66	4.24

Table C.6: Cooling and behavioural characterization assessment metrics for each household under the DAR-Fixed4 model

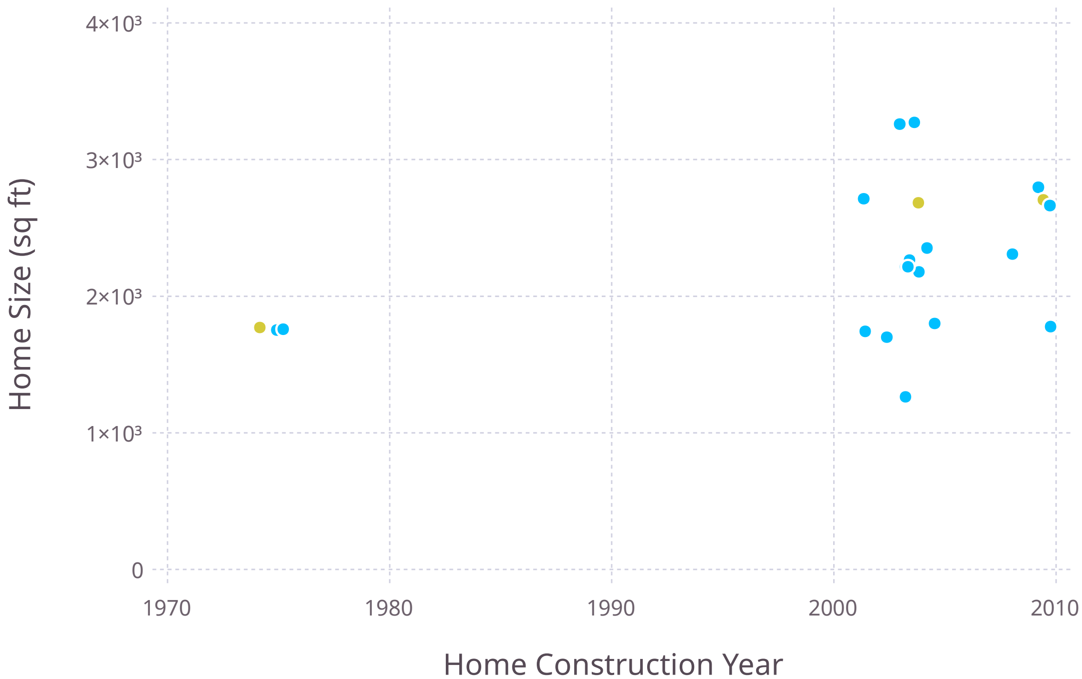

Figure C.3: Best clustering assignment of EHMS homes using physical parameters from the DAR-Fixed4 model

C.4 Deseasonalized Autoregressive Two-Regime Model (DAR-Fit2)

ID	Heating Activity Accuracy	Heating Activity Precision	Heating Activity Recall	Heating Activity F1	Heating Load RMSE (overall)	Heating Load RMSE (active)	Heating Load MAPE
A	0.63	0.55	1.00	0.71	0.08	0.09	1.06
B	0.89	0.95	0.89	0.92	0.25	0.29	0.89
C	0.61	0.52	0.99	0.68	0.14	0.15	1.04
D	0.82	0.88	0.88	0.88	0.36	0.39	0.92
E	0.80	0.85	0.91	0.88	0.21	0.21	0.87
F	0.62	0.57	1.00	0.72	0.21	0.22	0.99
G	0.72	0.66	0.98	0.79	0.20	0.22	0.84
H	0.84	0.87	0.93	0.90	0.41	0.44	0.89
I	0.58	0.44	0.99	0.61	0.15	0.17	1.04
J	0.68	0.98	0.61	0.75	0.15	0.17	0.93
K	0.94	0.99	0.93	0.96	0.16	0.17	0.66
L	0.64	0.55	0.91	0.69	0.09	0.10	0.80
M	0.13	NA	0.00	NA	0.09	0.09	1.00
N	0.87	0.96	0.89	0.92	0.23	0.25	1.04
O	0.65	0.34	0.97	0.50	0.12	0.16	1.31
P	0.09	NA	0.00	NA	0.29	0.31	1.00
Q	0.71	1.00	0.67	0.80	0.12	0.13	1.02
R	0.24	0.11	0.97	0.19	0.09	0.09	2.57
S	0.48	0.30	0.94	0.46	0.14	0.16	1.82
T	0.22	0.00	0.16	0.01	0.06	0.06	2.91
U	0.98	NA	0.00	NA	0.06	0.41	1.00

