A combined engineering and statistical model of UK domestic appliance electrical load profiles

Conference or Workshop Item

Accepted Version


It is advisable to refer to the publisher’s version if you intend to cite from the work.

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur
CentAUR

Central Archive at the University of Reading

Reading's research outputs online
A combined engineering and statistical model of UK domestic appliance electrical load profiles

David Lim*, Runming Yao

School of Construction Management and Engineering, University of Reading, Whiteknights PO Box 219, Reading, UK

Abstract

The development of a combined engineering and statistical Artificial Neural Network model of UK domestic appliance load profiles is presented. The model uses diary-style appliance use data and a survey questionnaire collected from 51 suburban households and 46 rural households during the summer of 2010 and 2011 respectively. It also incorporates measured energy data and is sensitive to socioeconomic, physical dwelling and temperature variables. A prototype model is constructed in MATLAB using a two layer feed forward network with backpropagation training which has a 12:10:24 architecture. Model outputs include appliance load profiles which can be applied to the fields of energy planning (microrenewables and smart grids), building simulation tools and energy policy.

Keywords: Artificial Neural Networks, Domestic appliance energy consumption, Load modeling

*Corresponding author: Tel: +44 (0) 118 378 7182; E-mail address: d.v.lim@reading.ac.uk

1. Introduction

Humanity arguably faces its greatest challenge to date: climate change (IPCC, 2007, IPCC, 2007a). Under the Kyoto Protocol (UNFCCC, 2009), the European Union agreed to a collective emission reduction of 8% with the UK’s target set at 12.5% in the period 2008 to 2012. Embracing the challenge, the UK set an internal emission reduction target of by 20% by 2010 from a 1990 baseline. Having passed the Climate Change Act in 2008 (DECC, 2008), long term targets now stand at 80% by 2050. In 2008, UK energy consumption by sector showed transport 38%, industry 20%, domestic 29% and other 13% (DECC, 2009). Disaggregating domestic consumption shows: space heating 56%, hot water 26%, lighting and appliances 15% and cooking 3% (DECC, 2009a). Consequently, in 2008 approximately 27.2 MtCO$_2$(e) was produced from lighting and appliances, clearly presenting significant scope for emission reductions in this sector.

Since 1970, appliance ownership and associated energy consumption has increased at an extraordinary rate, so much so that Lomas et al. (2007) state that the use of energy in UK homes for lighting and electrical equipment is increasing and can exceed that used for space heating. One of the UK’s leading organisations on carbon reduction, the Energy Saving Trust (EST), draws attention to consumer electronics, stating that it is the single most significant growth area of electricity consumption within the home (Owen, 2007), and that by 2020, entertainment, computers and gadgets will account for an extraordinary 45% of electricity used in the home. Furthermore, consumer electronics in particular will be the biggest single user of domestic electricity, overtaking the traditionally high consuming sectors of cold appliances and lighting. This phenomena is driven by factors such as: increased income (Roberts, 2008), decreasing household size (DCLG, 2007), consumer trends towards ‘bigger and better’ (Crosbie, 2008, Boardman et al., 2005) and an ever increasing product range (Owen, 2006). Electrical load profile analysis at household level will become increasingly important and essential for future energy planning, particularly for integrating microrenewables (Wright and Firth, 2007), sizing decentralized power plants and demand side management programmes (Paatero and Lund, 2005). Further work is required to improve our understanding of domestic electricity consumption to assist decision makers with climate change mitigation strategy.

The aim of this study is to develop a prototype model which can predict a dwelling’s diurnal appliance energy load profile for a given demographic, socioeconomic, physical and climatic characteristic. The model is
primarily developed for engineers and academics who can apply it to the following fields: (1) scenario modelling; (2) appliance efficiency evaluation; (3) building simulation tools; (4) energy policy (including demand-side management programmes); (5) supply and demand issues (micro renewables/energy planning); and (6) Micro grids and Smart grids. The model’s defining feature is its ability to capture product efficiency improvement by integrating a data pre-processing stage which uses a simple engineering method.

1.1 Modelling domestic energy consumption

Parameterization of domestic energy consumption within the literature encompasses social, economic, behavioural and physical factors. Modelling domestic energy use is situational; therefore the order of parameter significance depends on the characteristics of the modelled scenario. Different studies conclude of varying parameter significance, however commonly cited parameters at a micro level include: household size, occupancy pattern, income, floor area, dwelling type/vintage, location, climate, tenure and appliance ownership amongst others. Challenges concerning data procurement severely restrict the inclusivity of domestic energy models. Furthermore, energy modellers are often mindful of the balance which must be struck between computational efficiency versus accuracy. This can frequently result in limited parameter inclusivity. The model’s inputs reflect the most salient variables of domestic energy consumption as identified in the literature and uses Artificial Neural Networks to simulate load profiles.

