Matchstick: A Room-to-Room Thermal Model for Predicting Indoor Temperature from Wireless Sensor Data

Carl Ellis
School of Computing and Communications
Lancaster University
United Kingdom
c.ellis@comp.lancs.ac.uk

Mike Hazas
School of Computing and Communications
Lancaster University
United Kingdom
hazas@comp.lancs.ac.uk

James Scott
Microsoft Research Cambridge
United Kingdom
jws@microsoft.com

ABSTRACT

In this paper we present a room-to-room thermal model used to accurately predict temperatures in residential buildings. We evaluate the accuracy of this model with ground truth data from four occupied family homes (two in the UK and two in the US). The homes have differing construction and a range of heating infrastructure (wall-mounted radiators, underfloor heating, and furnace-driven forced-air). Data was gathered using a network of simple and sparse (one per room) temperature sensors, a gas meter sensor, and an outdoor temperature sensor. We show that our model can predict future indoor temperature trends with a 90^{th} percentile aggregate error between 0.61–1.50°C, when given boiler or furnace actuation times and outdoor temperature forecasts. Two existing models were also implemented and then evaluated on our dataset alongside Matchstick. As a proof of concept, we used data from a previous control study to show that when Matchstick is used to predict temperatures (rather than assuming a preset linear heating rate) the possible gas savings increase by up to 3%.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms

Keywords

Thermal Modelling; Prediction; Forced Air; Radiators; Underfloor Heating; Home Automation

1. INTRODUCTION

Home space heating systems use the largest share of energy for domestic homes in the United Kingdom. In 2009, space heating accounted for 62% of the total domestic energy consumed [1, 2]. Next to transportation, space heating is the second most energy intensive end-use in the UK. In the United States, the situation is similar with domestic space heating using 56% percent of the domestic energy share [3, 4].

Homes in the UK and US are typically equipped with a programmable thermostat, which occupants can use to specify desired temperatures ("setpoints" or "setbacks") for particular time intervals in the day. During the times when heating or cooling is required, the programmable thermostat actuates the heating, ventilation, and air conditioning (HVAC) infrastructure to bring the ambient temperature near the desired setpoint. The ambient temperature is typically that reported by a sensor contained in the programmable thermostat (often in a hallway or living room).

Prior work in the sensor network and ubiquitous computing communities has worked to improve upon such strict timer-based heating, using occupancy-reactive and arrival-predictive control [5, 6, 7]. These methods yield savings and may improve comfort during occupied times when compared to static, programmed timer schedules. However, these utilised simple heating models for their houses. Some rely upon a single measure of indoor temperature (rather than per-room) and all assume a constant, linear increase in temperature for heating periods (e.g. 0.3°C per 10 min).

Complementary to both traditional programmable thermostats and these energy approaches, a heating model could allow future temperature trends to be predicted using the current heating schedule. This allows heat controllers to verify the expected outcomes of their decisions, and adapt to different conditions. Heating controllers need to be able to answer the question "If the heat turns on now will the house be warm enough?". This might be answered by assuming a constant heating rate, but this fails to account for weather and inter-room effects.

This paper proposes that simple temperature sensors (one per room), combined with real-time algorithms can be applied to live data to enhance control solutions. The parameterised model we propose has two defining features. First, it recognises that different spaces heat and cool in different ways and at different times—not only due to insulation, but also due to the thermal masses in the heating infrastructure. Second, it automatically identifies rooms which appear to have a thermal relationship. We employ this model to provide two contributions. (1) We characterise the model's predictive performance, showing the two-hour lookahead er-

IPSN'13, April 8–11, 2013, Philadelphia, Pennsylvania, USA.

ror to be 1.50°C or better (90% confidence level) for all four houses. (2) We highlight the energy savings opportunities which would have been possible by detailing (for two houses) how the predictive model would have turned on heating later but still brought the house to setpoint by the desired time.

Compared to methodologies others have applied in buildings research [8, 9, 10, 11] our method is notable in that it uses data from deployments in occupied family homes across two countries; and it has a longer viable forecast length. We evaluate our model on a month of per-room data for four houses with different construction and heating systems, and analyse the error characteristics down to the room level.

2. RELATED WORK

There are two general types of approach for modelling a building's internal thermal interactions: process-driven and data-driven. These are also known as forward system identification and inverse system identification [12], respectively.

Process driven solutions use complex system equations based upon thermodynamic principles, and materials science. Detailed surveys provide the inputs needed for these equations, such as thermal conductance values, heat capacity, material thickness, solar incidence, and room dimensions.

Process-driven approaches have received much attention in the literature, and many tools exist to aid in their development and use. Large buildings use building information models (BIMs), which is designed at the architect's office, refined during the construction process, and then handed over to the building manager. BIMs include exact dimensions, locations, and materials of the building's components. With such detailed survey data available, process-based approaches can input a this data into a specific HVAC simulation to assess and predict the heating within the building.

There are trade industry software platforms which perform process-driven modelling, such as EnergyPlus, maintained by the US Department of Energy.[1] It should be noted that in general, EnergyPlus works over climate-sized timescales of months or years, and is designed to inform choices among HVAC technologies, layout and configuration. By contrast, data-driven solutions have the capacity to be much more fine-grained, informing real-time daily or hourly control.

Dounis et al [8] used the TRNSYS [13] simulator to model temperature whilst actuating windows to control indoor air quality. A single room was simulated with an RMS error of 0.29°C, but with residuals as high as 1.5°C.

Data driven approaches use models that are based on derived Equivalent Thermal Parameters (ETPs) [14] instead of parameters from a survey. ETPs are derived by statistical regression or neural networks [12] and used to find parameters which fit a training period of data. The parameters can be refined or updated as more data is gathered.

Coley and Penman [9] used a recursive least squares algorithm to build an inverse thermal model of a single room of a school. This model used sensor data every thirty minutes and it took ten days for the model parameters to converge. The parameter space was restricted so that the dynamic effects from input could overcome the inertia of past predictions (akin to the gain of a Kalman filter). The model had an RMS error of 1°C. It was envisioned that significant change in model parameters against historical data would indicate deterioration of the building.

Smith et al [15] used a per-zone linear regressive model to predict a room's temperature 30 minutes into the future. The model used vent level actuation of an American home and supported zonal interactions. Environmental effects such as outdoor temperature, wind, and sunlight were captured and modelled during system off times and used in subsequent heating periods. The predictive model error was within ±2°F (±1.11°C) of ground truth.

Hybrid solutions normally begin with a detailed survey, on which an initial model is based. Live data is then fed in for refinement and to track changing values in the face of environmental and occupant-driven influences.

Mejri, Barrio, and Ghrab-Morcos [10] created a simulation system for modelling office buildings. Using sensor networks to record temperature and electrical load every hour, combined with room volumes, a whole-building thermal model was designed. This model was only applied to office buildings as the interiors and heating schedules rarely change. In simulations, the whole-building model showed an RMS error of 0.7°C.