Table C.7: Heating characterization assessment metrics for each household under the DAR-Fit2 model

ID	Cooling Activity Accuracy	Cooling Activity Precision	Cooling Activity Recall	Cooling Activity F1	Cooling Load RMSE (overall)	Cooling Load RMSE (active)	Cooling Load MAPE	Behavioural Load RMSE	Behavioural Load MAPE
A	0.86	0.27	0.97	0.42	0.18	0.43	1.00	0.45	0.88
B	0.89	0.74	0.88	0.80	0.39	0.66	0.87	0.64	0.94
C	0.89	0.50	0.88	0.64	0.37	0.80	0.83	0.58	0.58
D	0.87	0.54	0.88	0.67	0.47	0.94	0.82	0.66	1.95
E	0.90	0.42	0.76	0.54	0.18	0.46	0.90	0.43	0.70
F	0.92	0.70	0.69	0.69	0.25	0.59	0.90	0.44	1.19
G	0.86	0.41	0.85	0.55	0.23	0.49	1.29	0.46	4.02
H	0.85	0.72	0.58	0.64	0.52	0.99	0.95	0.84	0.47
I	0.86	0.50	0.94	0.65	0.36	0.70	1.04	0.42	2.41
J	0.58	0.17	0.99	0.29	0.18	0.25	0.63	0.38	0.32
K	0.93	0.63	0.94	0.75	0.36	0.89	0.81	0.54	0.84
L	0.76	0.13	0.98	0.22	0.17	0.33	0.88	0.33	0.55
M	0.90	0.08	0.80	0.15	0.13	0.41	0.94	0.20	0.62
N	0.86	0.56	0.83	0.66	0.82	1.57	0.83	0.81	1.09
O	0.70	0.38	0.97	0.55	0.32	0.46	0.96	0.47	1.02
P	0.83	0.31	0.66	0.42	0.66	1.37	0.90	0.68	18.47
Q	0.71	0.29	0.99	0.45	0.62	0.96	1.07	0.67	1.28
R	0.24	0.98	0.16	0.28	0.35	0.37	3.24	0.40	1.85
S	0.48	0.95	0.34	0.50	0.66	0.75	3.82	0.75	2.16
T	0.22	0.95	0.22	0.36	0.14	0.14	1.41	0.26	0.77
U	0.38	0.99	0.37	0.54	0.53	0.53	0.92	0.52	3.07

Table C.8: Cooling and behavioural characterization assessment metrics for each household under the DAR-Fit2 model

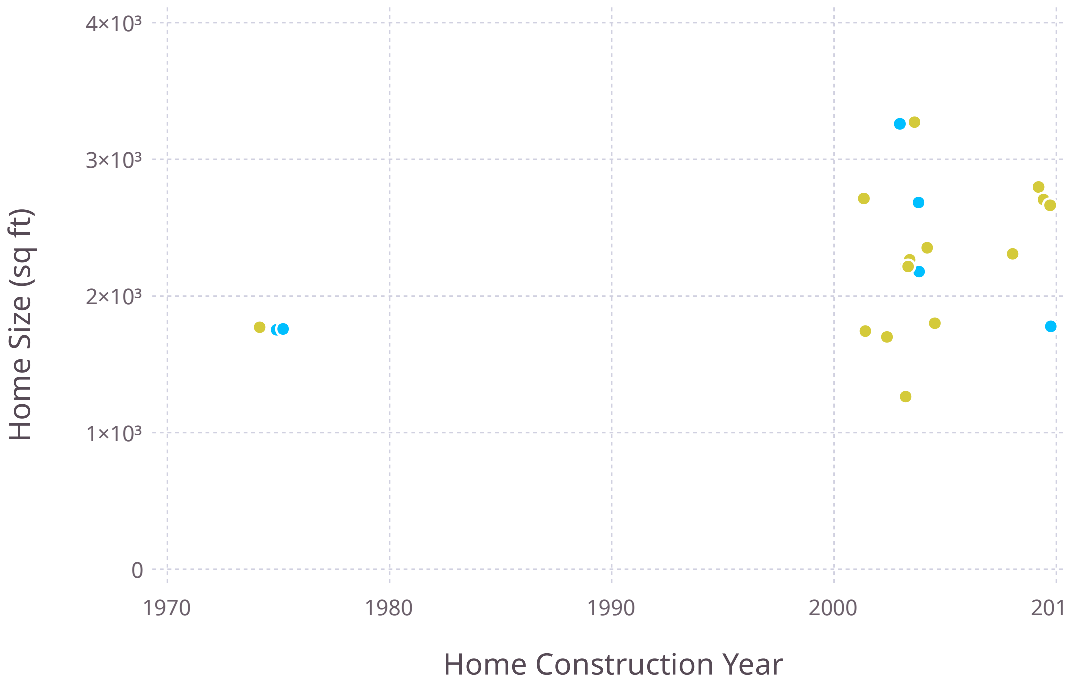

Figure C.4: Best cluster assignment of EHMS homes using physical parameters from the DAR-Fit2 model

