Control event matrices for stochastic heating use in urban-scale energy simulations

Peter McCallum[1], David P. Jenkins[1], Paraskevi Vatougiou[1]
[1]Urban Energy Research Group, Heriot-Watt University, UK

Abstract

The prevalence of Urban-scale Building Energy Modelling (UBEM) in recent literature emphasises the significant shift of focus in how we consider energy demands from the built environment. This change has of course been driven by the major transitions taking place in our energy systems, resulting from far-reaching decarbonisation initiatives. As is always the case for numerical modelling, a model is only as good as the data used in its formation. This work, contributes to one facet of the growing basis of knowledge around formal Uncertainty Analysis in building modelling – the aleatory uncertainty associated with the stochastic behaviour of people.

A hybrid rule-based, clustering method is presented, which identifies events in households that correspond to a changes between setback and comfort temperatures. To catalogue the resulting data, 'control event matrices' are compiled to record the probabilistic relationships between the timings of set point changes. This particular scheme has been designed to provide a lightweight and extensible model; it facilitates both the automated processing of large scale smart meter datasets, and the stochastic regeneration of control schedules for UBEM. The demonstration data (AECOM, 2018) featured 1,300 dwellings, and resulted in data structures five orders of magnitude smaller than the original dataset. The basis for generating such data is to provide a methodology for interpreting, anonymising and condensing vast records of smart meter data, to potentially offer near real-time observation of widespread behavioural change around heating practices and technology adoption, and augment traditional Time Use Survey methods.

Introduction

The way we use and generate energy is changing rapidly. Of the many factors that have taken us to this point, decarbonisation commitments have impacted every aspect of the whole energy system, from the consumers to the major generators and transmission network operators. The wholesale and spot markets continually undergo adaptations to respond to the needs of grid control rooms, while generators, utilities and aggregators balance innovative grid services products with peak plant assets and grid-scale batteries. Throughout the UK's energy system, organisations must now respond to variability of almost 50GW of distributed renewable plant, a rapidly evolving energy marketplace, and uncertain energy demand transformations.

As one of the major facets of energy demand uncertainty, the transition of the residential sector is extremely complex. Policy makers have played a significant role in decarbonising the UK's energy system, and have demonstrated how responsive the public can be to schemes such as Feed-in-tariffs and the Renewable Heating Incentive. Restrictive measures around fossil fuel use are also now in place. On one hand, the onus is placed on house builders to find new solutions. On the other, Distribution Network Operators (DNOs) must contend with a range of potential demand scenarios that strain the electricity network far beyond intended design.

Favourable scenarios for the UK residential sector include a large uptake of heat pumps (National Grid, 2020), although other electricity-based heating systems are also expected to become more widespread. Through periods of significant change resulting from current and future policy on fossil fuel-based heating (at present the phase-out affects new homes from 2025), DNOs require more advanced techniques to understand what demand will 'look' like in coming years, to plan ahead for potential reinforcement or other measures. In combination with energy system and operational network models, Urban Building Energy Modelling (UBEM) is becoming an increasingly important tool to design and plan for the future.

Physics-based thermodynamic modelling of buildings theoretically offers greater fidelity over other method; however, the input configuration of these models presents major issues. Definition of physical parameters across building stock is complex, often involving incomplete data with limited detail. In the context of the two classes of uncertainty in numerical modelling – aleatory and epistemic (Helton et al., 2010) – the process of creating the stock description of a UBEM generally results in the latter, due to structural and algorithmic behaviour of the numerical representations describing the underlying mathematics, inconsistency or lack of input and calibration data, and necessary simplifications for the benefit of tractability.

Temporal aspects of modelling transpire in aleatory uncertainty, as we cannot know deterministically how occupants will respond in any given dwelling. The control of heating (or cooling) systems across any stock, however, is of fundamental importance in UBEMs. Adequate representations of both the patterns and temporal diversity of system control are necessary to provide reasonable predictions of the intra-day demands on a network.

Stock models have been scrutinised on this basis for a number of years. Huebner et al. (2013) found from measurements that it was common for heating to be used outside typically assumed hours, and that diversity in stock models is poorly represented. Kane et al. (2015) also found various patterns more complex than the widely used BREDEM and Cambridge Housing Models, and provided results on the number of heating periods throughout the day at around 250 dwellings. Various other studies have employed direct controller measurement at tens to hundreds of dwellings, investigating underlying cause of diversity and controller-override behaviours (Bruce-Konuah et al., 2019; Hanmer et al., 2019).