Spindler and Norford [11] created a multi-zone (each zone had more than one room) model for mixed-mode cooling strategies. The building was surveyed to measure material values and air flow properties, combined with a week of live thermocouple data to refine the accuracy of the survey. The model accounted for thermal resistance and mass effect by using an empirically-chosen number of measurements into the past. The RMS error of most zones was 0.3–0.4°C.

Oldewurtel et al [16] discussed Stochastic Model Predictive Control (SMPC) which augments a process driven building system with a weather model that has stochastic errors. Current building measurements were fed back into the control system along with weather forecasts. The control algorithm was applied to a range of simulated buildings. It was shown that using weather predictions improved control decisions. A weather prediction horizon of one day could be used with only a 5% deviation from the ideal control schema.

Compared to the literature, we provide a model which relies on **simple and sparse (one per room) sensors**. Further, we model each room **individually** to take into account different heating elements. We also show a savings analysis when using the model with sensor data from **real occupied homes**, rather than a test cell.

3. DEPLOYMENTS

Sensors were deployed and data gathered from two homes in the United Kingdom, and two homes in the United States. A variety of sensors were used: in the UK homes custom sensors based on the .NET Gadgeteer [17] framework; and in the US homes iButton Thermochron sensors. The UK homes' data is from a previous heating control study [7], where each room's radiator could be actuated independently. All the deployments were over various winter periods in 2010–2011. The houses were designated UK1, UK2, US1, and US2. The deployment characteristics are summarised in Table 1.

In the UK deployments, a wireless sensor network was deployed which communicated its measurements using an 802.15.4 radio network to a server PC located in the house. Per-room temperature data was logged at a rate of once every five seconds; whole-house gas measurements were taken directly from the utility meters using a pulse counter sen-

[1] http://apps1.eere.energy.gov/buildings/energyplus/

Table 1: House Characteristics

Name	Location	Date	Floors	Rooms	Heating System
UK1	Southern UK	28^{th} Jan – 25^{th} Feb 2011	2	13	Underfloor/Wall-mounted
UK2	Southern UK	2^{nd} Feb – 2^{nd} Mar 2011	3	15	Wall-mounted
US1	Northwest US	8^{th} Dec – 6^{th} Jan 2012	3	12	Forced air
US2	Northwest US	8^{th} Dec – 6^{th} Jan 2012	2	15	Forced air

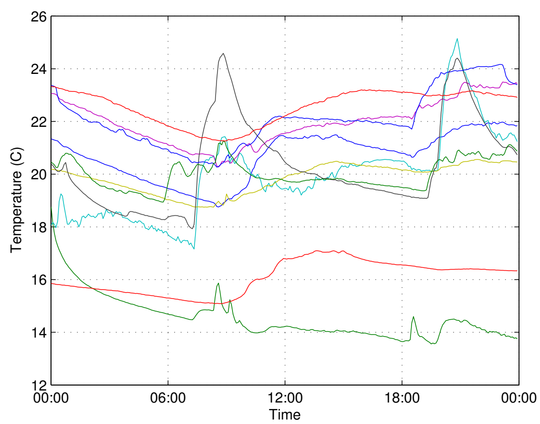


Figure 1: House with primarily underfloor heating (UK1); high thermal mass and per-room control

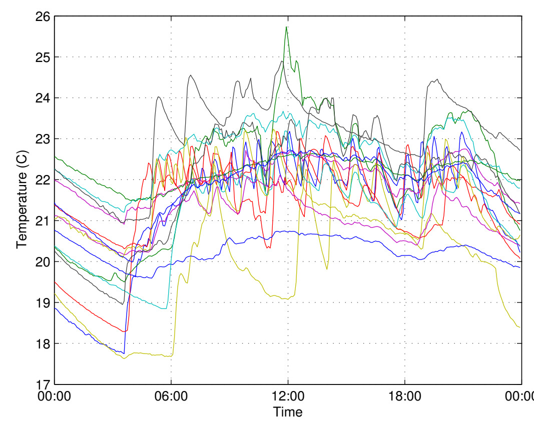


Figure 2: House with wall-mounted radiators (UK2); low thermal mass and per-room control

sor, and had a resolution dependent on the particular utility meter. Outdoor temperature was gathered from a local weather station deployed on the roof of a research building in the same city. Thermostatic radiator valves (TRVs) were actuated by using HouseHeat FHT-8Vs and controlled by the central PC. Readings were downsampled to one measurement per five minutes, for use in our experiments.

In the US deployments, twenty iButton Thermochrons were deployed, with at least one sensor in each room of the houses. For large rooms in open plan designs (e.g. a "great room"), two or three sensors were used. Furthermore, one iButton was placed outside to gather outdoor temperature, and another was placed directly on the furnace in order to sense actuation times. (The furnaces consume a near-constant amount of gas when they are on, so it was not necessary to meter the gas directly.) The sensors sampled temperature once every ten minutes, for a total of four weeks.

3.1 Building Characteristics

UK1 is a recently constructed detached two-floor building with a gas boiler, and underfloor heating on the ground floor. Rooms on the upper floor are in their own heating zone and are heated using TRV-equipped radiators. UK2 is a three-floor mid-terraced 19^{th} century house with wall-mounted convection radiators. US1 and US2 are in the north-west of the USA and utilise forced air heating systems, powered by a furnace.

Looking at a sample of per-room temperatures for each type of heating infrastructure (Figures 1, 2 and 3), it is clear that there is a high temperature variance between rooms located within the same building (up to 5°C). And, underfloor heating in UK1 creates very different temperature trends, compared to the convection heaters of UK2 and the US houses. Convection radiators rapidly bring rooms up to temperature while underfloor heating causes a gradual temperature change. This can be explained when looking at the properties of the different heating systems.

Radiators heat up quickly and are in direct contact with the air, causing the room to heat quickly. Similarly, forced-air systems literally pump hot air into a room. Underfloor systems are embedded in concrete floors which heat very slowly. Effects like these provide good motivation for modelling on a per-room basis, and this is especially true when the heating infrastructure has different heat transfer properties in different rooms. In Figure 1, the upper floor radiators' spiky temperature trends can be seen alongside that of the more gradual underfloor heating.

Figures 2 and 3 show temperature trends from houses with low thermal mass heating infrastructure: UK1 and US1, respectively. However, UK1 has per-room control that adapts to occupancy which means that different rooms might be heating up and cooling off simultaneously, as opposed to all rooms heating and cooling at the same time as in US1.

4. MODELLING

Having surveyed the literature, we decided to use a regression based optimisation model rather than a Kalman filter. We made this decision based on scalability, as optimisation based models can be split by room and trained separately, while existing Kalman filter–based models have been formulated such that all room parameters are trained together. This means that optimisation methods can scale pseudo-linearly while Kalman filters scale in a cubic manner, due to matrix inversions.