C.5 Deseasonalized Autoregressive Three-Regime Model (DAR-Fit3)

ID	Heating Activity Accuracy	Heating Activity Precision	Heating Activity Recall	Heating Activity F1	Heating Load RMSE (overall)	Heating Load RMSE (active)	Heating Load MAPE
A	0.70	0.60	0.99	0.75	0.08	0.09	0.94
B	0.89	0.94	0.90	0.92	0.25	0.29	0.88
C	0.52	0.47	1.00	0.64	0.13	0.14	1.16
D	0.83	0.85	0.94	0.89	0.35	0.37	0.96
E	0.81	0.82	0.97	0.89	0.20	0.20	0.89
F	0.51	0.50	1.00	0.67	0.20	0.20	1.05
G	0.57	0.55	1.00	0.71	0.17	0.18	0.94
H	0.82	0.82	0.97	0.89	0.39	0.40	0.86
I	0.54	0.41	0.99	0.58	0.14	0.16	1.07
J	0.80	0.80	1.00	0.89	0.12	0.12	0.83
K	0.89	0.89	1.00	0.94	0.12	0.12	0.95
L	0.44	0.44	1.00	0.61	0.10	0.10	1.79
M	0.13	NA	0.00	NA	0.09	0.09	1.00
N	0.81	0.81	1.00	0.90	0.21	0.21	2.18
O	0.18	0.18	1.00	0.31	0.14	0.14	4.42
P	0.09	NA	0.00	NA	0.29	0.31	1.00
Q	0.88	0.88	1.00	0.94	0.13	0.13	5.63
R	0.09	0.09	1.00	0.17	0.16	0.16	9.17
S	0.23	0.23	1.00	0.38	0.16	0.16	5.20
T	0.01	0.01	1.00	0.03	0.09	0.09	6.56
U	0.02	0.02	1.00	0.04	0.06	0.06	0.91

Table C.9: Heating characterization assessment metrics for each household under the DAR-Fit3 model

ID	Cooling Activity Accuracy	Cooling Activity Precision	Cooling Activity Recall	Cooling Activity F1	Cooling Load RMSE (overall)	Cooling Load RMSE (active)	Cooling Load MAPE	Behavioural Load RMSE	Behavioural Load MAPE
A	0.88	0.30	0.95	0.45	0.19	0.46	0.98	0.44	0.68
B	0.88	0.73	0.89	0.80	0.39	0.65	0.87	0.61	0.87
C	0.89	0.49	0.89	0.63	0.36	0.78	0.83	0.53	0.45
D	0.86	0.52	0.90	0.66	0.47	0.90	0.81	0.58	1.63
E	0.90	0.42	0.77	0.54	0.18	0.45	0.89	0.40	0.59
F	0.92	0.70	0.69	0.69	0.25	0.58	0.91	0.41	0.98
G	0.85	0.39	0.86	0.54	0.23	0.46	1.43	0.39	2.50
H	0.85	0.65	0.74	0.70	0.51	0.91	0.93	0.78	0.43
I	0.85	0.49	0.94	0.65	0.36	0.69	1.06	0.40	2.18
J	0.57	0.17	0.99	0.29	0.18	0.25	0.60	0.37	0.27
K	0.94	0.65	0.93	0.76	0.35	0.88	0.78	0.46	0.53
L	0.76	0.13	0.98	0.22	0.17	0.32	0.87	0.32	0.40
M	0.98	0.17	0.33	0.23	0.13	0.78	0.94	0.21	0.72
N	0.86	0.56	0.83	0.66	0.82	1.57	0.83	0.78	0.95
O	0.68	0.37	0.97	0.53	0.29	0.41	0.84	0.43	0.70
P	0.84	0.31	0.65	0.42	0.66	1.39	0.91	0.74	23.21
Q	0.71	0.29	0.99	0.45	0.61	0.95	1.11	0.63	0.92
R	0.24	0.98	0.16	0.28	0.35	0.37	3.36	0.35	1.14
S	0.47	0.95	0.32	0.48	0.66	0.75	3.87	0.70	1.58
T	0.22	0.95	0.22	0.36	0.14	0.14	1.50	0.23	0.46
U	0.45	0.99	0.45	0.61	0.52	0.52	0.89	0.49	2.82

