High-resolution stochastic integrated thermal-electrical domestic demand model

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

- This is an Open Access Article. It is published by Elsevier under the Creative Commons Attribution 4.0 Unported Licence (CC BY). Full details of this licence are available at: http://creativecommons.org/licenses/by/4.0/

Metadata Record: https://dspace.lboro.ac.uk/2134/20103

Version: Published

Publisher: © The Authors. Published by Elsevier

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
High-resolution stochastic integrated thermal–electrical domestic demand model

Eoghan McKenna *, Murray Thomson

Centre for Renewable Energy Systems Technology (CREST), Wolfson School of Mechanical, Manufacturing and Electrical Engineering, Loughborough University, LE11 3TU, UK

Highlights

- A major new version of CREST’s demand model is presented.
- Simulates electrical and thermal domestic demands at high-resolution.
- Integrated structure captures appropriate time-coincidence of variables.
- Suitable for low-voltage network and urban energy analyses.
- Open-source development in Excel VBA freely available for download.

Abstract

This paper describes the extension of CREST’s existing electrical domestic demand model into an integrated thermal–electrical demand model. The principle novelty of the model is its integrated structure such that the timing of thermal and electrical output variables are appropriately correlated. The model has been developed primarily for low-voltage network analysis and the model’s ability to account for demand diversity is of critical importance for this application. The model, however, can also serve as a basis for modelling domestic energy demands within the broader field of urban energy systems analysis. The new model includes the previously published components associated with electrical demand and generation (appliances, lighting, and photovoltaics) and integrates these with an updated occupancy model, a solar thermal collector model, and new thermal models including a low-order building thermal model, domestic hot water consumption, thermostat and timer controls and gas boilers. The paper reviews the state-of-the-art in high-resolution domestic demand modelling, describes the model, and compares its output with three independent validation datasets. The integrated model remains an open-source development in Excel VBA and is freely available to download for users to configure and extend, or to incorporate into other models.

Corresponding author. Tel.: +44 (0) 7958 531 842.
E-mail address: e.j.mckenna@lboro.ac.uk (E. McKenna).

1. Introduction

The widespread electrification of heat in the domestic sector, through the replacement of gas boilers with heat pumps, is expected to present a major challenge to the operation of electricity distribution networks, due to the large and potentially diversified nature of these loads [1]. The cost of having to reinforce existing electricity networks to accommodate these heat pumps and other low-carbon technologies could be very considerable [2] and thus it is vital to make best use of existing network assets, and to ensure that any reinforcement is based on an accurate assessment of need. This assessment is particularly difficult in the case of low-voltage networks (those which connect from the distribution transformers to the individual dwellings through, for example, 400 V three-phase street mains and single-phase 230 V service connections). Conventional low-voltage network design procedures are not well suited to the task, as they typically use rather simple representations of the varying demand and rely heavily on experience, which is not yet available with widespread low-carbon technologies [3]. To address this, high-resolution models of domestic demand are being developed that can provide a suitable basis for future low-carbon network studies. These models are often based on a core representation of occupancy within...
buildings and produce high-resolution stochastic disaggregated end-use energy demand data at the level of the individual dwelling.

CREST’s existing high-resolution model of domestic electricity demand was developed for this purpose and was based on a two-state active-occupancy model and accounted for electricity consumption associated with lighting, appliances and generation associated with photovoltaic arrays [4–7]. The model has been widely adopted in academia and industry, for example [8–16].

This paper describes the extension of the existing electrical demand model into an integrated thermal–electrical demand model that can provide a convenient basis for studying future network challenges associated with the electrification of heating. The new model includes the previously published components associated with electrical demand and generation (appliances, lighting, and photovoltaics) and integrates these with an updated occupancy model [17], a solar thermal collector model [18], and new thermal sub-models including a low-order building thermal model, stochastic external temperatures, domestic hot water consumption, thermostat and timer controls and gas boilers. The integrated model remains an open-source development in Excel VBA and is freely downloadable [19].

The following section reviews the requirements for domestic demand modelling for low-voltage network applications and describes the features that characterise the state-of-the-art. The integrated model has been developed to include all of these features, and is described in Section 3. Section 4 demonstrates the model’s capture of the appropriate correlation between sub-model outputs through an example of a single day’s simulation. To validate the model, Section 5 compares the model output with independent empirical data.

2. Defining the model requirements

The purpose of CREST’s demand model is primarily for application in low-voltage network analyses. Other models developed elsewhere of similar structure have been used for similar purposes [20–22]. The fitness of such models should therefore be considered in terms of their ability to produce the type of output that is required for this application.

A critical aspect of demand modelling for this purpose is the appropriate representation of the timing of that demand. It is natural that the electricity consumption of an individual dwelling can vary greatly from one moment to the next as appliances within that dwelling are switched on or off by the occupants or automatically. This behaviour is to a large extent random and unpredictable, but it must be taken into account in the design of electricity networks and in particular when considering the low-voltage network. Similar considerations are relevant to the design of gas distribution networks, water supply, sewage and local transport systems, and so there is some overlap of interest in models for these purposes, which we will come back to in Section 2.9.

Recognising that it is not possible to predict the exact behaviour of individual occupants or appliances, the aim of stochastic demand modelling is to provide simulated data that has the right statistics overall, so that it is suitable for the task in hand – in our case, low-voltage network design. A critical precursor to the modelling therefore is the careful consideration of exactly which statistics need to be got right, and, equally important, which aspects may safely be approximated. There is a considerable risk in this type of modelling of attempting to include too much detail in some areas and so creating a model that is too computationally intensive or that requires input data that is simply not available. The choices of what should be included in a model are largely a matter of judgement and the following sections describe the perceived priority requirements that have guided the development of the model presented in this paper.