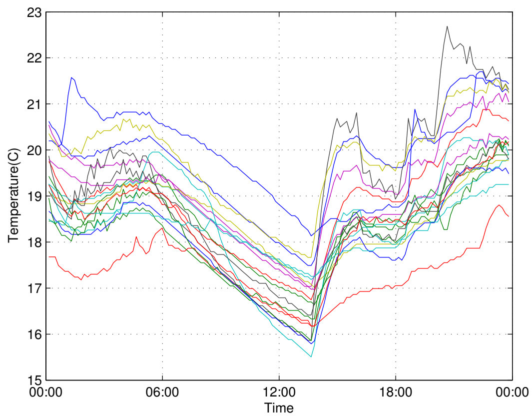


Figure 3: House with forced air convection heating (US1); low thermal mass and whole-house control

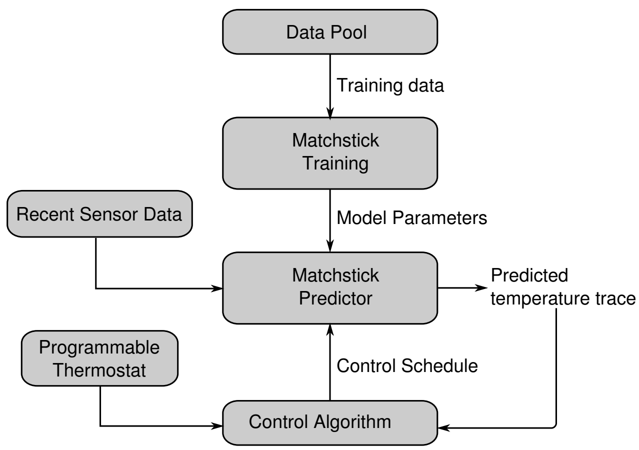


Figure 4: Flow diagram of Matchstick being used by a control algorithm

4.1 Model Overview

Our proposed model, *Matchstick*, takes into account room-to-room interactions, thermal mass delays, and outside temperature. It uses a non-linear transformation of gas use to better reflect the thermal mass present in each rooms' heating element and structure. We now discuss each of these model components and then discuss how Matchstick fits to each room in a building using logged data to periodically retrain, and how Matchstick predicts future thermal states.

Matchstick fits between the heating scheduler, such as a programmable thermostat or an occupancy predictor which dictates setpoint and setback times, and the controller of a heating system. Once it has trained on historical data, Matchstick uses current sensor measurements combined with the proposed heating schedule and then predicts what will happen. Then, using this knowledge, a heating controller can adjust the schedule until it reflects the intent of the program; saving gas and/or increasing comfort. A high-level image of how Matchstick interacts with sensor networks and control algorithms is shown in Figure 4.

An interesting feature of the domestic heating systems that we have observed is the delay between thermal energy

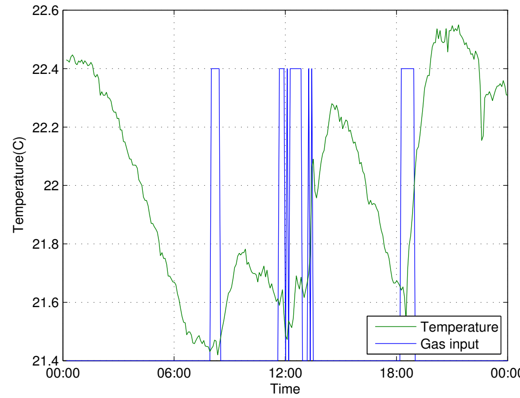


Figure 5: Temperature change and gas input for a room with underfloor heating

input, change of the heating element temperature, and ambient indoor air temperature. This is due to the thermal mass of the heat delivery system. The delay can be easily observed when looking at the plots of gas usage and room temperature, as shown in Figure 5. The room temperature continues to rise long after the boiler has stopped firing.

Air temperature is slower to rise with heating infrastructures involving large thermal masses and/or low conductivity (e.g. the underfloor system in UK1).

To account for this, we create a recursive non-linear transform function $g()$, which takes the raw gas usage for a heating system and the current valve state (in our study, either fully open or closed) for that radiator and outputs the thermal energy transfered into the air at time t:

$$\begin{array}{rcl} g(G_t, R_{T_n}) & = & \sigma(t).{R_{T_n}}^2 \\ \sigma(t) & = & \sigma(t-1)(1-{R_{T_n}}^2) + G_t \end{array} \quad (1)$$

These equations model the energy-storing nature of heating infrastructure, and how heat continues to be stored and radiated once the energy input (in our case, natural gas usage G_t) is off. $\sigma(t)$ represents the thermal energy stored in the room's heating element (i.e. the metal radiator or concrete floor) at time t. The amount of heat which is emitted is dependent on the thermal time constant (R_{T_n}) of the heating element, which is represented as a scalar value between 0 (no storage: energy output is immediate and equal to the gas energy input at that time), and 1 (infinite storage: all energy is stored and does not contribute to room air temperature).

R_T is empirically determined by searching the solution space $[0, 1]$ and finding the value which gives the smallest mean squared error when training the model with historic data. Figure 6 shows the energy output when compared to a sampled room's raw gas input for different values of R_T.

Other techniques to model heat dynamics use infrastructural temperatures (e.g. pipes, radiators) directly. To properly parameterise this property would have required at least one additional sensor per room, which goes beyond what one might reasonably expect in a typical home.

As shown in Section 3.1, there is a large temperature variance between different rooms in a house. As such there will be internal interactions between rooms which can be modelled as thermal flows. To model these flows, first a map of neighbouring rooms is needed. Then, these interactions

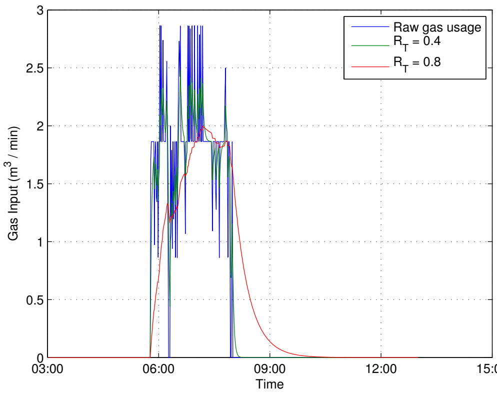


Figure 6: Equivalent gas output for different vales of R_{T_n}

must be expressed in the final system equation. However, the definition of a neighbouring room can be ambiguous. Further, depending on the building materials and heating layout, some neighbouring rooms may have little thermal flow.

To make the system model representative of reality, and to create a mapping of *significant* thermal flows between neighbours, we take an initial list of potential thermal neighbours for each room and trim them down using statistical methods.

The initial list of physically proximate neighbours is a necessary restriction to be placed upon the model training. Starting with an initial list of every room would show correlation over indirect properties such as sharing the same heating zone or having similar radiator settings. The list of neighbours does represent a small amount of survey data which must be provided to the model. However, much of this data can be obtained through sensor adjacency information, determined in an algorithmic way such as those described by Lu and Whitehouse [18]. In this way, the initial neighbour list can be generated automatically for each room, eliminating a dependence on user-supplied survey data.