Table C.10: Cooling and behavioural characterization assessment metrics for each household under the DAR-Fit3 model

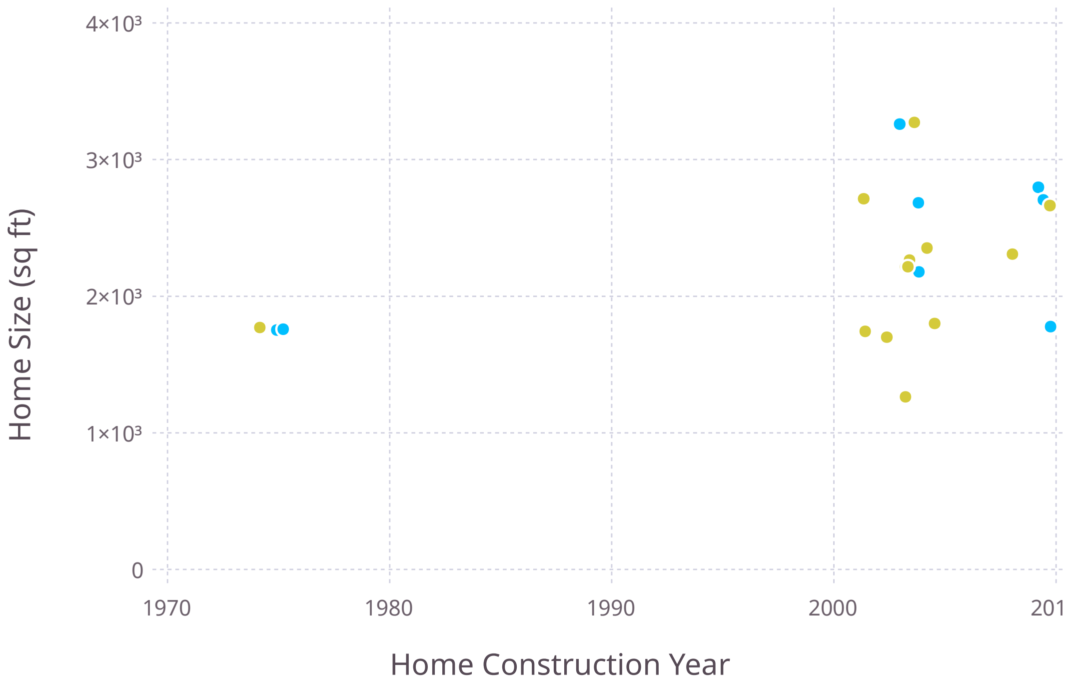

Figure C.5: Best cluster assignment of EHMS homes using physical model parameters from the DAR-Fit3 model

Appendix D

Model Extensions

D.1 State space modelling framework

A model fully addressing the specifications of Section 3.1 must be able to track a variety of both known and unknown values system variables. Some influential factors, such as external temperature or time of day, are clearly measurable and can be incorporated straightforwardly into a model. Other factors, such as a household's internal temperature or the current output of a heating or cooling system, are not widely observable but may be defined in terms of other known quantities or their own previous values. Finally, the system's overall electricity use may be defined in terms of the values of these other known or unknown quantities.

State space modelling provides an established mathematical framework for defining relationships between time-dependent system input, output, and unobserved internal state variables. This general framework has been applied to study, parametrize, and predict the evolution dynamical systems across a range of disciplines, from general chaos theory and econometric time series to ecological population dynamics and engineering control theory. In a state space model, unknown continuous-valued system states evolve interdependently over discrete time steps, with close parallels to a system of first-order differential equations (continuous state variables in continuous time) as well as hidden Markov models (discrete state variables in discrete time). An interpretation of the general state space framework for the context of residential energy use is depicted in Figure D.1.

If the relations governing interactions between system states and known inputs and outputs are linear, state estimates can be obtained through computationally-efficient classic Kalman filtering and smoothing, while in the more general nonlinear case, extensions such as the unscented Kalman filter and smoother must be applied (Durbin and Koopman, 2012).