2.1. High temporal resolution

Electricity demand of individual dwellings is typically characterised by long periods of low to medium demand when multiple small appliances are in use, and occasional spikes of high demand due to kettles and the like. Looking to the future, heat pumps and electric vehicles will likely make these spikes much broader (longer duration) and it is the cumulative effect of this (rather than a big increase in peak demands of the individual dwellings) that presents the main challenge for distribution networks. For the moment however, our focus is on the modelling techniques that can be used to simulate this spikiness. It is important that it is dully represented because it has significant effect on actual customer voltages and network losses (particularly in service cables).

In order to dully represent this ‘spikiness’ it is necessary to use a sufficiently high temporal resolution. The voltage drops and energy losses in electricity networks are all dependent on the instantaneous power flows and can be significantly underestimated if, for example, half-hourly average demand values are used in their calculation [23]. The required temporal resolution to minimise such errors is dependent on the typical switching rate of appliances within dwellings and the desire for modelling precision has to be balanced with the practicality of dealing with large amounts of data. As a compromise, a time resolution of one minute is often used [15,20] and is selected for the model presented in this paper. It should be noted, however, that in practice there may be a limit on the resolution of input data e.g. occupancy data based on time-use surveys is often of 10 min resolution [17]. Readers are referred to [24,25] for analyses of the impact of data averaging on domestic energy demand modelling at the low-voltage distribution network level.

2.2. Demand diversity

Low-voltage network analyses are typically conducted on individual low-voltage network feeders serving up to about 100 dwellings [26]. While there is a need to simulate individual dwellings at high-temporal resolution, the aim is not for the exact prediction of any one specific dwelling, but rather the statistical accuracy of the group.

Network planners often base the design of electricity distribution networks on the ‘after diversity maximum demand’ [26] – the maximum demand, per dwelling, as the number of dwellings connected to the network approaches infinity. While a single dwelling might have a maximum demand in excess of 10 kW, as the number of dwellings is increased the time-coincident maximum per dwelling rapidly approaches the after diversity maximum demand, typically around 2 kW for non-electrically heated dwellings in the UK.

Of the statistics that should be accurately represented, therefore ‘diversity’ is particularly important. We note, however, that the term is ambiguous and deserves clarification. Diversity can refer to the timing of individual demands, i.e. whether they are coincident or correlated in time. Diversity can, however, also refer more generally to the presence of a ‘spread’ or probability distribution of another property of interest, such as magnitude or duration. To be clear, therefore, when we need to be specific we will explicitly mention the type of diversity referred to e.g. ‘time-diversity’. We note also that by ‘demand’ we can be referring to individual dwellings, as well as to the individual appliances or fixtures within a dwelling. To be clear, when it is important to distinguish between these, we will refer to the former explicitly as dwellings and the latter as ‘loads’. 
An important measure of model ‘fitness’, therefore, is the accurate representation of the demand diversity. To do this, a model needs to simulate the loads connected to the network and the timing of their use. Regarding timing, this is a question of accounting for the independence of (or dependence between) individual loads. For example, imagine a hypothetical network of many identical loads cycling on and off. If the operation of every load is completely random and independent of the operation of the others, then the loads will be maximally diversified in time. If, however, every load is operated through a single common controller, which switches them in unison, then the opposite will be true – the loads will be maximally correlated in time. If either of these examples represented how domestic loads were operated in the real world, then the modelling task would be trivial. The complication comes from the fact that demand lies instead somewhere between the two extremes and that methods are needed to represent aspects of both in the model, as described in the next sections.

2.3. Dependency within and between dwellings

Of the dependencies between loads, two are particularly important to represent: dependency of loads within a dwelling, and dependency of loads between dwellings. One of the main dependencies of loads within a dwelling is related to the behaviour of residents, as the operation of many appliances and lighting will be dependent on the residents being at home and awake (‘active occupancy’). As such, ‘activity-based’ models are important for accounting for the appropriate time-correlation of loads within a dwelling. Section 2.5 discusses this type of model further.

Between dwellings, one of the main dependencies of loads is related to shared environmental variables. For example, if a sunny day suddenly turns overcast, then all dwellings with PV will experience a correlated drop in output, as well as an increased probability of occupants using lighting. It is important therefore for a model to capture the shared dependency of demand on the weather, both in terms of diurnal and seasonal variability.

The CREST demand model has been constructed so that these dependencies are only represented (as described in Section 3) and has output with the appropriate statistics to make it ‘fit’ for application in low-voltage network analysis (see Section 5).

2.4. Stochastic modelling

As mentioned above, there is a need to account for the (at least partially) random nature of demand and the fact that while there may be underlying dependencies between certain loads, the precise timing is subject to random variation. Furthermore, there is the need to account for the diversity of other properties of interest, such as the variation of appliance ownership, the randomness of patterns of occupancy, variations in thermal comfort etc.

To address this, stochastic programming techniques are used to produce output that has appropriate probability distributions to those found in the real world. For example, Monte Carlo methods are used to assign parameters by generating random draws from probability distributions of empirically observed data, such as the number of occupants in a dwelling. Another common approach is the use of the Markov-chain technique to generate representative stochastic sequences of discrete random variables, e.g. dwelling occupancy \[17,20,27–31\]. The authors note that while the Markov-chain technique is commonly used, the literature suggests that first order Markov-chains may be unsuitable for modelling certain aspects of domestic behaviour. Readers are referred to the following references for further discussion and analysis \[17,22,30,31\].