To determine the thermally significant neighbours, a recursive likelihood test is performed. Initially, the model is fitted with no neighbours, and then a likelihood-ratio test is performed against a model with each possible neighbour fitted. The likelihood ratio test allows a p-value to be computed for rejection of the null hypothesis (which in this case is the reduced model). If the null hypothesis is rejected, the most likely neighbour is added to the reduced model, and the process is repeated until the null hypothesis is accepted. The neighbouring rooms added to the model are then classed as a room's significant neighbours. The p-value used to reject the null hypothesis was p=0.2.

4.2 Fitting the Matchstick model

Matchstick is an adaptive model which, for each day and room, determines which rooms have significant thermal connections and the thermal resistance of the heating elements. It also determines how the outdoor temperature affects each room. Model regression was performed by using MATLAB's `lsqcurvefit` function, which solves non-linear least squares problems. A high-level diagram of the Matchstick training procedure is shown in Figure 7.

Matchstick takes its training data, and initially without

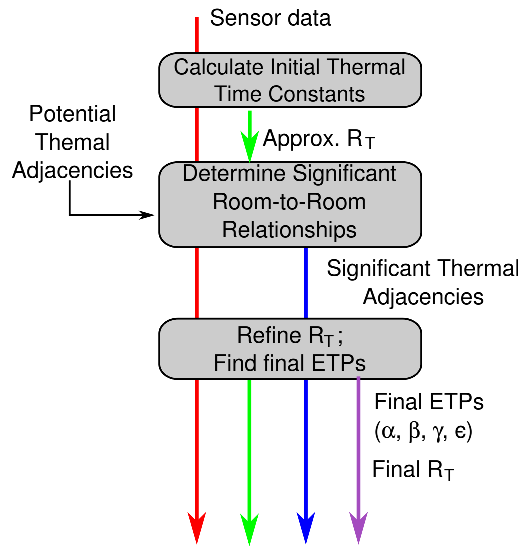


Figure 7: Flow diagram of Matchstick training, showing where variables are created and the flow through the procedure. Thermal resistance, adjacencies, and the training data are used for predictions.

setting any significant neighbours for any room, performs a search on the R_{T_n} thermal time constant for each room. Then, using the R_{T_n} values from the first stage, each room then takes its list of potential thermal neighbours and performs likelihood tests to form a list of thermally significant neighbours for each room. Finally, the R_{T_n} space is searched again, this time with thermally significant neighbours specified in the model. This allows for more accurate R_{T_n} values as heat gains from neighbouring rooms correlate with the neighbouring thermal data, rather than the gas input. The final calculated model ETPs are then used as parameters for prediction, along with the final R_{T_n} values, and the lists of thermally significant neighbours.

The mathematical form of Matchstick's system equations is as follows:

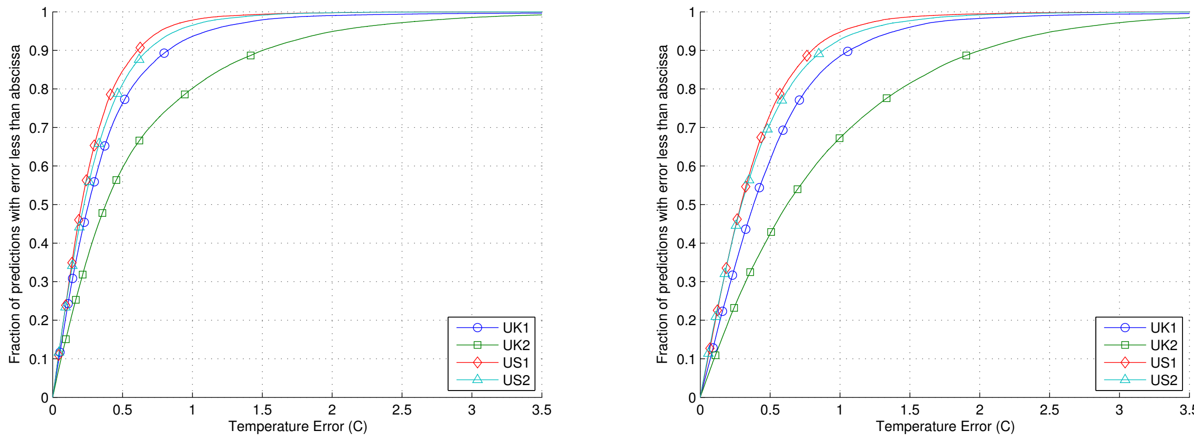
$$\begin{aligned}\forall n \in N,\ T_n(t+1) \quad = \quad & \alpha_t.T_n(t) + \alpha_g.g\Big(G(t), R_{T_n}\Big) + ... \\ & \textstyle\sum_{j \in \mathrm{neigh}(i)} \Big(\beta_{nj}.T_j(t)\Big) + ... \\ & \gamma_o.TO(t) + \epsilon,\end{aligned} \tag{2}$$

where N is the set of all rooms, T_n is the temperature of room n, G is the gas used, and TO is the outside temperature. The ETPs [14] in the model represent loss of heat from the room (α_t), heat transfer from the heating system (α_g), transfer of heat from thermally significant neighbouring rooms (β_{nj}), and the heat transfer with the outside (γ_o). The system model describes how the last known temperature and thermal output affects a room, together with neighbouring room thermal flows, and environmental measurements.

Given the present temperature for each room (measured by sensors), and with forecasts for the outside temperature, this system model can then be used to predict how rooms in a house will react under different heating schedules. For example, if a room only needed to be at its set point 30 minutes before the end of a heating schedule, then the heating can be switched off earlier. A control algorithm can use

Table 2: Summary of per-room temperature prediction error

House	90th percentile range		
	0.5 hour prediction window	2 hour prediction window	6 hour prediction window
UK1	0.1°C– 1.2°C	0.3°C– 1.6°C	0.5°C– 1.8°C
UK2	0.2°C– 2.0°C	0.6°C– 2.5°C	1.2°C– 2.8°C
US1	0.1°C– 0.8°C	0.3°C– 1.2°C	0.6°C– 1.3°C
US2	0.1°C– 0.6°C	0.3°C– 1.1°C	0.5°C– 1.5°C



(a) 2 hour prediction window

(b) 6 hour prediction window

Figure 8: Error residuals aggregates across all houses and all rooms

Equation 2 to extrapolate room temperatures given reduced heating time to determine the ideal new schedule. This can also be done in a feedback control system so long as the predictive look ahead window is large enough that control decisions will have an active effect on the house within the window.

5. EVALUATION

This section investigates the accuracy of the Matchstick model. First, we characterise the predictive accuracy of the model. Second, we analyse how the predictive accuracy changes for different rooms in different houses. Third, we investigate the effect of the model's training aspects, such as the training length and the effect of neighbouring room selection.