In addition to unknown time-varying state variables, a system may also be characterized by unknown static parameters (for example, a building's level of insulation or average energy use at a specific time of day). Depending on the nature of their interactions with other variables, these parameters may be treated as special system states that are constrained to remain fixed throughout time. For example, exogenous regression coefficients can be

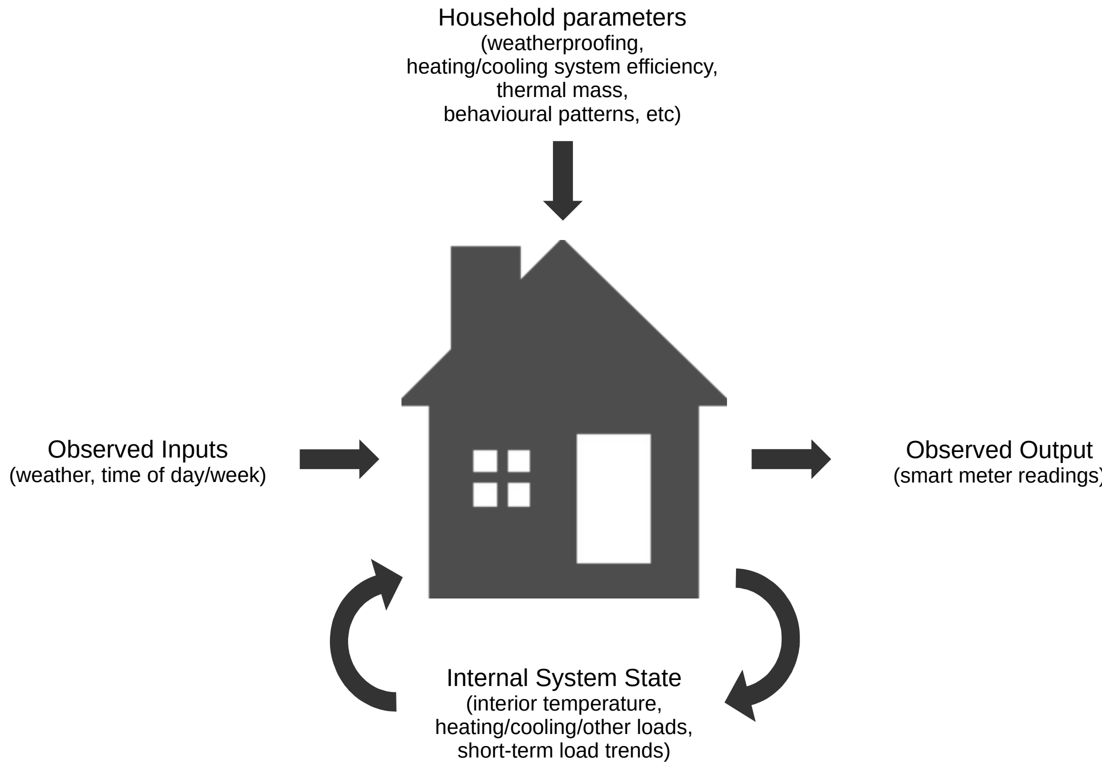

Figure D.1: Qualitative state space representation of household energy use

efficiently estimated by formulation as constant states in a linear system (Durbin and Koopman, 2012), or parameters may more generally be expressed as states in a parallel nonlinear system to be estimated through joint or dual estimation processes (Haykin, 2004). Alternatively, such fixed parameters may be fit to the data through an iterative process such as EM estimation, or maximum likelihood estimation via general nonlinear optimization methods (Durbin and Koopman, 2012).

Another benefit of a state space approach to system modelling is its modularity. Models may be developed and tested for independent subcomponents of a system and then combined into a single overall system representation while still maintaining the ability to later introduce interactions between these components. The following sections take advantage of this property by developing and validating heating, cooling, and behavioural models independently in advance of their eventual integration.