Time Use Surveys (TUS) such as the UK 2000-2001 survey (Ipsos-RSL/ONS, 2003) have long provided valuable data on a much bigger scale, into the tens of thousands. The widely used Richardson et al. (2008) model for stochastic residential energy behaviour used this data as its basis; the same survey data continues to serve this purpose (McKenna et al., 2020). More recent large scale TUS include the UK 2014-2015 survey (Gershuny and Sullivan, 2017); application of this data was demonstrated by McKenna et al. (2015) via the CREST model. An occupancy-integrated archetyping scheme was also presented by Buttitta et al. (2019). Similarly, Dutch and Belgian TUS data was used to develop the powerful stochastic modelling tool StROBe (Baetens and Saelens, 2016), which applies an occupancy chain method to interpret activity in the household and approaches aleatory uncertainty using probabilistic methods.

These surveys, however, are vast undertakings that are likely to be limited in frequency to a decade or longer. To augment and support rich TUS data, other methods also exist to infer temporal behaviour signals for urban-scale modelling. In particular, there have been numerous studies on pattern identification, data mining and Machine Learning (ML) methods linked to electricity smart meter data (Silva et al., 2011; Benítez et al., 2014; Granell et al., 2015; Ramos et al., 2015; Wang et al., 2016), generally treating total energy consumption or non-heating energy consumption. With regard to heating specific energy consumption, there are significant opportunities to mine valuable insight from gas meter records, an area of underdevelopment in the literature.

The modelling work presented in McCallum et al. (2020) is extended here. The 'ParaDwell.jl' model was previously described, in which automated methods are used to build parametric descriptions of residential building stock. Large quantities of building simulation models are generated to provide physical archetypes (in EnergyPlus), reflecting building form, materials, and heating system. In contrast to Buttitta et al. (2019), the household is intentionally distinct from physical stock descriptions, in order to introduce probabilistic control methods. At present, ParaDwell is restricted to thermal modelling in heating climates, and takes set point control schedules as a principal temporal input. This paper covers a hybrid rule-based, clustering method for inferring heating-use patterns from gas smart meter data, which can support urban-scale simulation models such as ParaDwell.

Descriptions of the method specification and detailed process are provided in the next section. A specific dataset is described in the subsequent section, which is then applied to the method, and examined in terms of a detailed single-site illustration, and broader stock-scale results.

Specification and method

Process requirements

The outline requirements of the presented method are provided in Table 1.

	Requirements
1.	To provide the basis for probabilistic temporal control functions and corresponding *temporal diversity* in physics-based, energy demand models;
2.	To automate processes which utilise large quantities of smart meter data to interpret underlying behaviours, useful for Urban Building Energy Modelling;
3.	To augment and complement both traditional Time Use Survey data and the evolving Machine Learning based methods for energy modelling, to provide a logic-based scheme that can be interpreted and enhanced with new methods;
4.	To facilitate modelling following cultural adaptations to behaviour due to changes in energy practices and propagation of new technology (e.g. heat pumps, demand side response, batteries)
5.	To provide robust and lightweight code modules and data schemas, which are inherently extensible and adaptable;
6.	To ensure model(s) are replicable, accessible, versatile and portable (through open source code).

Table 1: Control event interpretation requirements.

To meet these requirements, a hybrid rule-based, clustering method was developed. By providing a method to process and anonymise smart meter data, this scheme is designed to collapse useful outputs (control schedules) into lightweight, extensible data structures (control event matrices).

Control event interpretation method

Figure 1 outlines the semi-analytic process to interpret control events. In particular, the two-level iterative procedure is described in the following sections, which captures variation between each day (of N_d) across $N_\mathbf{d}$ dwellings. The specific cleaning process required for the applied dataset is not described; this must be tailored for every new dataset (of $N_\mathbb{D}$). More details are provided on the relevant distribution matrices (†) under different scenarios in the Application and Results section. Notional implementation considerations are discussed later in the conclusions.

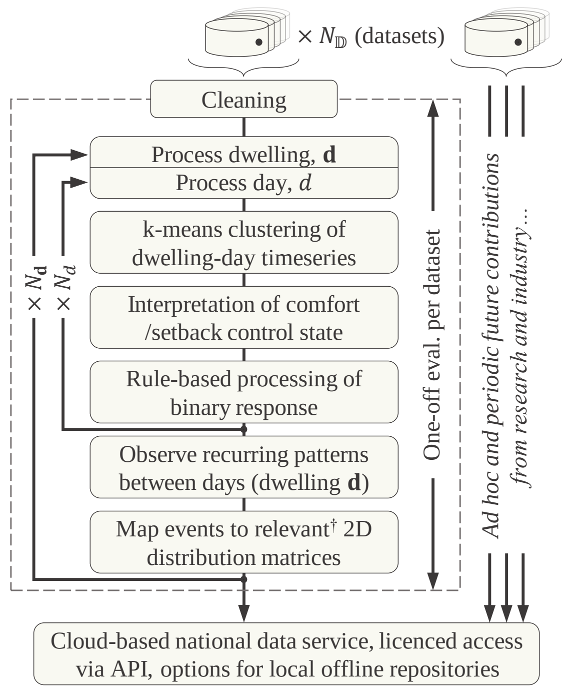


Figure 1: Method process, including notional implementation.