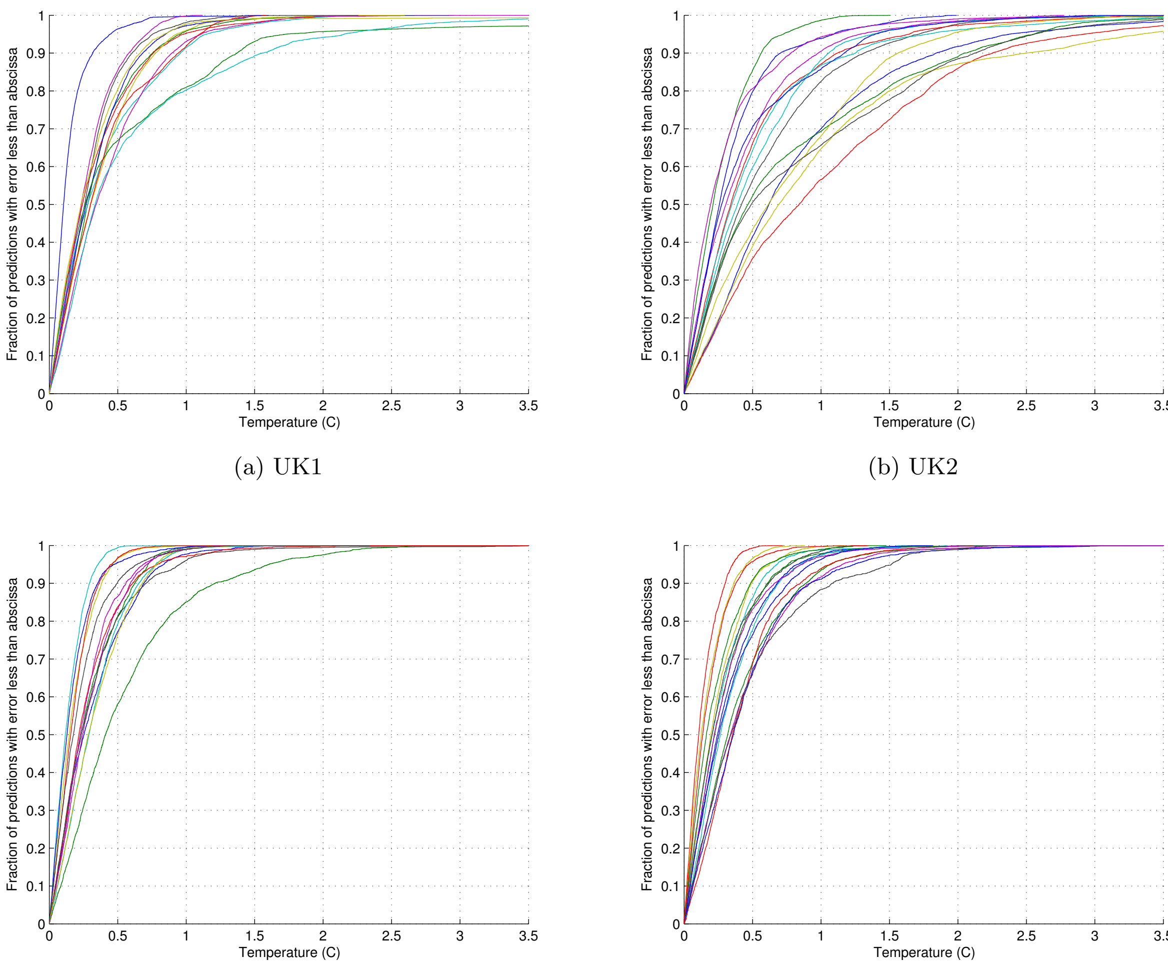
To evaluate the accuracy of the model, we used a total of four weeks of wintertime gas consumption and per-room temperature data for each of the four houses, and outdoor temperature data for the two cities. We give the predictive accuracy of the model across three weeks of data, using a sliding window of the prior seven days as training data. The predictor operates on the present and past per-room temperature readings, and we evaluate the success of its extrapolation for specific times in the future. In our evaluation, we supply the predictor with two types of future knowledge: the future gas inputs (which can be derived from the intended/programmed schedule of the boiler or furnace) and the future measured values of outside temperature.

While our use of outside temperature might be seen as relying upon an "oracle", it is important to note that weather forecasts in most locations are sufficiently accurate for our purposes. For example, in the UK where weather is notoriously variable, the three-hourly temperature forecasts are accurate to within ±2°C (95% confidence).[2] Note that this level of error is similar to that which might arise from using temperature measurements taken at another site in the same locale (since not all homes/neighbourhoods have an outdoor weather station), or which arises from measurements using inexpensive, consumer-grade sensors (typically ±1°C). Regardless, we comment below on the sensitivity of prediction accuracy to errors in outside temperature forecasts.

The experimental evaluation procedure is as follows, for each day. At the beginning of the day (midnight), the model is trained (Figure 4), using the seven previous days of data. For each time step t (from 0–24 h) we take each room's current temperature and predict p hours into the future, by modelling each time step (five-minutely in the UK; ten-minutely in the US) until the time $t + p$ is reached. The predicted per-room temperatures for time $t + p$ are then stored. By modelling the future predictions starting from all time steps in the day in this way, we create a temperature trace made entirely of predictions p hours into the future. This is compared to the temperature ground truth to create temperature prediction error distributions. We varied the prediction lookahead p between one-half and six hours.

Figure 8 shows the cumulative error distribution for all the residuals from all rooms for all houses. Two lookahead windows are shown (2 and 6 h) to illustrate the difference

[2]`http://www.metoffice.gov.uk/about-us/who/accuracy/forecasts`



(a) UK1 (b) UK2

(c) US1 (d) US2

Figure 9: Cumulative temperature error distributions for each house; each line is a separate room in the house (Prediction length = 2 h)

in residuals. The RMS and 90$^{\text{th}}$ percentile error ranges for the two hour prediction window are 0.28–0.62°C and 0.61–1.50°C respectively. The RMS and 90$^{\text{th}}$ percentile error ranges for the six hour prediction window are 0.38–0.88°C and 0.61–1.50°C respectively.

Of interest to control system algorithm designers is prediction accuracy and how the accuracy is related to the length of the forecast. In order to address this question, we evaluated using the same method as above but set p over a range of different windows. The predictive windows we looked at were between half an hour and 6 hours lookahead. The error residuals for each prediction window p were combined across the week to get a better error distribution, and then the one-tailed 90$^{\text{th}}$ percentiles were plotted against p in Figure 10. As expected, shorter prediction windows have smaller error residuals, with $p = 2$ h giving 90$^{\text{th}}$ percentile errors of under 1.5°C for each house. However, for heating infrastructures like underfloor heating, we recommend a predictive window of at least four hours.

To investigate the effect of weather forecast error on prediction error, the above experiment was repeated with a six hour prediction window, for outdoor temperature offsets of both -2°C and +2°C. The change in the 90$^{\text{th}}$ percentile aggregate error was no more than ±0.05°C over all the houses.

While the literature mostly deals with zonal control rather than per-room, we have errors for each room. The residuals for each room are grouped by house and plotted in Figure 9. These temperature errors are for two-hour predictions.