D.2 Thermal extensions

D.2.1 Bounded variable internal temperature model

A more realistic assumption regarding the objectives of a building's heating and cooling systems is that interior temperature is allowed to vary in time, but is constrained to fall within some comfortable range of values, $T_{min} \leq T_{int}(t) \leq T_{max}$. In this context, no active intervention is required when the building's internal temperature is within the comfort range. Outside of this range, active intervention is required to both stabilize interior temperature from deviating further and restore it to a minimum / maximum acceptable value. Recall the general heat flow model:

$$C_{th}\frac{\delta T_{int}(t)}{\delta t} = \tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) + q_h(t) + q_c(t)$$

Decomposing heat requirements into stabilizing and restorative elements gives:

$$q_h(t) = \begin{cases} q_{h_{\text{stabilize}}}(t) + q_{h_{\text{restore}}}(t), & T_{int}(t) < T_{min} \\ 0, & T_{int}(t) \geq T_{min} \end{cases}$$

$$q_c(t) = \begin{cases} 0, & T_{int}(t) \leq T_{max} \\ q_{c_{\text{stabilize}}}(t) + q_{c_{\text{restore}}}(t), & T_{int}(t) > T_{max} \end{cases}$$

Stabilizing interior temperature such that $q_{net} = 0$ (before the effects of other active heating components) is a similar exercise as the constant-interior-temperature case described previously, only now $T_{int}(t)$ is allowed to vary in time and not guaranteed to a desirable value (hence the need for a second restorative term):

$$q_{h_{\text{stabilize}}}(t) = -\tilde{U}\big(T_{ext}(t) - T_{int}(t)\big)$$
$$q_{c_{\text{stabilize}}}(t) = -\tilde{U}\big(T_{ext}(t) - T_{int}(t)\big)$$

Restoring interior temperature to some desirable value is accomplished by additional active heat transfer beyond that required to simply maintain interior temperature at a constant value. There are many possibilities for the design of such a system, but for simplicity a simple proportional controller will be assumed:

$$q_{h_{\text{restore}}}(t) = -k_h\big(T_{ext}(t) - T_{min}(t)\big)$$
$$q_{c_{\text{restore}}}(t) = -k_c\big(T_{ext}(t) - T_{max}(t)\big)$$

In the long-timestep discrete time case, one possible interpretation of such a controller design is that the resulting system output is the product of the fixed output of the heating system (proportional to k) and the proportion of the time period that the system is required to be active (proportional to $T_{ext}(t) - T_{max}$).

Combining these components, we get:

$$q_h(t) = \begin{cases} -\tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) - k_h\big(T_{ext}(t) - T_{min}\big), & T_{int}(t) < T_{min} \\ 0, & T_{int}(t) \geq T_{min} \end{cases}$$

$$q_c(t) = \begin{cases} 0, & T_{int}(t) \leq T_{max} \\ -\tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) - k_c\big(T_{ext}(t) - T_{max}\big), & T_{int}(t) > T_{max} \end{cases}$$

or, in terms of electrical load,

$$P_h(t) = \begin{cases} -\frac{\tilde{U}}{e_h}\big(T_{ext}(t) - T_{int}(t)\big) - \frac{k_h}{e_h}\big(T_{ext}(t) - T_{min}\big), & T_{int}(t) < T_{min} \\ 0, & T_{int}(t) \geq T_{min} \end{cases}$$

$$P_c(t) = \begin{cases} 0, & T_{int}(t) \leq T_{max} \\ \frac{\tilde{U}}{e_c}\big(T_{ext}(t) - T_{int}(t)\big) + \frac{k_c}{e_c}\big(T_{ext}(t) - T_{max}\big), & T_{int}(t) > T_{max} \end{cases}$$

In this case, $T_{int}(t)$ is time-varying and requires an explicit representation. This can be obtained by solving the differential equation given by the original thermal model:

$$C_{th}\frac{\delta T_{int}(t)}{\delta t} = \tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) + q_h(t) + q_c(t)$$

Alternatively, and as is more relevant in this context, the equation can be discretized via forward differencing and represented as a recursion relation:

$$C_{th}\frac{T_{int_{t+1}} - T_{int_t}}{\tau} = \tilde{U}\big(T_{ext_t} - T_{int_t}\big) + q_{h_t} + q_{c_t}$$

The discrete-time thermodynamic system can then be expressed as:

$$T_{int_{t+1}} = T_{int_t} + \frac{\tau}{C_{th}}\tilde{U}\big(T_{ext_t} - T_{int_t}\big) + \frac{\tau e_h}{C_{th}}P_{h_t} + \frac{\tau e_c}{C_{th}}P_{c_t}$$

$$P_{h_{t+1}} = \begin{cases} -\frac{\tilde{U}}{e_h}\big(T_{ext_{t+1}} - T_{int_{t+1}}\big) - \frac{k_h}{e_h}\big(T_{ext_{t+1}} - T_{min}\big), & T_{int_{t+1}} < T_{min} \\ 0, & T_{int_{t+1}} \geq T_{min} \end{cases}$$

$$P_{c_{t+1}} = \begin{cases} 0, & T_{int_{t+1}} \leq T_{max} \\ \frac{\tilde{U}}{e_c}\big(T_{ext_{t+1}} - T_{int_{t+1}}\big) + \frac{k_c}{e_c}\big(T_{ext_{t+1}} - T_{max}\big), & T_{int_{t+1}} > T_{max} \end{cases}$$