Dwelling-day binary response

Data is sequentially processed into 'dwelling-days', cycling through each day, d, at each dwelling, $\mathbf{d}$. These timeseries are initially processed to delineate between comfort and setback heating states directly from the smart meter data. Common interpretations that heating can only be on or off depending on awake/asleep/absent occupation, are not applied. More generally, occupancy is not used to strictly dictate heating use for the following reasons:

1. setback conditions are permitted during both absence and periods of sleep, to reflect heating system response during severe cold spells (selected systems can still be forced 'off' if required by lowering the setback temperature further);
2. setback conditions during occupied-awake periods are also permitted, making allowances for intentional setback (this could result from energy saving practices for a partially occupied dwelling during certain times, or fuel poverty, should data is available).

The data-driven comfort/setback interpretation was via k-means clustering on the truncated timeseries' of gas energy consumption for each dwelling-day, $E_{g,\mathbf{d},d}$. Tests were carried out to empirically determine stable and consistent behaviour using the k-means method for this purpose; three seeds were both necessary and sufficient to cluster cumulative gas records (specifically in kWh/30min) in a way that clearly segregated the heating system idle state from the two other clusters, representing measurable gas consumption. This generates a half-hourly sequence of binary setback/comfort control states, θ. A dwelling-day sample is provided in Figure 2.

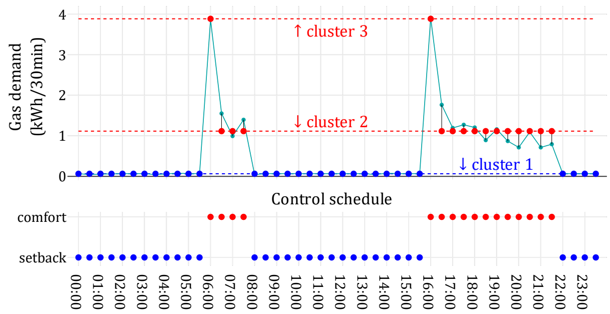


Figure 2: Clustering of gas consumption values for heating control evaluation, one 'dwelling-day'.

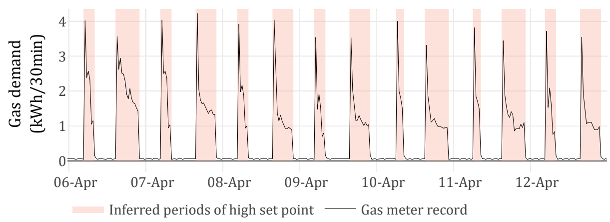


Figure 3: Week-long gas meter record from a single site, with periods of machine-inferred high set point.

A week-long example of this process is provided in Figure 3. Through each dwelling-day that is processed, a new truncated timeseries $E_{g,\mathbf{d},d}$ is used in order to characterise behaviour effectively. Between different days, the cluster centres will change.

Dwelling-day event schedules

A daily control schedule or list of setpoint change events, $\Theta_{\mathbf{d},d}$, is then established for every dwelling-day. This takes the form of a list of times, corresponding to changes in the comfort/setback control states. Changes in control state for any complete gas record can be found using Eqn. (1).

$$\forall\ t : \theta(t) \neq \theta(t-1) \Rightarrow \Theta = t \tag{1}$$

For each dwelling-day this becomes $\Theta_{\mathbf{d},d}$. The number of heating events $\mathbf{N}_{event,\mathbf{d},d}$ in any given dwelling-day is simply the length of this control schedule.As $\Theta_{\mathbf{d},d}$ is simply a record of time-of-day, it is also necessary to record whether the set point is high or low at midnight (both are permitted).

Dwelling-day regime classification

The binary heating response is then categorised as one of five control regimes ($\mathbf{R}_{\mathbf{d},d}$):

1. Regimental and coherent controller-led patterns, which repeat consistently day-to-day over the medium to long term, recognising the specific day of the week;
2. Stochastic, ad hoc heating system use, which may or may not follow roughly consistent patterns day-to-day;
3. Heating systems which are turned on continually, 24-hours per day;
4. Idle heating systems, which may arise due to seasonal effects, because the household do not use the heating system (for various reasons), or because a dwelling is empty;
5. Corrupted data, including partially complete days (<3% of data).