As the Figures show, the error results are encouraging. Per-room 90$^{\text{th}}$ percentile errors are summarised in Table 2 for three prediction lengths. Large prediction windows are explored as heating infrastructures, such as underfloor heating, can take more than an hour to heat a room and can still effect the air temperature hours after the system has been turned "off".

Sometimes, Matchstick trains on sensor data but gives poor predictions. We found that this was primarily due to non-measured sources of background heat gains (or losses) in the building. These sources can vary over time in a way

Table 3: Literature comparison: Using other models on our data

House	0.5 hour prediction window - 90$^{\text{th}}$ percentile			2 hour prediction window - 90$^{\text{th}}$ percentile		
	Matchstick	Smith et al. [15]	Coley and Penman [9]	Matchstick	Smith et al. [15]	Coley and Penman [9]
UK1	0.51°C	0.42°C	2.47°C	0.83°C	1.66°C	8.22°C
UK2	0.94°C	1.12°C	2.92°C	1.50°C	39.81°C	7.75°C
US1	0.36°C	0.44°C	1.19°C	0.61°C	7.80°C	3.76°C
US2	0.36°C	0.84°C	0.97°C	0.68°C	204.51°C	2.57°C

that does not correlate with the model input data. We give some specific examples below.

In UK1, there are two rooms (hall and utility) in Figure 9a which have notably larger errors than the rest of the rooms. The hall contains the front door, which will be frequently opened allowing a very fast heat exchange between the air in the hall and the outside. The utility is connected only to the hall and contains machines for doing laundry. The hall and utility were never heated during the study.

In UK2, while the errors across all rooms are typically larger than in the other houses, there is a clear higher error with certain rooms. These rooms are the utility room, kitchen, downstairs toilet, living room, and the downstairs landing. All of these rooms are on the ground floor, and the utility room and the downstairs landing both have doors to outside. The kitchen also contains cooking equipment which will give off heat using energy we do not measure.

In US1, there was only one room which had significantly larger error than the rest of the rooms: the living room. This was a great room (a large two-story space) and the dynamics of the larger body of air may have unknown effects upon temperature at the point measured by the sensor.

In US2, there was also only one room which had a particularly larger error than the rest of the rooms: the office. This room had an external wall, its door was kept closed, and it held a desktop computer which gave off heat.

These casual heat gains can lower correlation between gas use and temperature rise in neighbouring rooms, leading to poorer fits. Ways to address these issues are discussed in Section 7.

In order to evaluate our model against the literature, the modelling algorithms of Coley and Penman [9], and Smith et al. [15] were implemented and run on our dataset. Coley and Penman's algorithm was designed to run on whole house data, so we treated each room in a house as a building-like structure. The work by Smith et al. focused on zoned American homes and duct actuation, but the equivalent radiator valve data for UK homes was used instead; individual rooms were considered a zone. The US homes did not have per-room control so the vents were modelled as being always open. All the models were given one week of training data and followed the same evaluation protocol as Matchstick.

Table 3 shows the reported error for a 0.5–2 h lookahead, for each house. The model of Smith et al. has comparable error to Matchstick at the 0.5 h prediction window, but model predictions soon diverge as lookahead increases. Each room relied upon the predictions of others in their model, and this could cause predictions to diverge as increased error is introduced with each time step. Coley and Penman's model doesn't diverge at larger prediction windows, but the overall error is five to ten times that of Matchstick. This could be because the model does not capture neighbouring interactions. Without a deeper analysis we can only speculate on why errors are worse than Matchstick's with larger prediction windows.

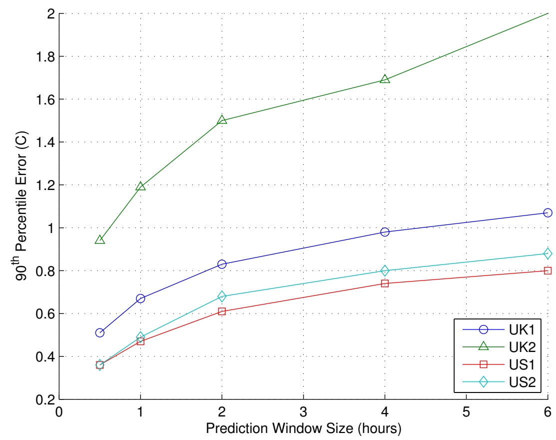


Figure 10: Prediction error as forecast length increases

5.1 Model Tuning

There are a number of aspects which affect how Matchstick reacts to training data. In order to explore the effects of changing these, we took the second week of evaluation data and recorded the accuracy effects as we changed them. The aspects we explore are the length of training data, and how to select initial neighbouring rooms to be passed to the model.

A learning algorithm needs training data in order to understand how its inputs guide its outputs. However, the amount of training data to use is not immediately clear, with arguments for and against large and small training data windows. Small training windows better map parameters to more recent trends and allow the model to better handle drastic changes in the house environment (building work, improvements, or furniture rearrangement). Larger training windows have the advantage of being more robust to bad data (sensor failure, anomalous readings) and create a model which is tailored to a more 'typical' day of a house.

We decided to investigate how a training window that ranges from 1–7 days affects general prediction accuracy. Using the experimental framework from the main evaluation, we gathered error residuals for the week of data, but for different lengths of training window.

As Figure 11 shows, prediction accuracy gets better with more training data. UK1 and UK2 show the most improvement, with each additional day of training. By contrast, the US homes have similar accuracy statistics regardless of training window length. It is worth pointing out that the

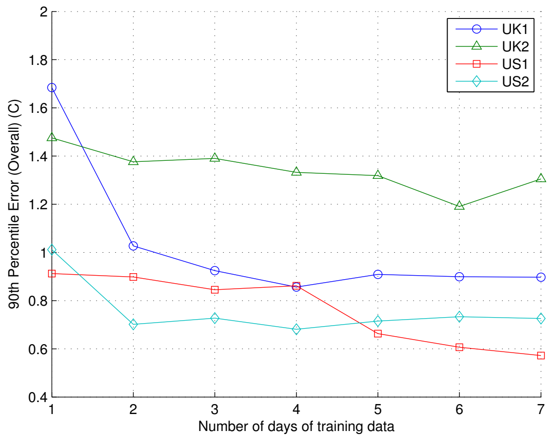


Figure 11: Prediction error as training window increases

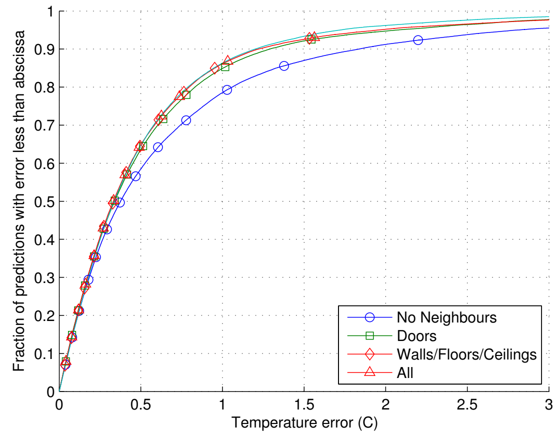


Figure 12: Prediction error when using different policies to set the initial neighbour list (5 hour prediction window)

data for the UK houses used in this experiment had per-room control. The longer training window helps account for situations where infrequent or inconsistent occupancy can result in insufficient training data. This allows for more accurate data correlation, than in shorter training windows. If the relationship between gas and temperature is not reflected in the training data, then this cannot be used by the model.