Stochastic models are often built around the fundamental technique of comparing software-generated random numbers against pre-determined probabilities, and thus the model construction is mainly a matter of determining those probabilities. One way to approach this would be to analyse measured data, to determine its spikiness etcetera, and to use a random-walk algorithm to simulate it; this would be a ‘top-down’ model, which is relatively easy to construct if sufficient measured data is available. An alternative is a ‘bottom-up’ model in which the spikiness is created by simulating the switching on and off of individual appliances, and this is the approach taken in this paper. Swan and Ugursal [32] provide a fuller discussion of bottom-up versus top-down approaches to demand modelling.

The advantages of the bottom-up approach include that it provides the ability to consider future technology changes, such as replacement of gas boilers with heat pumps, or fuel cell CHP units [33], and that it provides the means to achieve appropriate correlations between dwellings as discussed above. We note a further benefit of stochastic modelling is that it facilitates a model to be self-contained, as the use of large datasets can be avoided by instead using the statistics that summarise and characterise them.

2.5. Activity-based models

People use energy to satisfy needs for comfort, cleanliness, convenience, work, transport, etc. These requirements and the social practices that give rise to them are key factors in determining how energy is used and how this might evolve [34,35]. While historically this ‘human dimension’ has been absent from energy systems models, there is increasing recognition of the value of integrating social and behavioural insights into energy models [36].

Demand models of the type presented in this paper go further than many energy models in addressing this gap. A characteristic feature is their basis on a core representation of the occupancy and activity of individuals within a dwelling, for example [5,20,37–39]. As mentioned above, the purpose of this approach is to represent the dependencies between loads, thereby achieving the appropriate time-diversity of demand. In such models, the representation of occupancy and activity is commonly based on time-use surveys – large nationally-representative surveys detailing the location and activity of people at 10 min resolution for a single 24 h period. See Widén et al. [40] for a review on this subject.

While these models represent steps in the challenge of integrating the human dimension into energy systems models, there is still considerable progress to be made in aligning engineering modelling with social science theory. For further discussion on this subject see [36,41–43].

2.6. Accuracy and computational efficiency

The preceding requirements (bottom-up structure, representation of diversity, high-resolution output, and large numbers of dwellings) all contribute to model complexity and could easily lead to excessive computational run-times. A ‘fit’ model is therefore one that achieves a reasonable balance between model accuracy, complexity, and computational efficiency. For example, the model presented here uses many simplifying assumptions to simulate various aspects of domestic demand, and it could easily have been made more complex. It is, however, useful to re-iterate that the aim is for statistical accuracy of the group rather than absolute accuracy of any one individual building. Given its purpose, therefore, it is important to question whether any added complexity is justified in terms of its impact on the accuracy of the statistics that are important for the application. The Markov-chain technique mentioned previously is a good example of a method that balances accuracy and computational efficiency adequately for the purposes of modelling aspects of occupancy and activity [17,30].
2.7. Low-order thermal models

A particularly challenging area, in terms of its computational burden, is the simulation of the dynamic thermal behaviour of dwellings. There are various well-established and sophisticated packages available for dynamic thermal modelling of buildings [22,44,45]. They are widely used for detailed studies of specific individual buildings under specific conditions (e.g. weather and occupancy profiles). In such models, each building may be represented at the level of its components, such as walls, roofs and windows. The difficulties of adopting such models within a stochastic multi-dwelling model can include the extremely detailed input data requirements and excessive computational runtimes.

Generally, models achieve computational efficiency by using ‘reduced-order’ or ‘low-order’ building thermal models. Examples include those that use one [46], two [47–50] or three thermal masses [14,51], to those where the number is determined by the internal building geometry e.g. the number of internal rooms [15,38]. Identifying suitable reduced order models to adequately model building thermal dynamics for the purposes at hand is an area of considerable complexity and readers are referred the following publications for further information and methods [52,53].

2.8. Transparency and reproducibility

Energy models play a critical role in underpinning national energy policies, and therefore should be accessible and open to independent review [36,54]. Furthermore, the challenge of tackling such a complex domain would benefit from a strategy of model integration [41] and, more generally, learning between modellers. The inner workings of models should therefore be explained in detail and, ideally, the models themselves and input data made accessible. The authors’ models are available as open-source downloads and it is good to see a similar approach being taken elsewhere e.g. the ‘Integrated District Energy Assessment by Simulation’ (IDeAS) and ‘Stochastic Residential Occupant Behaviour’ open-source tools from Leuven University [20,22].

2.9. Broader applications: urban energy system modelling

Low-voltage network analysis can be said to be concerned with an aggregation of dwellings at the scale of a neighbourhood [20]. The next scale up can reasonably be considered that of a district or city. Models at this scale are generally called ‘urban energy system models’ and are concerned with improving the understanding and performance of the systems and infrastructures that mediate between the service requirements of an urban population and the resulting consumption of raw fuels and resources [41]. They have a broad focus that includes modelling all aspects related to urban resource consumption including thermal and cooling demand, transport, climate, water and waste, domestic and non-domestic, and energy supply (e.g. district heating). The state-of-the-art models being developed take a bottom-up ‘microsimulation’ approach where the activities and resource demands of each individual within an urban area are accounted for e.g. [37,55–58]. Urban energy models generally have less of a requirement for high temporal resolution than low-voltage network focussed demand models, and so typically have longer time resolutions, e.g. 5 min [38] to one hour [59], though it should be noted that urban energy systems models are being developed to produce output at very high resolution e.g. seconds [41].

There is, as a result, a large overlap of requirements between the two applications of low-voltage network analysis and urban energy analysis. Indeed, according to the categorisation proposed by Keirstead et al. [41], the integrated model presented here could reasonably be counted as a specific example within the ‘building design’ category of urban energy system models. Therefore, while the CREST model has been developed primarily for low-voltage network applications, it can also serve as a suitable basis for application within the broader field of urban energy systems analysis.