D.3 Coupled thermo-behavioural models

The previous section considered thermal and behavioural electrical loads in isolation, but in reality, the two components are interdependent. Heating and cooling performance affects occupant behaviours (for example, an individual's decision whether to stay at home or go elsewhere on a hot summer afternoon), and occupant behaviours affect heating and cooling requirements as well (the number of occupants in a building and the nature of their activities influences interior temperature and thus heating requirements). In general, such interactions are highly variable and difficult to model both precisely and in a generalizable manner. However, some interactions can be modelled according to general physical and statistical relationships: interplays in which occupant behaviour influences heating requirements (but not vis-versa) are considered in this section.

While thermal electrical loads are applied explicitly towards creating heat transfer, behavioural loads also influence internal temperatures. The power consumed by all behavioural loads is eventually dissipated as heat, whether it be in a stove, an electronic device, lighting, or friction in some mechanical process. In many cases, this dissipated heat is released inside the building envelope, contributing to raising internal temperatures (there are of course notable exceptions to this general principle, including electrically-heated water that is disposed into sewer systems without heat recovery, outdoor appliances, and stored energy that is used outside the home (electric vehicles, etc). Increased behavioural loads may also indicate higher or more active building occupancy, correlating with increased heat dissipation from non-electrical occupant activity.

A simple model of incidental heat released by occupant activity can be expressed as:

$$q_{\text{behavioural}} = \alpha P_o$$

where α is some proportionality constant relating electrical behavioural loads to indirect heat production. Augmenting the basic passive heating model with this term gives:

$$q_{\text{passive}}(t) = q_{\text{conductive}}(t) + q_{\text{behavioural}}(t) = \tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) + \alpha P_o(t)$$

The constant and extrema-bound thermal models can now be re-derived to incorporate this coupling with behavioural loads.

D.3.1 Behaviour-dependent constant interior temperature model

The constant-interior-temperature relation can be restated with an updated q_{passive} incorporating behavioural considerations:

$$C_{th}\frac{\delta T_{int}}{\delta t} = 0 = q_{net}(t) = q_{passive}(t) + q_h(t) + q_c(t)$$
$$q_{passive} + q_h + q_c = 0$$
$$\tilde{U}\big(T_{ext}(t) - T_{int}\big) + \alpha P_o(t) + q_h(t) + q_c(t) = 0$$

The new behavioural term shifts the equilibrium point at which $q_{\text{passive}} = 0$ and no active intervention is required. Solving for the new value:

$$\tilde{U}\big(T_{ext}(t) - T_{int}\big) + \alpha P_o(t) = 0$$

$$T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) = T_{int}$$

Now, $q_h(t) > 0$ becomes necessary when $T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) < T_{int}$, and $q_c(t) < 0$ is required when $T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) > T_{int}$. Formalizing this:

$$q_h(t) = \begin{cases} -\tilde{U}\big(T_{ext}(t) - T_{int}\big) - \alpha P_o(t), & T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) < T_{int} \\ 0, & T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) \geq T_{int} \end{cases}$$

$$q_c(t) = \begin{cases} 0, & T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) \leq T_{int} \\ -\tilde{U}\big(T_{ext}(t) - T_{int}\big) - \alpha P_o(t), & T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) > T_{int} \end{cases}$$

In terms of electrical power, the relations are

$$P_h(t) = \begin{cases} -\frac{\tilde{U}}{e_h}\big(T_{ext}(t) - T_{int}\big) - \frac{\alpha}{e_h} P_o(t), & T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) < T_{int} \\ 0, & T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) \geq T_{int} \end{cases}$$

$$P_c(t) = \begin{cases} 0, & T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) \leq T_{int} \\ \frac{\tilde{U}}{e_c}\big(T_{ext}(t) - T_{int}\big) + \frac{\alpha}{e_c} P_o(t), & T_{ext}(t) + \frac{\alpha}{\tilde{U}} P_o(t) > T_{int} \end{cases}$$

D.3.2 Behaviour-dependent extrema-bound interior temperature model

As in the constant-temperature case, the extrema-bound interior temperature model can be updated to include behavioural coupling as:

$$C_{th} \frac{\delta T_{int}(t)}{\delta t} = \tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) + \alpha P_o + q_h(t) + q_c(t)$$