More explicitly, the logic definitions in Eqn. (2) define each regime. Straightforward conditions for $\mathbf{R}_{\mathbf{d},d} = 3,4$ require all gas consumption records for the dwelling-day to be above or below 500W for continuous/idle heating, respectively. Furthermore, if more than eight heating events occur in a day, these days can also be classified as continuous/idle heating depending on the standard deviation of $E_{g,\mathbf{d},d}$.

A series of conditional checks are then carried out on $\Theta_{\mathbf{d},d}$ to determine whether or not it is controller-led ($\mathbf{R}_{\mathbf{d},d} = 1$), assuming that if any condition fails, ad hoc heating-use applies ($\mathbf{R}_{\mathbf{d},d} = 2$):

1. An even number of heating events must occur. This ensures that any day that starts with a low set-point condition, also ends with a low set-point condition (or equivalently, high set-points at the start and end of day). A common pattern features four events: one pair of events in the morning (low, to high, then back to low), with a second pair of events in the evening.
2. $\Theta_{\mathbf{d},d}$ is compared against the following four days (where these exist in the data): d-7, d-1, d+1, d+7. If Θ is identical on *any* of those four days, that day is potentially considered controller-led.
3. To confirm $H^r_{\mathbf{d},d} = 1$ (as controller-led), an additional requirement is imposed to ensure $\Theta_{\mathbf{d},d}$ recurs a significant number of times (10 occasions in the current work).

Application and results

Smart meter trails data

A large open dataset (2007-2010 sample of smart meter trials, AECOM (2018), accessible through the UK Data Service, SN:7591) was assessed for underlying heating patterns in order to infer the timing events for comfort/setback control setpoints. This dataset provided historical gas consumption records (in kWh/30min) at ∼1,300 dwellings, with an average time-span covering ∼530 days.

Single site illustration

A single example site is to used illustrate the semi-analytic process undertaken on the gas smart meter data. This particular site exhibited well defined patterns with respect program-led dominated heating use, as shown in Figure 4. Of the five distinct categories of heating control behaviour, three are discernible directly from the timeseries:

3. heating plant in use 24 hours per day (11%);
4. near-zero demand – heating is off (4%);
5. missing data (0.2%).

Category 3 corresponds to periods during the particularly cold winters of 2009 and 2010, in the UK. Category 4 patterns appear to be when the dwelling was vacant during distinct spells in summer (likely to be holidays). Given that the computational processes for identifying these periods (along with the single missing day in April 2009)
are trivial, the main task is to categorise the remaining 84% of data (regimes 1 and 2). As well as determining whether each day is characterised by a controller-led pattern or ad hoc user control, each day needs to be described in terms of the time-schedule of controller events.

Regarding the periods interpreted as continuous heating, aside from clearly having both a raised minimum and average gas demand, a significant feature of this behaviour is that the maximum demand has also dropped. This indicates that the hydronic

$$\mathbf{R}_{\mathbf{d},d} = \begin{cases} 1, & \mathbf{N}_{event,\mathbf{d},d} = \{2,4,6,8\} \quad \wedge \quad \exists\, n : \Theta_{\mathbf{d},d}(t) = \Theta_{\mathbf{d},d+n}(t), \text{where } n = \{\pm 1, \pm 7\}; \\ 2, & \mathbf{N}_{event,\mathbf{d},d} \leqslant 8 \quad \wedge \quad \forall\, n : \Theta_{\mathbf{d},d}(t) \neq \Theta_{\mathbf{d},d+n}(t), \text{where } n = \{\pm 1, \pm 7\}; \\ 3, & \min\limits_{\{\mathbf{d},d\}} E_g > 0.5 \quad \vee \quad (\mathbf{N}_{event,\mathbf{d},d} > 8 \quad \wedge \quad \sigma_{\mathbf{d},d} > 0.1); \\ 4, & \max\limits_{\{\mathbf{d},d\}} E_g < 0.5 \quad \vee \quad (\mathbf{N}_{event,\mathbf{d},d} > 8 \quad \wedge \quad \sigma_{E_g,\mathbf{d},d} \leqslant 0.1); \\ 5, & \text{(where any data are missing)} \end{cases} \tag{2}$$