2.10. Summary of model requirements

The model requirements are summarised in Tables 1 and 2 under general and thermal-specific categories. This is the set of guiding principles that were used to judge what should be included in the model, and what should be omitted or simplified in the interests of computational efficiency and in recognition of input data availability limitations.

3. Model description

3.1. Overview

The integrated thermal–electrical demand model is constructed from several sub-models as shown in Fig. 1. The occupancy model generates stochastic sequences of occupancy for each dwelling, which form a basis for the calculation of appliance, lighting and water-fixtures switch-on events. These are aggregated to determine the dwelling’s electricity and hot-water demands. The thermal demand model simulates the dwelling’s thermal dynamics and gas demands given the climate data, internal heat gains, and dwelling-specific building fabric data.

The occupancy, irradiance and external temperature models are the principle means of ensuring the integrated model has appropriately correlated output variables. For example, a dip in modelled irradiance will simultaneously affect four sub-models for every dwelling; solar thermal collector and PV models will show a reduction in output, passive solar gains will reduce in building thermal models, and finally, the use of lighting will be more likely in dwellings that are actively-occupied.

The integrated model retains several sub-models from the previously published model: the irradiance model, the PV model, and the appliance and lighting models (which are part of the electrical demand model). While the new model includes more input data than previously, it remains self-contained, and does not require any further data from the user to run. It also retains a bottom-up structure and is configured to provide output for a single day – this corresponds to the length of diary entries of the time-use survey on which the occupancy model is based.

The new features of the integrated model are sub-models to represent thermal and hot water demands, an external temperature model, a solar thermal collector model and an updated occupancy model. The new model also supports the capability of producing output for many dwellings. The irradiance and temperature sub-models have therefore been configured to provide a
common set of irradiance and temperature data for a group of dwellings, so that the irradiance and temperature dependent sub-models of each dwelling have appropriate time-diversity. Finally, model output has been extended to include not only high-resolution results for each dwelling but also daily totals and high-resolution aggregated results. The following describes the individual components in more detail.

3.2. Occupancy

The occupancy model used here is described fully in a previous publication [17]. It uses a first-order Markov chain technique to create stochastic profiles of dwelling occupancy using transition probability matrices based on UK time-use survey data [60]. It is similar to the ‘active-occupancy’ model [7] that was used previously (active-occupancy means ‘at home and awake’), as both are based on UK time-use survey data and use a first-order Markov-chain technique. The difference is that the updated occupancy model now has four-states: each resident can be at home and active, at home and asleep, away from home and active, or away from home and asleep. This enhancement allows improved representation of casual heat gains from people as they sleep. A second enhancement corrects a slight under-representation of people who stay at home all day; this was considered important because of its relationship to space heating demands. In the four-state model, as previously, dwellings can have up to five residents, and their behaviour is appropriately correlated in time: people may leave or arrive home at the same time. As with the previous version of the model, the present version combines the output from the occupancy model with activity probability profiles.
3.3. Electrical demand

The electrical demand model comprises the appliance and lighting demand models described fully in previous publications [5,6]. The authors note that the previous version included a basic representation of electric storage heaters. These were treated similar to appliances, with switch-on probabilities calibrated to produce annual demands that matched up with typical values, with no representation of temperatures and thermal dynamics. With the development of the present thermal demand model, the simple electric storage heaters that were present in the previous version have now been removed. The only other modification is the addition of electrical loads due to thermal components e.g. pumps associated with the operation of the heating system and solar thermal collector.

3.4. Domestic hot water

Fig. 2 shows the structure of the hot water demand model. Its operational concept follows that of the previously published appliance model. At the beginning of the run, each dwelling is stochastically assigned a set of water fixtures (basins, kitchen sinks, baths and showers) based on the probability of their presence in dwellings [61]. These fixtures are each assigned to an associated activity; Basins, showers, and baths, are assigned to the ‘washing and drying’ activity, while kitchen sinks are assigned to the ‘cooking’ activity. The likelihood of occupants undertaking each activity varies throughout the day, and this is represented by an “activity profile” based on the time-use survey [60]. This concept and the profiles used are the same as the appliance model.

As the model runs, the times of turning on each water fixture are stochastically determined, based on a probability combining occupancy and activity profiles. The volume of hot water drawn for each turn-on event is also determined stochastically and adjusted such that, on average, the total hot water draw per fixture per day matches with empirical data [62].

3.5. Thermal demand and heating system

The thermal demand model has four parts: the building thermal model, the hot water cylinder model, the heating controls, and the heating system. Fig. 3 indicates the structure and data flows. The heating system is currently implemented as a gas boiler. This has been chosen because it represents the predominant type of heating system in the UK today and therefore offers the widest availability of measured operational data against which the integrated model can be validated (Section 5).

3.5.1. Building thermal model

The building thermal model is illustrated in Fig. 4, using the electrical circuit analogy in which voltage is analogous to temperature and current is analogous to heat flow [63]. It has three thermal capacitances, C_v, C_i, and C_m, representing the thermal masses of the building, indoor air, and heat emitters (radiators or underfloor heating). The thermal conductances between these masses are shown as rectangles with heat transfer coefficients H_{va} and H_{im}. The external air temperature is represented as a variable voltage source $\Theta_a$, and $H_{va}$ represents heat transfer from the building fabric. Ventilation heat loss from the indoor air is represented by a conductance $H_v$ the value of which is dependent on the air exchange rate. Passive solar gains, casual heat gains and heating system gains are represented as current sources $\phi_v, \phi_i,$ and $\phi_h$, respectively.