1.2 Energy modeling techniques

Energy models can be classified using various criteria from purpose, to data resolution, to mathematical approach. Nevertheless, two general approaches are formally recognized - ‘top-down’ and ‘bottom-up’. Top-down models commonly use national aggregated data, adopt econometric or technological approaches and are generally used for supply-side issues. Bottom-up models require high resolution data, use statistical or engineering techniques aimed at estimating aggregated consumption and are widely used for demand-side analysis (Swan and Ugursal, 2009). Engineering methods include distributions, archetypes and samples (Swan and Ugursal, 2009), whilst statistical methods commonly feature regression techniques, Conditional Demand Analysis (CDA), Markov Chain (MC) and Artificial Neural Networks (ANNs) amongst others. Intuitively, each method displays a different set of advantages and disadvantages. Although it is beyond the scope of this study to provide a full discussion on such methods, it is worth noting that artificial neural networks are commonly used for load modelling and the literature demonstrates their superior capability over conventional methods such as regression, engineering and time series (Karatasou et al., 2006) when applied to multivariate modeling.

1.3 Artificial Neural Networks (ANNs)

ANNs is a biomimetic technique proposed in 1943 by McCulloch and Pitts (1943). The concept was inspired by the anatomy of the human brain and consists of a network of neurons or processing units arranged in layers. Commonly used networks for short term load forecasting (STLF) are ‘multi-layer perceptrons’ (MLPs) (Beccali et al., 2004). MLPs have an input layer, one or more hidden layers and an output layer. They are generally classified as ‘feedforward’ networks as there are no feedback connections. The generic network form can be seen in Fig.1 below.

![Generic form of a MLP](image)

Figure 1. Generic form of a MLP (Beccali et al. 2004)

Each neuron has input weights, a transfer and activation function and an output. Networks are trained to recognize data patterns within ‘input output pairs’ facilitated by a training algorithm. Equation 1 below expresses the output of a neuron:

$$ y_j = f \left( \sum_{i=1}^{n} w_{ji} x_i - b \right) $$

Equation 1

From the anatomy of a neuron, the network is arranged in layers, with the input layer connecting to the hidden layer to the output layer.
where: $y_j$ is the output of generic neuron $y$ belonging to layer $j$; $x_i$ are the input signals to the neuron, $w_{ij}$ is the synaptic weight associated with the connection between the generic neurons belonging to layers $j$ and $i$ respectively; $b$ is the bias term (another neuron weight); and $f$ is the activity function.

ANNs are renowned for their accuracy in reproducing lifelike characteristics of energy use and capturing the effects of occupant behaviour (Kreider et al., 1995). Fischer (2010) states that: ‘over the past ten years, the use of ANNs has become a benchmark for comparing new emerging methods’. Park et al. (1991) initially used ANNs for domestic load forecasting in the early nineties, achieving a 4% error when predicting demand 24 hours ahead of time. Two years later, ASHRAE initiated an energy modelling competition in 1993 which involved predicting energy consumption of a single building; the top six models used ANNs, suggesting the technique is superior to regression methods (Kreider and Harberl, 1994). Considering the aforementioned, the selection of ANNs is regarded as a logical selection in which to develop the model in order to reproduce the required features.

2. Data sets, inputs and data pre-processing

The network requires ‘input output data pairs’ to effectively train it in a supervised manner. Inputs represent determinants of energy consumption whilst the outputs correspond to the hourly appliance loads. This data is acquired through field survey and measurement, and is processed before presenting it to the network.

2.1 Data sources

The model uses four main sources of quantitative data: (1) a questionnaire; (2) diary-style appliance use data; (3) weather data; and (4) measured appliance energy data.

2.1.1 Questionnaire

A total of 51 households in the urban area of Greater London and 46 households in the rural region of Dorset, South West England were invited to participate in a two-part survey during the summer of 2010 and 2011 respectively. The first part of the survey consisted of a single questionnaire used to establish household characteristics. Such characteristics included demographic and socioeconomic factors, physical dwelling characteristics, appliance ownership and occupancy patterns. Respondents could complete the survey online, via a dedicated survey website (www.appliancesurvey.org) or by post.