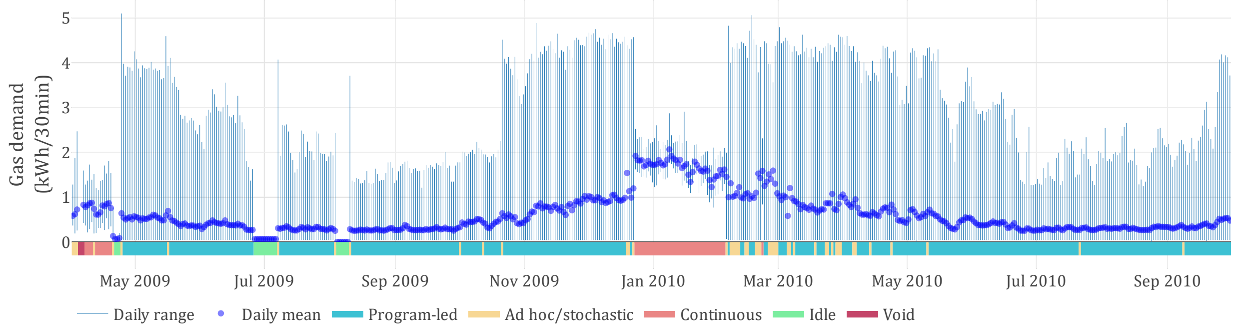
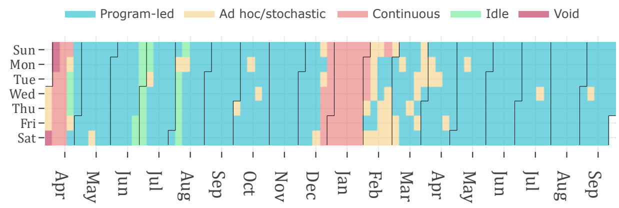


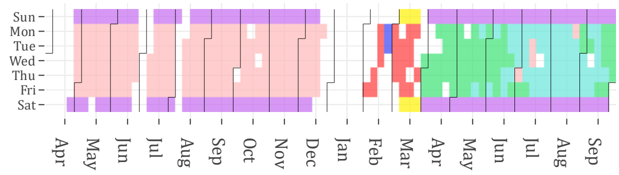
Figure 4: Gas smart meter records for 1 of 1297 dwellings; data from AECOM (2018), SN:7591, site:10662.

system operates in a reasonably stable state, seeing a narrow-banded primary (mixed) return temperature. This contrasts with the program-led and ad hoc regimes, which experience large spikes when pre-heating the closed-circuit water from ambient temperatures after several hours of heating standby. Observing these spikes (or lack thereof), allows for analytic distinction between impulsive heating calls from the occupant(s), and other stochastic gas draws due to domestic hot water (DHW) heating and gas appliance use (there is, however, no current process to separate these other processes from each other).

Of the 84% occurrence of heating regimes $\mathbf{R}_{\mathbf{d},d} = 1, 2$, Figure 5 shows this site is dominated by program-led control. Clear distinction between weekdays and weekends is also clear.



(a) Control regimes



(b) Distinct 'program-led' schedules (7 variants)

Figure 5: Calendar heatmaps for site:10662

Stock analysis

Figure 6 shows the inferred control regime categorisation using the main 18 month period in the data records (raw data from AECOM (2018)). Eqn. (2) was again used to identify and record the daily temporal patterns in binary high/low states for each site, and to interpret how each day related to other days at the same site to infer the underlying control regime. Some general comments are below:

- the most prevalent control regime is through ad hoc boiler use (average of 58% through one year);
- seasonal variation is pronounced for boiler idling conditions as expected (at around 2% in February, rising above 30% in July); in summer, this largely displaces ad hoc patterns;
- continuous-use heating occurs in around 10% of dwellings through summer rising close to 30% in early January, also offsetting ad hoc patterns;
- consistent program-led patterns were identified in around 15% of dwellings; with a slight increase in winter.

A number of more subtle points are also worth emphasising.

- Most sites experience a mix of all five control regimes. Seasonal changes due to system shutdown or continuous use are the most obvious instances of behavioural change; however, weekly ripples in program-led cases in winter suggest that controls are more likely to be overridden by the occupants (or become obscured by DHW demands) at weekends.
- Although the inferred control regime can indicate persistent use of heating through summer for some sites, the demand on the heating system is still far less in summer months (Figure 6 shows occurrence of time-windows rather than magnitude of demand).
- The greatest difficulty and shortcoming of this method is the interpretation of ad hoc boiler use. Program-led, continuous and idle patterns are generally clearly defined when they occur; however, stochastic gas usage patterns can reflect impulsive heating calls from the occupant(s), domestic hot water use or gas appliance use (the latter having considerably less impact). It is worth noting, however, that all three require the occupant to be present in the home.
- Other complications include cases where the heating set point is achieved in a program-led dwelling, the gas demand can stop earlier than usual and the corresponding day no longer reflects neighbouring days (appearing as ad hoc boiler use).