The parameters for heat transfer coefficients and thermal capacitances are determined by calibrating them against the output of a reference software model, as in [14,52]. This is known as a ‘grey-box model’, where the system of capacitances is identified based on prior knowledge of the physical structure of the dwelling, and the parameters for the individual network components are determined based on calibrating against observed behaviour of thermal dynamics of the building that is to be represented by the reduced-order model [53]. Models based on three capacitances have been shown to offer considerable improvement in terms of accuracy over those based on two capacitances, while four capacitances offer less of a step-change in accuracy, and yet have increased computational burden. Three capacitances was therefore felt to be a suitable compromise and is therefore equivalent to models such as those of Cooper [57,64] and Lauster [51]. The accuracy of the model is analysed in Section 5.

Three buildings representing typical UK building types (detached, semi-detached, terraced) were selected based on work conducted as part of the SUPERGEN HiDEF project [65], and the ESP-r software was used to generate calibration data. Two insulation levels for each building type were chosen (the default and an ‘improved’ level of insulation), resulting in six sets of parameters. These sets of parameters are available at run-time.
and specified by the user to characterise the thermal properties of the dwellings to be simulated. The emitter heat transfer coefficient is calculated as the value required to maintain an interior temperature of $20^\circ C$ with an external temperature of $0^\circ C$ and emitter temperature of $50^\circ C$, as in [14].

The external air temperature is provided as an output of the temperature model. Heat gains within the building are provided as outputs from the heating system model, solar irradiance from the irradiance model, and casual gains from the occupancy and electrical demand models. Casual gains are calculated the same as described in [14], with the exception that the four-state occupancy model used here allows the number of dormant and active occupants to be specified explicitly rather than assumed.

3.5.2. Hot water cylinder model

The electrical circuit analogue of the hot water cylinder model is shown in Fig. 5 and is based on a single thermal mass, $C_{b}$, as in [57,64]. Hot water demand is represented by a variable heat transfer coefficient $H_{\text{demand}}$ which is calculated from the output of the hot water demand model. The cold water inlet temperature is assumed to be a constant $10^\circ C$. Heat gains to the cylinder are

---

Fig. 3. The structure of the thermal demand model.

Fig. 4. Building thermal model.
provided by the heating system and solar thermal collector models. The hot water cylinder thermal capacitance is simply based on its volume, which is taken as 125 l. When representing a combi boiler system, the cylinder volume is set to a low value (e.g. 5 l). The authors note that the modelling of a hot water cylinder using a single capacitance is a simplification of reality and that for complex processes such as hot water stratification to be modelled a more complex model should be used.

For clarity, the hot water cylinder sub-model has been shown separately from the building thermal sub-model, but their solution is fully integrated: the hot water cylinder is subjected to thermal losses to the indoor air temperature node, which is part of the building thermal model above. In the model, therefore, the circuits are treated and solved as one.

The electrical circuit analogue allows the specification of a set of first-order differential equations (see Appendix A) which can be solved to find the variables of interest [50], in this case the temperatures of the building, indoor air, emitter, and cylinder nodes. In the model, Euler’s method is used to calculate the temperatures as a function of the same temperatures in the previous time step, the outside air temperature, heat gains, and the heat transfer coefficients and thermal capacitances appropriate for the specific dwelling being modelled.

3.5.3. Heating controls

The heating controls model represents a typical arrangement of thermostats and timers, as shown in Fig. 6. The outputs control the operation of the boiler, heating system pump, and whether heat is used for space heating or DHW. The heat emitter temperature serves as a proxy for the boiler return temperature. The thermostat set points are assigned stochastically to each dwelling based on empirical distributions for indoor air temperature [66] and hot water delivery temperature [67]. Deadbands are assigned to the thermostats: 5 °C for the emitter and hot water cylinder, and 2 °C for space.

Timer settings for space heating are based on empirical distributions for the probability of space heating being switched on for weekdays and weekends [66]. A first order Markov chain technique is used to stochastically assign sequences of timer settings such that over a large number of runs the timer state probabilities match the empirical probability distributions. This was achieved by deriving transition probability matrices between timer states for each half-hour based on the state probabilities given by the empirical probability distributions and a ‘parameter of mobility’ as described in [68]. A constant parameter of mobility of 0.25 was chosen, which results in an average number of ‘on’ hours for timers of 9.86 h and 10.12 h for weekdays and weekends respectively. This is close to the average of 10 h reported for both weekdays and weekends [66]. Average durations for ‘on’ sequences were 2.56 h for weekdays and 2.62 h for weekends. The resulting sequences of timer settings are randomised by ±15 min to desynchronise the timer-clocks of individual dwellings. The authors note that this approach is a simplification of reality due to data limitations, for example by assuming that the entire dwelling is heated as a single zone to a single temperature set-point. A more realistic approach to heating control settings would be based on the stochastic output from the occupancy model, and would account for multiple zones, as for example is implemented in the IDEAS and StROBE models [22], provided appropriate data was available.

3.5.4. Heating system – gas boilers

The heating control variables are used to determine the operation of the gas boiler and how its output is distributed between space and water heating, as shown in Fig. 7. The gas flow rate, lower calorific value of natural gas and boiler thermal efficiency determine the boiler’s net heat output. Gas flow rates are based
on typical values from boiler manufacturer datasheets [69], while the thermal efficiency is assumed 75% to reflect the average boiler efficiency in the UK [70]. The lower calorific value of natural gas was taken as 40 MJ/m$^3$.