The decomposed heat requirements remain identical:

$$q_h(t) = \begin{cases} q_{h_{\text{stabilize}}}(t) + q_{h_{\text{restore}}}(t), & T_{int}(t) < T_{min} \\ 0, & T_{int}(t) \geq T_{min} \end{cases}$$

$$q_c(t) = \begin{cases} 0, & T_{int}(t) \leq T_{max} \\ q_{c_{\text{stabilize}}}(t) + q_{c_{\text{restore}}}(t), & T_{int}(t) > T_{max} \end{cases}$$

Restorative heat also remains the same, however the heat flow required for stabilizing interior temperature (such that $q_{net} = 0$ before the effects of other active heating components) has now shifted as in the constant temperature case:

$$q_{h_{\text{stabilize}}}(t) = -\tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) - \alpha P_o(t)$$
$$q_{c_{\text{stabilize}}}(t) = -\tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) - \alpha P_o(t)$$

Combining these components, we get:

$$q_h(t) = \begin{cases} -\tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) - \alpha P_o(t) - k_h\big(T_{ext}(t) - T_{min}\big), & T_{int}(t) < T_{min} \\ 0, & T_{int}(t) \geq T_{min} \end{cases}$$

$$q_c(t) = \begin{cases} 0, & T_{int}(t) \leq T_{max} \\ -\tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) - \alpha P_o(t) - k_c\big(T_{ext}(t) - T_{max}\big), & T_{int}(t) > T_{max} \end{cases}$$

or, in terms of electrical load,

$$P_h(t) = \begin{cases} -\frac{\tilde{U}}{e_h}\big(T_{ext}(t) - T_{int}(t)\big) - \frac{\alpha}{e_h} P_o(t) - \frac{k_h}{e_h}\big(T_{ext}(t) - T_{min}\big), & T_{int}(t) < T_{min} \\ 0, & T_{int}(t) \geq T_{min} \end{cases}$$

$$P_c(t) = \begin{cases} 0, & T_{int}(t) \leq T_{max} \\ \frac{\tilde{U}}{e_c}\big(T_{ext}(t) - T_{int}(t)\big) + \frac{\alpha}{e_c} P_o(t) + \frac{k_c}{e_c}\big(T_{ext}(t) - T_{max}\big), & T_{int}(t) > T_{max} \end{cases}$$

The definition of $T_{int}(t)$ has also changed. The relevant behaviour-coupled differential equation is now:

$$C_{th}\frac{\delta T_{int}(t)}{\delta t} = \tilde{U}\big(T_{ext}(t) - T_{int}(t)\big) + \alpha P_o(t) + q_h(t) + q_c(t)$$

Once again, the equation can be discretized via forward differencing and represented as a recursion relation:

$$C_{th}\frac{T_{int_{t+1}} - T_{int_t}}{\tau} = \tilde{U}\big(T_{ext_t} - T_{int_t}\big) + \alpha P_{o_t} + q_{h_t} + q_{c_t}$$

Finally, the discrete-time thermodynamic system of equations can be expressed as:

$$T_{int_{t+1}} = T_{int_t} + \frac{\tau}{C_{th}}\tilde{U}\big(T_{ext_t} - T_{int_t}\big) + \frac{\tau\alpha}{C_{th}}P_{o_t} + \frac{\tau e_h}{C_{th}}P_{h_t} + \frac{\tau e_c}{C_{th}}P_{c_t}$$

$$P_{h_{t+1}} = \begin{cases} -\frac{\tilde{U}}{e_h}\big(T_{ext_{t+1}} - T_{int_{t+1}}\big) - \frac{\alpha}{e_h} P_{o_{t+1}} - \frac{k_h}{e_h}\big(T_{ext_{t+1}} - T_{min}\big), & T_{int_{t+1}} < T_{min} \\ 0, & T_{int_{t+1}} \geq T_{min} \end{cases}$$

$$P_{c_{t+1}} = \begin{cases} 0, & T_{int_{t+1}} \leq T_{max} \\ \frac{\tilde{U}}{e_c}\big(T_{ext_{t+1}} - T_{int_{t+1}}\big) + \frac{\alpha}{e_c} P_{o_{t+1}} + \frac{k_c}{e_c}\big(T_{ext_{t+1}} - T_{max}\big), & T_{int_{t+1}} > T_{max} \end{cases}$$