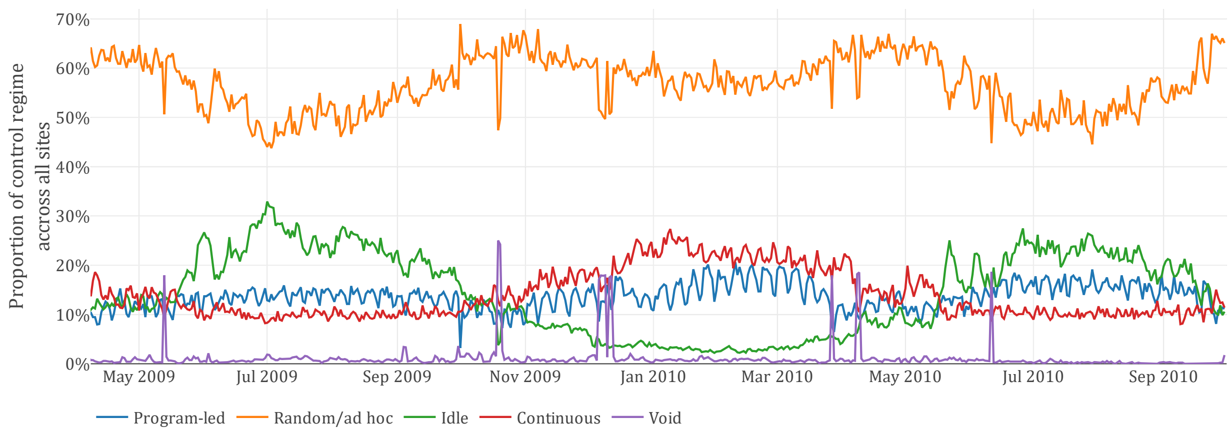


Figure 6: Inferred control regime by percent for all 1297 sites, using the CEI method.

In spite of these short comings, the method demonstrated that, at around 200 sites, coherent program-led patters reoccurred consistently throughout the year. Addressing the above shortcomings would only increase the proportion of program-led cases.

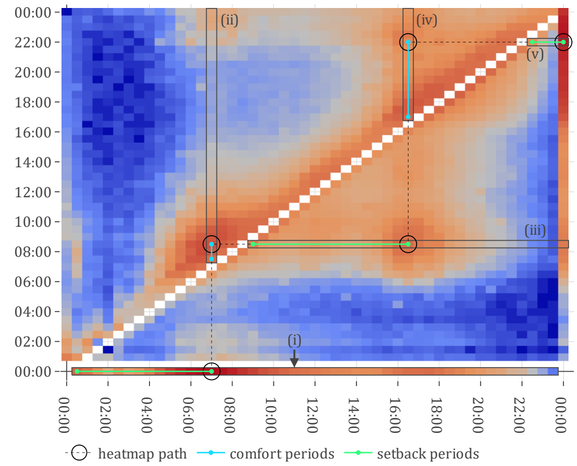
Control matrices

Having extracted inferred heating controller use patterns from the smart meter data, two options are available for deployment of this method for urban-scale building energy modelling.

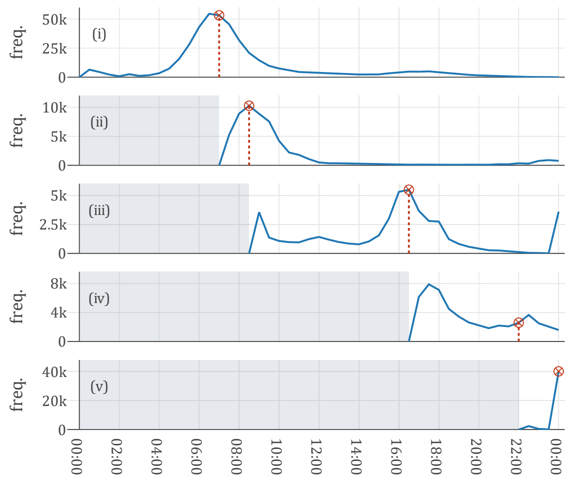
1. Large registers of dwelling-day control schedules ($\Theta_{\mathbf{d},d}$) can be generated via a one-off process. These can then be stored in a database for later (repeated) use.
2. One-off evaluations can be used to interpret $\Theta_{\mathbf{d},d}$, before mapping to a series of 'control event matrices', to build a lightweight stochastic description of behaviour.

Both options are extensible, in that they can be treated as 'living' databases that allow ongoing additions over long periods of time as new data becomes available. The proposed control event matrices, however, are extremely compact, making this option far better suited to distributed computing. As these are statistical distributions, they also offer more permutations than fixed registers of control schedules.