The model does not include any other use of gas. The most obvious exclusion is gas cooking. This exclusion however will have a relatively small impact on the gas consumption of the dwelling as its effect will be reduced contribution to casual heat gains from gas cooking, and which will be compensated by increased usage of the boiler with corresponding increased gas usage. In the model, therefore, the gas that would have been used in cooking will instead be used in the boiler to heat the internal space (at least during the heating season). We also note that gas usage related to cooking is small compared to its usage for heating.

3.6. Solar and external temperature sub-models

The integrated model incorporates solar photovoltaics (PV) and a solar thermal collector model, both of which are based on an irradiance model that produces one-minute resolution stochastic irradiance data using a first-order Markov chain technique based on historic irradiance data from Loughborough. These are all described fully in previous publications [4,18].

A new addition to the model is the external temperature sub-model that produces stochastic one-minute resolution external air temperature data which is based on historic temperature data and which is appropriately correlated with the one-minute resolution irradiance data. The temperature model has been developed in order to maintain the self-contained nature of the model. The

---

**Fig. 6.** Heating controls.

**Fig. 7.** Block diagram for regular gas boiler heating system.
model can however be readily adapted as required by the user such that it uses weather data from a variety of sources e.g. Meteonorm.

The external temperature is simulated in two steps. First, an average temperature for the day is simulated, and then subsequently this is used as a basis for constructing the profile of how the temperature deviates around this average throughout the day. The average daily temperature is generated based on the long-term average daily temperature for central England [71] for the day of year selected by the user, combined with a stochastic deviation around the average to add an appropriate amount of randomness.

Starting from this average daily temperature, the second step is to construct the temperature profile over the course of the day. This is done by first assigning a maximum and minimum temperature for the day which is made to be proportional to the total cumulative irradiance for the day, so overcast days will tend to have fairly a flat temperature profile around the average, while sunny days will exhibit more deviation. The profile is then shaped in a piece-wise fashion. During the hours of daylight, the temperature changes are proportional to the incident irradiance, while outside of daylight hours the temperature changes are dependent on the clearness index, so that for example temperatures will fall slower at night when it is overcast. This shaping of the temperature profile provides a degree of correlation between irradiance and air temperature variables, but note that this has been implemented so that the average temperature for the day is unchanged.

4. Correlation of output variables: simulation example

A critical feature of the CREST demand model is that the operation of its sub-models is appropriately correlated with respect to time. This is demonstrated in Fig. 8, which shows an example of the output of the model for a winter day for a 2-person detached dwelling with a regular gas boiler, a hot water cylinder, PV and a solar thermal collector system. Fig. 8 (top) shows the simulated occupancy profiles for the dwelling. The combined occupancy state for the dwelling is shown (how many residents are at home), the activity state is shown (how many are active), and the combined ‘active occupancy’ state is shown (the logical AND of the occupancy state and the activity state). The residents are at home and asleep at night, wake and leave the dwelling in the morning, return home in the late afternoon, and fall asleep in the early evening. Fig. 8 (upper middle) shows the power flows associated with the electrical components of the model. Appliance and lighting use is appropriately correlated with the periods when the occupants are at home and active. The PV system produces power during the hours of daylight, with a maximum in the middle of the day, and periods of reduced output throughout the day due to simulated cloud cover.

Fig. 8 (lower middle) shows the simulated temperatures for the dwelling, while Fig. 8 (bottom) shows the associated heat flows. The hot water cylinder temperature gradually reduces at night due to thermal losses, followed by more rapid reductions due to

---

![Figure 8](https://via.placeholder.com/150)

**Fig. 8.** Example simulation output for a single day in winter.
hot water draws soon after the residents awake. When the hot water thermostat set point is reached, the boiler switches on to provide heat to the cylinder. Further domestic hot water related boiler firing can be seen in the evening. Note that the model captures the time-coincidence of active-occupancy related thermal and electrical demands.

Returning to the temperature graph (lower middle), the solar thermal collector starts at the same temperature as the outside air, and heats up with incident irradiance. When the collector temperature exceeds the hot water cylinder temperature, the pump activates and transfers heat from the collector to the cylinder. The building and indoor air cools throughout the night. No heat is provided to the space by the boiler until the timer switches on in the morning. Subsequently, the boiler fires at regular intervals to maintain the indoor air temperature. The emitters heat up quickly, with the indoor air and building thermal masses showing delayed thermal response as appropriate. Casual heat gains follow the pattern of occupancy, appliances and lighting components, while the passive solar gain has the same pattern as the PV output. Again note the model captures the appropriate time correlation between passive solar gain, and the output of the collector and PV.

5. Validation of the model

To validate the model, the output of the thermal and hot water sub-models are compared against independent data sets which were not used in their calibration. Simulated gas demands are compared with data from the Energy Demand Research Project Early Smart Meter Trials [72] and the gas-boiler control group of the Carbon Trust Micro-CHP Accelerator [73]. Hot water demands are compared with the data from the Energy Saving Trust Measurement of Domestic Hot Water Consumption in Dwellings [67]. These will be referred to as “EDRP”, “Carbon Trust”, and “EST” datasets respectively. The electrical components were validated previously [7], with particular emphasis on validating the demand diversity, and are unchanged.

5.1. Building thermal dynamics

The accuracy of the identification and parameterisation of the building thermal model is assessed by comparing the output of the model with calibration data generated by the building thermal simulation package ESP-r. The calibration data consists of one-minute resolution data including ambient temperature, indoor temperature, and heating gains for each building type over 11 days in January. The dwelling is heated to 20°C between the hours of 8 am and 10 pm. A winter dataset with heating was chosen as it has been shown that calibration data without heating can result in unreliable parameterisation [52].