2.1.2 Appliance energy-use diary

The second part of the survey involved respondents recording their appliance use for a typical week (reflecting normal occupancy patterns) in summer by completing ‘appliance-diary-sheets’ (Figure 2). The sheet designs were tailored to the respondent’s appliance ownership type and batched according to their property layout. When an appliance was used, respondents indicated the appliance type, duration of use, time and household member. For kettle, washing machine and oven loads, respondents also had to indicate if the appliance was operating at ‘full’ or ‘half’ capacity.

Between three and six batches of sheets were issued to every household; the intention being that one batch is to be placed in each main area of their home. This approach was implemented to increase the likelihood of respondent engagement in order to reduce unrecorded events.

2.1.3 Weather data

Weather is highlighted in the literature as being an influencing, or significant factor of domestic energy consumption. This is particularly true for space heating demand but less so for appliance energy use.

2.1.4 Measured appliance energy data

Appliance loads can be placed into four categories: (1) continuous loads (e.g. alarm clock), (2) operational constant loads (e.g. kettle), (3) operational variable loads (e.g. washing machine) and (4) cyclical loads (e.g.
refrigerator). Calculating the energy consumption of continuous, operational constant and cyclical loads is relatively straightforward. However, power variation can occur due to appliance manufacturer specification (e.g. power ratings, capacities etc.) and other associated variables such as age, condition and environmental factors. To handle this, the model randomly selects an appliance power rating from a ‘power rating database’ constructed for commonly used appliances. The power rating database was composed by empirical measurement using two types of instrument: (1) Plogg Zgb automatic energy data logger and (2) Brennenstahl EM230 energy monitor.

Operational variable loads on the other hand are more complex to calculate, due to many factors which significantly affect energy consumption. For example, washing machine loads depend on type of wash programme used, washing capacity, hot or cold fill and the make and model of machine. Again, the model captures a degree of variation by randomly selecting an energy load profile from a second database, the ‘operational variable load database’, which accounts for some of the factors mentioned above.

### 2.2 Model inputs

The selection of model inputs is largely determined by the literature. Although twelve variables have been selected, it is noted that other less significant variables are highlighted within the literature, such as: education level, dwelling vintage, occupant gender, humidity and solar radiance. However, a trade-off is usually required; model complexity versus accuracy (and challenges in data collection). Table 1 shows the model inputs (variables) and their associated factors. It can be seen that each variable has a corresponding neuron number, value and possible coding scheme.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Input Neuron</th>
<th>Input</th>
<th>Value</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household size</td>
<td>1</td>
<td>Senior citizen (60+)</td>
<td>Number of people</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Adults</td>
<td>Number of people</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Children (3 - 17)</td>
<td>Number of children</td>
<td>n/a</td>
</tr>
<tr>
<td>Occupant presence</td>
<td>4</td>
<td>Total occupant hours at home</td>
<td>0 - 1</td>
<td>n/a</td>
</tr>
<tr>
<td>Income</td>
<td>5</td>
<td>Total household income</td>
<td>0 - 1</td>
<td>n/a</td>
</tr>
<tr>
<td>Household employment</td>
<td>6</td>
<td>Employment ratio</td>
<td>0 - 1</td>
<td>0 = none employed, 1 = all members employed</td>
</tr>
<tr>
<td>Dwelling size</td>
<td>7</td>
<td>Total usable floor area (m²)</td>
<td>0 - 1</td>
<td>n/a</td>
</tr>
<tr>
<td>Appliance ownership</td>
<td>8</td>
<td>Appliance ownership level</td>
<td>1 – 10</td>
<td>1 = low level ownership, 10 = high level of ownership</td>
</tr>
<tr>
<td>Dwelling type</td>
<td>9</td>
<td>Dwelling type</td>
<td>1 – 6</td>
<td>1-6 = detached, semi-detached, terraced, flat, detached</td>
</tr>
<tr>
<td>Tenure</td>
<td>10</td>
<td>Tenure type</td>
<td>1 – 3</td>
<td>1-3 = home owner, renting, social housing</td>
</tr>
<tr>
<td>Geographic location</td>
<td>11</td>
<td>Location type</td>
<td>1 – 5</td>
<td>1=inner city, 2=urban, 3=suburban, 4=rural, 5=remote</td>
</tr>
<tr>
<td>Weather</td>
<td>12</td>
<td>Mean external temperature</td>
<td>Degrees Celsius</td>
<td>n/a</td>
</tr>
</tbody>
</table>

### 2.3 Data preprocessing

As the neural network is trained using a supervised learning method, output values or ‘targets’, must be assigned to input values to form a ‘data pair’. The targets are calculated during a data pre-processing stage which uses a simple engineering method. The targets denote the hourly aggregated appliance load values. The aggregated hourly load is a sum of the minutely component loads (e.g. kettle + toaster + fridge etc.) which are calculated from selecting random values within the ‘power rating’ and ‘operational variable load’ databases described in section 2.1.2. This introduces a stochastic element to capture variation of behavioural factors and appliance specification. In doing so, notably, this process enables the modelling of: (1) appliance efficiency improvements; (2) appliance energy policy; and (3) energy trends.