As each home showed improvement (overall) with each extra day of training, we decided to use seven days of training in the main evaluation. Using seven days also allows for the training data to contain weekend and weekday heating schedules for better correlation in the model fitting. This minimises training artefacts due to irregular heating of specific rooms (such as a utility room or guest bedroom), which again was more common in the UK houses because heating for each room was based on occupancy.

Another aspect of training is the information about potential neighbours. Neighbouring rooms (or adjacent rooms) is an ambiguous term. Is a room neighbouring another if it shares a wall, or a door? To address which policy is better suited to build the initial list of potential neighbours, we did an evaluation of model accuracy (using the same procedure as above), but using four initial lists: (1) no potential thermal neighbours; (2) potential thermal neighbours share doorways; (3) potential thermal neighbours share walls; (4) all rooms are potential thermal neighbours.

Policies 1 and 4 can be used to generate an initial list of neighbours quite trivially. However, policies 2 and 3 require either manual entry or a further supporting algorithm for automatically detecting shared walls and/or doorways [18, 19].

Figure 12 shows the changes in error residuals with the different policies. As expected, setting no neighbours leads to poorer predictions (an error of about 0.5°C worse at the 90^{th} percentile with a six-hour prediction window). As the initial lists get more comprehensive, the accuracy increases. The error residuals for policy 3 is nearly identical, with policy 4 having a slight increase in accuracy. This means that allowing the model to select neighbours from potentially all rooms is comparable to providing a list of physically neighbouring rooms via another means (policy 2 or 3). However, using a full list (policy 4) means the algorithm will be considering possible neighbours which may have no physical relation, and will increase the time it takes to train.

6. SAVINGS ANALYSIS

In order to estimate the potential gas savings when applied to a control scenario, we took the temperature control events from a previous study [7] and used this data in combination with predictions from Matchstick to simulate and improve heating decisions. The previous study used an arrival predictor to work out: (1) when a given room is going to be occupied, and (2) when the room is going to be heated. The time difference between the two was determined by a static linear heating rate. For our savings analysis, we replaced the linear heating rate and instead used Matchstick to determine the ideal time to heat a room. The advantage of using this existing data is that the savings are directly comparable. As we can predict what the temperature will be at the expected time of arrival, Matchstick can work out the nearest time to the arrival point at which to start heating the rooms expected to be occupied. This allows us to minimise gas which took rooms above their set point using the linear heating rate.

The previous study used UK1, UK2, US1, and US2 to determine savings, but the US homes did not have per-room measurements (only a whole-house measure). As such, we perform the savings analysis on UK1 and UK2, which have per-room measurements, the predicted times of arrival for each room, and heating actuation times for each room. The previous study had two experimental conditions which were tested: scheduled heating and predictive heating. We only apply the savings analysis to the predicted days; the heating and expected times of arrival are dependent upon each other and can be leveraged for savings.

For each day of the study with a predicted condition, two events were considered for savings analysis: heat-on in the morning following the night, and heat-on in the evening when occupants come back from work. If the predicted day was a weekend and no one left the house, only the first heating period was used.

A simple control algorithm was used to determine the ex-

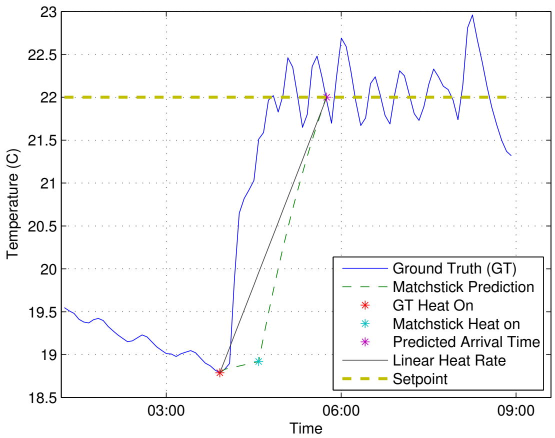


Figure 13: An example prediction trace showing how savings are calculated for a heating period

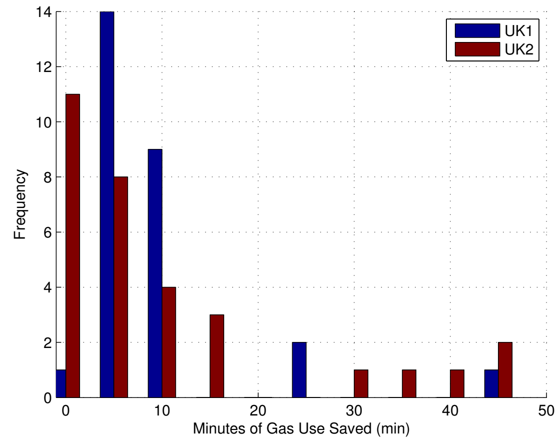


Figure 14: Distribution of gas saved in terms of minutes removed from heating periods

tra savings for heating periods using Matchstick. First, take the temperature data from when the room was originally to be heated. Then for each room, use Matchstick to predict what the room temperature will be for the time it is expected to be occupied. If the room temperature is higher than its set point, re-forecast with the heat turning on one timestep later by shifting the gas trace into the future. This effectively cut away at the gas which was being used keeping the room at a steady state, rather than the gas used to bring the room up to setpoint. This process is repeated until a room's temperature when occupied drops below its set point.

The latest possible time predicted, where all of the rooms are at their set points, is then used as the new time to turn the central heating on. Any gas used in the cut-away steady state period is saved.

6.1 Results

The above methods were used on 47 days of data from the two UK houses (27 for UK1 and 20 for UK2). Figure 13 shows an example of how the new heating times are calculated and Figure 14 shows the distribution of the minutes of gas use saved (as computed by the above method).

Overall, UK1 saved 3.3% of its total gas over the experimental period and UK2 saved 2.3%. The above experimental procedure was designed to save as much gas as possible while not impacting comfort. It's important to stress that these savings are in addition to those reported by the original study [7](8–18% across both houses) and that the savings are based on simulations using Matchstick.

7. IMPROVEMENT AND FUTURE WORK

As described in Section 5, if there are changes in the temperature which do not correlate with any heat input in the training period, or on the day Matchstick is predicting, bad prediction residuals are reported. Currently, we do not address this, but there is a scope for building **fall-back methods** for when Matchstick reports bad prediction residuals.