An example comparing the output of the model to the calibration data for one of the buildings is shown in Fig. 9 while Table 3 summarises the root mean squared errors (RMSE) and mean bias errors (MBE) for all building types. The ‘simulation’ values are the errors between the calibration data indoor air temperature and the modelled indoor air temperature given the same heating gains as the calibration data. The ‘one-step prediction’ shows the error where the model is used to predict the temperatures in the next time step in the calibration data. The mean bias errors are small and indicate that the model does not deviate from the calibration data over long time periods, a problem that has been noted in reduced-order grey-box models [52]. Root mean squared errors are of the same order of magnitude as can be expected for low-order models of this type and are acceptable for the purposes of the model. The authors note that the model does not capture all of the building thermal dynamics, for example the model does not achieve one-step prediction residual errors that pass the ‘white noise’ test, as described in [52,53]. While this is of limited relevance for the intended purpose of modelling larger aggregations of dwellings, it indicates that the model should not be expected to be used as a replacement for a detailed building thermal simulation package where a single specific dwelling is intended to be modelled at a high level of accuracy.

5.2. Aggregated demand profiles and totals

Fig. 10 compares the gas data in terms of daily profiles and cumulative distributions of daily totals. Seasonal variation is shown by comparing winter (upper graphs) and summer (lower graphs) data separately. The EDRP data consists of half-hourly gas smart meter data for 8062 dwellings for the months of January and July over three years, corresponding to 65 days for January and 93 days for July. The winter Carbon Trust data is at 5-minute resolution and for 18 dwellings for the months of December, January and February for the years 2006 and 2007 (90 days in total), while the summer data is for 20 dwellings for the months June, July, and August for the same years (68 days in total). The EDRP data can be considered reasonably representative of the UK average. The Carbon Trust data cannot be considered nationally-representative, in particular it will be characterised by the presence of new, high-efficiency gas boilers, but was included due to the relatively high-resolution of the data, to compare the ‘spikiness’ of the data. The model data consists of 104 dwellings for seven days in January and July (five weekdays and two weekends each) for a total of 728 dwelling-days each – this is the maximum number of dwellings that MS Excel spreadsheets can hold the high-resolution

![Fig. 9. Example comparison of calibration data and model.](image-url)
disaggregated data for. The number of occupants was stochastically assigned according to the distribution in the UK time-use survey [60]. The building-type was equally likely for the three default types included in the model (no ‘improved’ insulation buildings were included). The probability of boiler-type was the same as in the EST data (64% regular, 36% combi).

In general, Fig. 10 illustrates that the model captures much of the timing of heating demand, but results in lower overall demand than the validation datasets. Average daily gas demand for summer is 12.1 kW h, 15.2 kW h, and 15.7 kW h for the ERDP, Carbon Trust and model respectively, while for winter the average daily gas demands are 101.9 kW h, 75.5 kW h and 54.0 kW h. Evidently the

---

**Table 3**

Summary of root mean squared errors (RMSE) and mean bias errors (MBE) between modelled indoor air temperature and calibration data for simulation and one-step prediction.

<table>
<thead>
<tr>
<th>Building number</th>
<th>Simulation RMSE (°C)</th>
<th>Simulation MBE (°C)</th>
<th>One-step prediction RMSE (°C)</th>
<th>One-step prediction MBE (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.03</td>
<td>2.14E-02</td>
<td>0.09</td>
<td>5.73E-04</td>
</tr>
<tr>
<td>2</td>
<td>0.88</td>
<td>-3.40E-02</td>
<td>0.08</td>
<td>-7.81E-04</td>
</tr>
<tr>
<td>3</td>
<td>0.71</td>
<td>1.78E-02</td>
<td>0.08</td>
<td>9.19E-04</td>
</tr>
<tr>
<td>4</td>
<td>0.85</td>
<td>-2.33E-02</td>
<td>0.09</td>
<td>-8.05E-04</td>
</tr>
<tr>
<td>5</td>
<td>0.96</td>
<td>-7.76E-03</td>
<td>0.12</td>
<td>-2.00E-04</td>
</tr>
<tr>
<td>6</td>
<td>0.85</td>
<td>2.46E-02</td>
<td>0.11</td>
<td>6.62E-04</td>
</tr>
</tbody>
</table>

---

Fig. 10. Validation of the model’s gas demands.
six building types, in the arbitrary proportions included here, are more thermally efficient than the UK national average. We note that the model output could readily be adjusted to provide a better match, by scaling the building thermal parameters, as in [15], however this has not been done here as it is not a ‘grey-box’ parameterisation technique.

Returning to Fig. 10a and c, in terms of daily profiles, the model captures the characteristic morning peak in gas demand, with a slight under-representation of the initial ramp up. The evening peak is under-represented. There are a number of possible explanations. The model space heating is determined by timer settings based on a study of 248 homes [66]. If the heating patterns of this sample are different from those of the validation datasets then a discrepancy in timing will result. A second possible explanation is that the building types used in the model could have a higher thermal mass than found in the validation datasets (e.g. 33% will be detached dwellings), which could result in reduced thermal demands in the evening.

Comparing the distribution of gas demands across dwellings, the model appears to over-represent the average and under-represent the extremes. This is to be expected, however, as the construction of the model is inherently based on averages, with diversity introduced through stochastic techniques, which tends to result in an under-representation of the extremes [17]. Furthermore, the model has a limited number of input building types, which does not cover the full range to be found in the EDRP data.

The model’s hot water demand output is in broad agreement with the EST validation data as shown in Fig. 11. The model dataset is the same as the summer dataset (see Section 5.1), and shows hot water draws at one minute resolution. The EST data consists of measured hot water draws for 107 dwellings averaged over each hour of the day. The average is 122.4 l/day for the EST data and 117.5 l/day for the model output. The average for the model winter dataset (not shown) is 113.7 l/day. The daily profiles are similar, with the model capturing both morning and evening peaks. The model noticeably underrepresents the proportion of dwellings with very high hot water usage (>300 l/day).