Modelling these kinds of phenomena is an indicative limitation of neural network load models and is typical advantage widely reserved for engineering models.

### 2.4 Data normalization

To improve the computational efficiency of the network, input data can be scaled to an interval of -1 to +1 or 0 to 1. This study uses the scaling interval of 0 to 1 for continuous data, using equation 2 below.

$$ x_n = \left( \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \right) $$  \hspace{1cm} (2)$$

where:
- $x$ = value of the input/output unit
- $x_n$ = value of the scaled input/output unit
- $x_{\text{min}}$ = minimum value of the input/output unit
- $x_{\text{max}}$ = maximum value of the input/output unit
2.5 Data division

Data used for neural network development is commonly divided into two sets - a training set (for training) and a test set (for testing). Various authors have adopted a proportional split of 75%, 25%, training set and test set respectively (Aydinalp et al. 2002; Yang et al., 2005) and others a 80%, 20% split (Beccali et al. 2004). As the model is implemented in MATLAB R2011b, the data division default proportion is used. The default proportions are 75% training set, 15% test set and 15% validation set.

2.6 Network architecture

The model assumes a 12:10:24 architecture: 12 neurons feature in the input layer, ten in the hidden layer and 24 neurons for the output layer. The number of hidden neurons has no physical significance and can be arbitrarily increased or decreased to affect the network’s performance. Too many hidden layers can result in ‘overfitting’, whereas too few can reduce the network’s ability to map the target outputs (Cohen and Krarti, 1995). This is often a matter of trial and error when striving to reduce the model error whilst trying to maintain computational efficiency.

2.7 Selection of activation functions

This study looks at using the universal function approximators, ‘tan sigmoid’ and ‘linear’ activation functions in the hidden and output layer respectively. Such a configuration is widely used in energy modeling and is reported to be the most commonly used architecture (Tso & Yau, 2007).

3. Model development and applications

This section presents the model in schematic form, and discusses its applications in the real world.

3.1 Model schematic

Figure 3 below shows the model in schematic form, illustrating its main components and the various processes involved in data collection and data treatment.

Figure 3. Model schematic

Appliance load profiles form the basis of the outputs which can then be analyzed in various contexts as identified in the above figure and later discussed in section 3.3.

3.2 Artificial Neural Network Model

The prototype model was implemented in MATLAB R2011b using the neural network toolbox on a high specification laptop computer (Intel Core i7 processor @
1.6GHz, 16GB RAM). Figure 4 shows the neural network model illustrating the 12:10:24 architecture with each neuron connected to the proceeding layer. Network training was stopped after 100 epochs using 75% of the data which was randomly selected. The model achieves an overall R value of 0.933 which is considered very respectable for this level of modeling. Table 2 shows the mean squared error (MSE) and R values for the training, validation and test data sets. Figure 8 shows regression plots for all the data sets as seen in a MATLAB screen shot.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Mean Squared Error</th>
<th>R Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>2.706 e-3</td>
<td>0.998</td>
</tr>
<tr>
<td>Validation</td>
<td>5.566 e-3</td>
<td>0.711</td>
</tr>
<tr>
<td>Test</td>
<td>1.494 e-3</td>
<td>0.854</td>
</tr>
<tr>
<td>Overall</td>
<td>3.26 e-3</td>
<td>0.933</td>
</tr>
</tbody>
</table>

Table 2. Network performance

Figure 6 shows an example load profile for survey respondent number 22 generated during the data pre-processing stage. This data is subsequently simplified and averaged into hourly time steps to calculate the model targets used during network training. An example output can be seen in Figure 7. This particular profile corresponds to a low energy use household with single occupancy for weekday use.

Further model improvements will include experimentation with alternative network configurations, namely adjusting the number of neurons (in the hidden and output layer) and number of hidden layers. Increasing the number of neurons in the output layer will result in smaller time steps thus addressing issues concerned with time averaging effects (Wright & Firth, 2007). This is expected to be a key improvement area as specific appliance loads are associated with energy demand bursts. Using a similar experimental approach to Aydinalp et al. (2002) it is also worth applying different training algorithms and activation functions to help minimize the network error. Further work will also be conducted in developing data correction factors relating to the diary-style appliance use survey. Calculation of such factors will utilize measured energy data obtained from sub-metering commonly used appliances during survey periods to highlight the disparity between measured and recorded appliance uses. Unrecorded appliance use due to survey fatigue is an inherent feature of this type of survey exercise. Nevertheless, this can be partially addressed through introduction of such correction factors.