The parameters from a recent day with a good residual fit could be used, or a set of default model parameters, based on an average of all Matchstick's previous good fits. However, as Matchstick cannot tell if its predictions are bad until the time it has predicted has arrived, warning heuristics or methods must be developed to indicate that the model has a bad fit for the coming day. For example, when bad data is trained upon, the fitted model parameters can be significantly different from those of well-fitted models. This could be an indicator that the model has given a bad fit, but equally it could indicate that the sensors may be malfunctioning, or the local climate may have radically changed from a week ago (heat wave, cold snap).

Time-proven model parameters could be **shared** across similar houses, like in a housing estate where buildings are similar in construction. As noted by Coley and Penman [9], parameters which consistently change from what is expected can also indicate wear and tear in households, and could be used by an estate manager to aid with maintenance.

In the evaluation we have performed, we trained Matchstick at midnight of every day, used that fitted model for all predictions across the following day, and then retrained the following midnight. While certainly feasible to execute once per day, training is a non-trivial operation, and in a real scenario it would take too long to train Matchstick before every prediction, or as new sensor data arrived. We tested Matchstick's performance if retrained every six hours. This did give a marginal reduction of prediction error, but in a control scenario it would take up a lot of processing power which might be better spent elsewhere (e.g. occupancy prediction).

Currently Matchstick only supports heating infrastructure, but could very easily be extended to **support cooling** systems. However, this still needs to be tested on houses with active cooling infrastructure. Other forms of heat gains could be added to the predictive model too, so long as they are either controllable or predictable. Using a combination of non-intrusive load monitoring (NILM) [20] and occupancy tracking, it would be possible to determine likelihoods for certain electronic devices to be switched on, and their casual heat gains. Items such as computers, ovens, and TVs all have regular usage patterns and can heat a room easily. When investigating our test houses, kitchen and living room temperature rose noticeably with electricity usage.

None of the deployments used window or door sensors so we could not model (or predict) transient air filtration events such as windows opening, which may cause large thermal changes. However, our room-to-room model implicitly captures static, or near static, draughts such as doors which stay open or closed for a long time.

Ultimately, to fully verify the savings from using Matchstick and to explore the possible applications which can be built on top of it, the model **must be deployed in the wild**. Matchstick can also be used to reduce discomfort, as well as save gas. Due to the dependency on simple sensors, this could be as simple as deploying iButtons for a month (as in US1 and US2), and then using Matchstick to build an improved furnace actuation schedule. The other possibility is using a home sensor network to feed Matchstick. This could be used to provide live control decisions, as with our saving analysis.

8. CONCLUSIONS

In this paper we have introduced Matchstick, a data-driven adaptive model which relies on relatively sparse sensor deployments (one per room). For differing lookahead windows, we have evaluated how well the model predicts across three weeks of data in four houses, in two different countries. Our evaluation also characterises error distribution in a per-room fashion to give insight into which rooms of a house are more difficult to model and the reasons why.

We have shown that the model is comparable, and in many cases better, than the selected models from the literature. This accuracy was achieved despite the fact that our data was taken in real homes occupied by families, rather than using test cells. We have also shown that by using our model, rather than assuming a constant linear heat rate for warm-up periods, control systems can achieve gas savings by trimming down furnace or boiler actuation schedules.

9. ACKNOWLEDGEMENTS

We would like to thank our anonymous participants for their time and aid in deploying sensors in their homes. This work represents a primary contribution of Carl Ellis's PhD research which was funded by an EPSRC doctoral training grant, and EPSRC grant EP/I00033X/1.

10. REFERENCES

[1] L Pérez-Lombard, J Ortiz, and C Pout. A Review on Buildings Energy Consumption Information. *Energy and Buildings*, 40(3):394–398, 2008.

[2] Department of Energy and Climate. Energy Consumption in the United Kingdom. Technical report, HM Government, 2011.

[3] U.S. Energy Information Administration. Residential Energy Consumption Survey (RECS). Technical report, US Government, 2011.

[4] U.S. Energy Information Administration. Annual Energy Review. Technical report, US Government, 2011.

[5] M Gupta and S Intille. Adding GPS-control to Traditional Thermostats: An Exploration of Potential Energy Savings and Design Challenges. *In Proc. of Pervasive*, pages 95–114, 2009.

[6] G Gao and K Whitehouse. The Self-Programming Thermostat: Optimizing Setback Schedules Based on Home Occupancy Patterns. In Proc. of BuildSys, pages 67–72, 2009.

[7] J Scott, A J Brush, J Krumm, B Meyers, M Hazas, S Hodges, and N Villar. PreHeat: Controlling Home Heating Using Occupancy Prediction. In *Proc.of UbiComp*, 2011.

[8] A I Dounis, M Bruant, G Guarracino, P Michel, and M Santamouris. Indoor Air-Quality Control by a Fuzzy-Reasoning Machine in Naturally Ventilated Buildings. *Applied Energy*, 54(1):11–28, 1996.

[9] D A Coley and J M Penman. Second Order System Identification in the Thermal Response of Real Buildings. Paper II: Recursive Formulation for On-Line Building Energy Management and Control. *Building and Environment*, 27(3):269–277, July 1992.

[10] O Mejri, E Palomo Del Barrio, and N Ghrab-Morcos. Energy Performance Assessment of Occupied Buildings Using Model Identification Techniques. *Energy and Buildings*, 43(2-3):285–299, February 2011.

[11] H C Spindler and L K Norford. Naturally Ventilated and Mixed-Mode Buildings - Part I: Thermal Modeling. *Building and Environment*, 44(4):736–749, 2009.

[12] J Teeter and M Y Chow. Application of Functional Link Neural Network to HVAC Thermal Dynamic System Identification. *IEEE Trans. on Ind. Elec.*, 45(1):170–176, 1998.

[13] S A Klein. *TRNSYS: A Transient Simulation Program*. Eng. Experiment Station, 1976.

[14] R C Sonderegger. Diagnostic Tests Determining the Thermal Response of a House. *ASHRAE Journal*, 19:35–47, 1977.

[15] G Smith, T Sookoor, and K Whitehouse. Modelling Building Thermal Response to HVAC Zoning. In *Proc. of CONET*, 2012.

[16] F Oldewurtel, A Parisio, C N Jones, D Gyalistras, M Gwerder, V Stauch, B Lehmann, and M Morari. Use of Model Predictive Control and Weather Forecasts for Energy Efficient Building Climate Control. *Energy and Buildings*, 45:15 – 27, 2012.

[17] N Villar, J Scott, S Hodges, K Hammil, and C Miller. .NET Gadgeteer: A Platform for Custom Devices. In *Proc. of Pervasive*. 2012.

[18] J Lu and K Whitehouse. Smart Blueprints: Automatically Generated Maps of Homes and the Devices Within Them. In *Proc. of Pervasive*, volume 7319. 2012.

[19] Carl Ellis, James Scott, Ionut Constandache, and Mike Hazas. Creating a room connectivity graph of a building from per-room sensor units. In *Proc. of BuildSys*, pages 177–183, 2012.

[20] C Laughman, K Lee, R Cox, S Shaw, and S Leeb. Power Signature Analysis. *IEEE Power & Energy*, 1(2):56–63, 2003.