5.3. Gas demand – spikiness

An important requirement of the model is that it should properly represent the short duration peaks that are characteristic of the undiversified energy demand of individual dwellings. We call this ‘spikiness’. It is important, for example, in the sizing of ‘service’ connections (the cables or pipes that go from the ‘mains’ that run along the street to the individual dwellings). It is also important in the calculation of resistive energy losses and voltage drops in electricity cables or of pressure drops in gas pipes. Furthermore, if the model is to be extended in the future to consider the use of heat pumps, then the spikiness of the thermal demand will become significant in the electricity demand. The use of buffer tanks or other thermal storage, which may be more prevalent with heat pumps, will affect this spikiness, but the aim in the current work is to ensure that the spikiness of the underlying thermal demand is properly represented. With this in mind, the initial modelling of gas boilers, with minimal thermal storage, provides a good basis for this verification.

![Graph of Domestic Hot Water Profiles](image1)

![Graph of Domestic Hot Water Totals](image2)

Fig. 11. Validation of the model’s hot water demands.
Fig. 12 (top) shows an example of real gas demand for a single dwelling from the Carbon Trust data over a period of five days in December and compares this with simulated data (bottom). While there appears to be spikiness present in the example of model output, it is compared more thoroughly in Fig. 13 which compares cumulative distributions of changes in gas demand in each time step for the model and the Carbon Trust (winter) datasets. To provide a fair comparison with the Carbon Trust data, the model output was averaged over five minutes. Distributions are shown for single dwellings and for a six-dwelling aggregation. In general, the model broadly captures the shape of the distribution, with both data sets characterised by a high proportion of time with small changes in demand, and a considerable proportion with large changes. Compared to the validation data, the model over-represents large changes in gas demand, and under-represents small changes. This discrepancy can be explained by the fact the building types included in the model are more thermally efficient than those in the Carbon Trust data, and that the heating system model is relatively simple and does not include a requirement for minimum boiler run times. Both of these would result in the model exhibiting more cycling and therefore a higher proportion of time with large changes in gas demand, which is confirmed by Fig. 13. Again, we note that we are not expecting an exact match with the independent validation dataset. The six dwelling distributions are more in agreement and show that the model produces better results for aggregations of dwellings than for individual dwellings. This is adequate for the intended purposes of the model, and again we note that the model is not intended as a replacement for detailed models that simulate a single specific dwelling to high accuracy.

5.4. Simultaneity factors

To assess the ability of the model to capture the appropriate diversity of demand at the level of the low-voltage distribution network, this section analyses the simultaneity factor of the model output of varying aggregations of dwellings. The simultaneity factor for a group of dwellings is the peak demand of the aggregated group’s demand divided by the sum of the peak demands of the individual dwellings. Fig. 14 shows the simultaneity factors for electricity, gas and domestic hot water for the January model dataset for aggregations of dwellings from 2 to 30. Each value of dwelling aggregation shows a number of data points. The dataset consists of 728 dwelling-days, which results in 364 sets of 2 dwelling aggregations, reducing to 24 sets of 30 dwelling aggregations. The results for electricity, gas and water show similar trends. Smaller aggregations of dwellings have higher and greater spread of simultaneity factors than larger aggregations. For 30 dwellings, the average simultaneity factors are 0.21 for electricity, 0.26 for gas and 0.13 for domestic hot water. For an equivalent number of dwellings, Baetens & Saelens report similar simultaneity factors for simulated Belgian dwellings with the StROBE model of 0.26 and 0.13 for electricity and domestic hot water respectively [22]. Finally we note that the maximum electricity simultaneity factor for 30 dwellings results in an after diversity maximum demand of 2.17 kW for the model dataset, which shows the model tending towards the typical UK standard after diversity maximum demand of 2 kW. Generally, it can be said that the model can represent the diversity of high-resolution demand for the purposes of low-voltage distribution network analysis.
6. Conclusions

This paper has presented a major new version of CREST’s high-resolution stochastic energy demand model. As previously, it has a bottom-up activity-based structure to provide high temporal resolution and uses stochastic programming techniques to appropriately represent the diversity of demand. The major addition is the thermal model which has been validated against three independent datasets. The key feature is that the thermal model is fully integrated with the electrical model thus ensuring appropriate time correlation with dwelling occupancy. This is of critical importance for the model’s application to low-voltage network analysis and the electrification of heating. The model is again open-source, allowing users to inspect every internal detail and to modify or extend its operation for their own specific application.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council, UK, within the Transformation of the Top and Tail of Energy Networks project (EP/I031707/1).

Appendix A

Set of first-order differential equations used in the model to simulate building thermal dynamics:

\[
\begin{align*}
C_b \dot{\theta}_b &= -(H_{sb} + H_{eb}) \theta_b + H_{sb} \theta_i + H_{eb} \theta_o \\
C_i \dot{\theta}_i &= H_{bi} \theta_b - (H_s + H_{bi} + H_{em} + H_{loss}) \theta_i + H_o \theta_o + H_{loss} \theta_{cyl} + \phi_s + \phi_c \\
C_{em} \dot{\theta}_{em} &= H_{em} \theta_i - H_{em} \theta_{em} + \phi_{space} \\
C_{cyl} \dot{\theta}_{cyl} &= H_{loss} \theta_i - (H_{loss} + H_{dhw}) \theta_{cyl} + H_{dhw} \theta_{cw} + \phi_{water} + \phi_{collector}
\end{align*}
\]

References