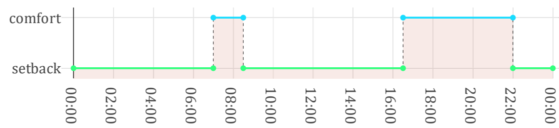
The matrix provided in Figure 7(a) was derived from the present analysis. The matrix is comprised of two halves, split by the diagonal (the diagonal itself is null). In contrast to a true 2-dimensional matrix, the contained values only have meaning when read along specific paths: the lower-right is read horizontally (left to right), and the upper left read vertically (bottom to top). Each matrix half contains 48 univariate distributions, which get progressively shorter when moving across the perpendicular axis. Each distribution describes the probability, with respect to clock-time, that the set point temperature will change from setback to comfort, or vice versa.



(a) Control event matrix, log colour scale.



(b) Discrete distributions, via the path taken in Fig. 7(a).



(c) A regenerated set point schedule for one day.

Figure 7: Application of the control event matrix

The example path which is overlaid on the matrix in Figure 7(a) determines its own route of alternating horizontal and vertical distributions. After the first distribution is evaluated, the resulting value is used to identify which distribution to follow next. Figure 7(b) shows the same set of univariate distributions as Figure 7(a), from (i) through to (v), which ultimately define the regenerated schedule instance in Figure 7(c). As the first distribution returned x=07:00 (along y=00:00) this represents the heating set point moving from setback to comfort, for the first time in the day. Next, follows the upward path at x=07:00, through distribution (ii), which determines that the heating enters the setback condition at y=08:30, and so on. The horizontal lines represent the process of identifying when the heating controller will resume comfort setting; vertical lines relate to when the heating reverts to setback.

The above process describes the application of these matrices, to stochastically regenerate synthesised heating control schedules in large volumes. The prior generation of the matrix itself is effectively the reverse of what is described above, constructed using the smart meter processing methods described in previous sections. A suitably large dataset is a prerequisite of this matrix preparation stage; in the present work ∼450k dwelling-days were used to create Figure 7(a).

This complete process naturally renders an anonymised data structure which, in the present work, was five orders of magnitude smaller than the original 5GB smart meter dataset. Regarding the alternative deployment option referenced earlier in this section (option 1), by retaining the complete register of control schedules, the resulting database would be around three order of magnitude smaller than the original data. While this discrepancy is already significant, these registers would continue to grow linearly with the every new addition. The control event matrices described here could conceivably represent tens of millions dwelling-days using less than 1MB. In the context of distributed computing, integrated Application Program Interface (API) calls within external models, would take in the region of a few seconds.

Conclusion

The basis for smart meter data use in this work is to allow for enhance learning from this vast and expanding data resource. Personal and commercial sensitivity concerns limit open access to this data; in the context of digitisation initiatives, such as the UK's Modernising Energy Data Access (Innovate UK, 2020), it is vital that this kind of data is used appropriately to help implement our future energy systems. By accessing this resource, it is possible to observe large-scale behavioural change, both on a historical basis and in near real-time, that can reflect rapid technology uptake, or significant societal transitions, like changes to home-working practices. This information can also be used to support scenario analyses involving the whole energy system.

The sample data for this application was a 1300-dwelling gas smart meter database (AECOM, 2018), which was approximately 5GB in size, and resulted in a set of matrices measuring less than 200KB. This itself is a fundamental feature of this method. As we continue to move away from desktop studies, where traditional tractability concerns relate to CPU time on local hardware, critical focus must adapt to become more flexible on CPU and GPU demands (i.e. using cloud services and parallel computing infrastructure), and more concerned with data volume. Most major software developments are fully embracing open source languages and libraries, are embedded in automated Continuous Integration (CI) frameworks, and link to arrays of external Application Program Interfaces (APIs) to exploit the most advanced analytics methods available. Single application, local machine-only, packages are now giving way to open source software ecosystems; to make efficient code, we must be concious how much data needs sent between integrated tools via API calls.

The intended application of this work is to maintain registers of set point control schedules built from many smart meter databases, in the form of control event matrices. In most circumstances, one-off calculations would be required to process the matrices for a newly accessed smart meter dataset. For this approach, however, it is essential that the metadata records adopt a structured and extensible system to ensure that users can understand precisely what data was used for each matrix, along with the applied algorithm version. Moreover, it would be imperative to ensure that no dataset is entered more than once.

To augment Time Use Survey data, the use of processed smart meter data could allow for more frequent reviews of behavioural trends, using large scale, geographically contextualised datasets potentially in near real-time. Significant events, such as severe climate shocks, natural disasters and epidemic/pandemic responses can all impact energy use. On a planning and scenario basis, this scheme provides an extensible platform for technology specific developments, for various uptake rates of heat pump for example.