3.3 Appliance database development

Development of the operational variable load database was constructed using measured data obtained during the survey process. Figure 5 shows eight types of domestic washing machine load profiles (on a 40°C wash cycle) in the database. Energy consumption variation across appliance type, caused by manufacturer specification and user behavior (programs or settings), is significant; in some cases consumption could vary by a factor a five. The model’s functionality randomly selects a load profile from its database during data pre-processing stages (for the relevant appliance) and uses this to create aggregated hourly load totals that are then used as ‘targets’ in which
3.3 Model applications

This section briefly discusses how the model can be used by academics and energy planning engineers to whom it is originally intended for. Discussion allows for parallels to be drawn between the academic and industrial worlds where the model’s applications can provide valuable links.

3.3.1 Scenario modelling

The model can be used to predict the appliance loads of an existing neighbourhood or future housing developments based on a given demography, socioeconomic characteristics of households and basic physical properties of the dwellings. A range of scenarios can be simulated with respect to planning issues, financial constraints or other determining factors. This type of scenario modelling is vital for cost planning from the perspective of construction and utility provision.

3.3.2 Appliance efficiency
Modelling technological change is traditionally the domain of engineering models, however due to the model’s data pre-processing stage, evaluating the outcome of improving appliance efficiencies can be achieved. For example, this application is useful to predict the impact of introducing a new EU Energy Efficiency Rating band or prohibiting manufacture of all ‘B’ rated washing machines for instance.

### 3.3.3 Building simulation tools

The model can be used as a basis for a ‘bolt-on’ to building simulation tools such as ESP-r, IES<VE> or EnergyPlus to help improve the prediction capability of actual energy use in domestic schemes. At present, modellers are commonly faced with adopting crude algorithms embedded in software programmes in an attempt to simulate occupant appliance use. It is also common for simulation tools to ignore the energy contribution of appliances altogether. It is the intention to investigate how the model could function on a universal platform allowing accessibility from a range of existing tools in the form of a ‘toolbox’.

![Figure 8. Regression plots for training, testing and validation datasets](image)

### 3.3.4 Energy policy

The model can be used to assess the impact of energy policy measures on appliance use; energy rating schemes being a prime example. In addition, the impact of local housing policy can be evaluated. For example, if a particular policy focussed on developing low cost homes for a certain demographic (e.g. one or two person households with key worker occupants earning between £15K and £25K in southern England) the model will be able to estimate the specific contribution from appliances on the electricity network.

#### 3.3.5 Supply and demand

The model can be applied to problems concerning energy supply and demand matching. For instance, analysis of microrenewables supply curves (from photovoltaic systems or micro wind turbines) can be compared with appliance demand profiles to assess the extent to which carbon offsetting can be achieved. Simulated appliance load profiles can also be analyzed in the context of Dynamic Demand Management and general demand management programs to reduce peak demand.

#### 3.3.6 Micro grids and smart grids

Domestic load modeling is applicable to many fields; however analysis at a micro level, as demonstrated within this study, will be extremely beneficial for developing smart grids and micro grids which are very sensitive to load variability. By nature, appliance use encompasses a wide range of services and prediction of peak loads will become increasingly important when meeting demand with technology which generates electricity intermittently such as microturbines. Currently in the UK, much debate ensues around these two grid aspects as it pushes forward in its infancy. Reliable data at high resolution is sparse and modeling the impact removing neighborhood loads away from the main electricity grid will be crucial for successful grid integration.

### 4. Conclusions

The authors demonstrate the suitability of artificial neural networks to model diurnal domestic appliance loads. They present a two layer feedforward network using backpropogation training with architecture 12:10:24. The model’s achieves an R value of 0.932 and its distinguishing feature is its ability to assess the impact of improving appliance efficiency which are anticipated for the future; a feature widely reserved for engineering models. A discussion of the model’s application highlights the importance of work in this field, particularly at a micro level. Improvements to the model can be achieved with further experimentation of network
architecture, in particular, increasing the quantity of output neurons to decrease the load profile time step, and use of different activation functions and training algorithms. Furthermore, a larger and richer data set could be used to enhance the model’s output in order to make generalizations of appliance energy use in the UK; this work is already underway.

Acknowledgements

The authors would like thank the UK Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Grant Scheme, The Comino Foundation and The Royal Academy of Engineering for jointly funding this research.

References