Whilst this work focused on gas smart meter records, future work is planned to use this hybrid rule-based/clustering method to contribute to electricity demand interpretation. Understanding around how dwellings with storage heaters impact the households energy practices; of how sub-optimal heat pump systems respond significantly differently to effective installations, because these are real issues facing DNOs, local authorities, aggregators and other service providers. Coupled with scenarios describing how rapidly technology propagates following new policy, market conditions or product and service offerings, data methods which observe elements of the underlying engineering and practice theory can augment existing machine learning approaches to provide new insight.

Acknowledgement

This work was conducted as part of the ReFELX Project (a PFER demonstrator, main funder: Innovate UK, grant: 104780), and within the scope of the National Centre for Energy Systems Integration (CESI, main funder: EPSRC, grant: EP/P001173/1).

References

AECOM (2018). Energy Demand Research Project: Early Smart Meter Trials, 2007-2010. *UK Data Service SN: 7591*.

Baetens, R. and D. Saelens (2016). Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour. *Journal of Building Performance Simulation 9*(4), 431–447.

Benítez, I., A. Quijano, J. L. Díez, and I. Delgado (2014). Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. *International Journal of Electrical Power and Energy Systems 55*, 437–448.

Bruce-Konuah, A., R. V. Jones, and A. Fuertes (2019). Physical environmental and contextual drivers of occupants' manual space heating override behaviour in UK residential buildings. *Energy and Buildings 183*, 129–138.

Buttitta, G., W. J. Turner, O. Neu, and D. P. Finn (2019). Development of occupancy-integrated archetypes: Use of data mining clustering techniques to embed occupant behaviour profiles in archetypes. *Energy and Buildings 198*, 84–99.

Gershuny, J. and O. Sullivan (2017). United Kingdom time use survey, 2014-2015. *UK Data Service SN: 8128*.

Granell, R., C. J. Axon, and D. C. Wallom (2015). Impacts of Raw Data Temporal Resolution Using Selected Clustering Methods on Residential Electricity Load Profiles. *IEEE Transactions on Power Systems 30*(6), 3217–3224.

Hanmer, C., M. Shipworth, D. Shipworth, and E. Carter (2019). How household thermal routines shape UK home heating demand patterns. *Energy Efficiency 12*(1), 5–17.

Helton, J. C., J. D. Johnson, W. L. Oberkampf, and C. J. Sallaberry (2010). Representation of analysis results involving aleatory and epistemic uncertainty. *International Journal of General Systems 39*(6), 605–646.

Huebner, G. M., M. McMichael, D. Shipworth, M. Shipworth, M. Durand-Daubin, and A. Summerfield (2013). Heating patterns in English homes: Comparing results from a national survey against common model assumptions. *Building and Environment 70*, 298–305.

Innovate UK (2020). Modernising Energy Data Access (MEDA). https://innovateuk.blog.gov.uk/2020/05/29/modernising-energy-data-access-and-the-winners-are/.

Ipsos-RSL/ONS (2003). Office for National Statistics, Ipsos-RSL, United Kingdom Time Use Survey, 2000. *UK Data Service SN:4504*.

Kane, T., S. K. Firth, and K. J. Lomas (2015). How are UK homes heated? A city-wide, socio-technical survey and implications for energy modelling. *Energy and Buildings 86*, 817–832.

McCallum, P. D., D. P. Jenkins, and P. Vatougiou (2020). An urban-scale residential stock model for grid-constrained regions. *Building Simulation and Optimization conference (BSO2020)*.

McKenna, E., S. Higginson, T. Hargreaves, J. Chilvers, and M. Thomson (2020). When activities connect: Sequencing, network analysis, and energy demand modelling in the United Kingdom. *Energy Research and Social Science 69*(September 2019), 101572.

McKenna, E., M. Krawczynski, and M. Thomson (2015). Four-state domestic building occupancy model for energy demand simulations. *Energy and Buildings 96*, 30–39.

National Grid (2020). UK Future Energy Scenarios.

Ramos, S., J. M. Duarte, F. J. Duarte, and Z. Vale (2015). A data-mining-based methodology to support MV electricity customers' characterization. *Energy and Buildings 91*, 16–25.

Richardson, I., M. Thomson, and D. Infield (2008). A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings 40*(8), 1560–1566.

Silva, D. D., X. Yu, D. Alahakoon, G. Holmes, and S. Member (2011). Consumption Analysis From Meter Data. *IEEE Transactions on Industrial Informatics 7*(3), 399–407.

Wang, Y., Q. Chen, C. Kang, and Q. Xia (2016). Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications. *IEEE Transactions on Smart Grid 7*(5), 2437–2447.