Residential remote energy performance assessment: estimation of building thermal parameters using interval energy consumption data

Michael Zeifman, Fraunhofer CSE
Kurt Roth, Fraunhofer CSE

ABSTRACT

Remote residential audits are intended to identify significant retrofit opportunities without home visits. Among such opportunities, the potential for improvement of building envelope thermal properties and/or installation of a high-efficiency HVAC system have some of the largest energy savings potentials for homes located in cooler climates. The state-of-the-art methods for remote auditing are still in their infancy in terms of limited validation and need for home characteristics that are unavailable to utilities (e.g., the building's square footage, volume, orientation, wall materials, HVAC model and age) and/or numerous similar homes in a neighborhood. In this paper, we discuss a new remote audit method that has the potential to estimate key building thermal parameters (overall U-value, ACH, HVAC efficiency) using only interval electricity and/or gas consumption data. Unlike PRISM and other approaches that assume steady-state conditions, our approach capitalizes on physics-based modeling of transient thermal response. In particular, we consider a lumped capacitance/mass heat transfer model for a building. Under some reasonable assumptions, a solution to the underlying equations can be used to relate HVAC run-time and weather data to key building physical parameters. We implement this approach using hour-resolution gas consumption data on eighty single-family households in the Northeast.

1. Introduction

Space-heating and cooling loads in residential buildings consume a significant share of primary energy in many developed countries. For example, in the US, these loads account for more than 9% of total primary energy consumption [DOE, 2011]. Studies suggest that these loads can be significantly reduced by the following three major categories [EEAC 2012, 2013]:

- building insulation,
- air sealing, and
- HVAC systems.

As the retrofit opportunities vary from home to home, the conventional way to identify home energy improvement opportunities and estimate prospective savings is to perform an on-site home energy audit. These are known to be inconvenient to homeowners, expensive, of inconsistent accuracy (usually only qualitative), and thus fail to scale [EEAC 2012, 2013].

A remote audit that would characterize energy saving opportunities and estimate potential savings with no on-site visit can dramatically improve the conventional practice. From utilities' point of view [Klint, 2016], such remote audits should:

- utilize data available to the utilities to approach customers with retrofit offers that are specifically tailored to their homes, and
- provide accurate estimates of the expected energy savings and costs.

Not only would such remote audits not inconvenience home owners, but also their results will be quantitative rather than qualitative.

A conventional approach, the Princeton Scorekeeping Method (PRISM), to identification of candidate homes for energy retrofits [Fels, 1986] is based on the analysis of energy bills and outdoor temperatures. Under certain limiting assumptions (e.g., constant indoor household temperature in either heating or cooling season, no supplemental HVAC sources, constant intrinsic heat gains, negligible building thermal mass), the monthly household HVAC energy consumption should be linearly proportional to the monthly-averaged outside temperature. The slope of this linear dependence is directly related to the PRISM's ultimate measure of household energy efficiency, the normalized annual consumption (NAC). However, since NAC (and the slope for this matter) integrates several different physical building parameters (e.g., R-values of exposed surfaces along with their areas, air change per hour, HVAC efficiency), it cannot be directly used to pinpoint a specific energy retrofit opportunity, e.g., replacement of an old HVAC system with a high-efficiency one or attic/wall insulation. Moreover, the PRISM model does not produce reliable estimates when the underlying assumptions are violated [Minehart and Meier, 1992]. The widespread of smart and/or wirelessly controlled thermostats that result in essentially non-constant indoor temperature is one of the reasons why the PRISM model is becoming less effective nowadays. On the other hand, the availability of interval energy consumption data opens up new opportunities for remote auditing.

In the commercial/industrial building sector, such organizations as First Fuel, Seldera and Retroficiency offer remote audits using interval electricity consumption data together with some additional building data and weather data. In the residential sector, the current offerings largely concentrate on using interval data from communicating thermostats (CTs) [Ecofactor, 2011], [GE, 2012], [Google, 2015], with few offerings by energy disaggregation companies that use smart-meter data. However, the offerings in the residential sector are usually geared towards relatively small energy efficiency improvements and HVAC fault detection (e.g., filter replacement) rather than the identification and characterization of the above-mentioned three major retrofit categories. Also, the accuracy of residential offerings usually cannot be verified in terms of error probabilities in detection or in terms of the prospective savings.

The algorithmic technologies underlying these approaches are not clearly defined. For example, a patent by Ecofactor [Ecofactor, 2011] states that it involves a "system for calculating a value for the operational efficiency of an HVAC system," but it does not clearly explain what particular algorithms are used by this "calculating system." Similarly vague descriptions appear in other patents [Google 2015], [GE, 2012].

What is apparently common in the remote-audit approaches is that a predictive model, connecting the data inputs with the retrofit-characterization outputs, is required for the underlying algorithms to work. Indeed, Gaasch et al. of Retroficiency clearly state that an accurate yet computationally efficient predictive model of a building thermal response is required for early-stage retrofit analysis [Gaasch et al. 2014]. Such a model should be complemented with a simple and reliable way to estimate the retrofit potential from the data using the model.

We have recently reviewed modeling approaches that can be used for development of algorithmic technology for utility-grade remote audits of single-family residences [Zeifman et al., 2016]. Briefly, the predictive models for building thermal response can be loosely divided into white-box, gray-box, and black-box categories [Afram and Janabi-Sharifi, 2015; Berthou et al., 2014].

White-box models are very detailed and accurate physics-based simulation tools, e.g., EnergyPlus. Since they typically require hundreds of parameters to describe a single building, both setting up the model and estimation of its parameters from experimental data to characterize the retrofit opportunities are time-consuming and, sometimes, ill-posed tasks, making the white-box models difficult to scale.

Black-box models rely on large training data sets and machine learning techniques to estimate building physical parameters and/or classify buildings by their retrofit opportunities. Because these models do not have a physical basis, their predictive ability is limited and restricted to homes whose characteristics are represented by those in the training data set. Because of their simplicity, these models can scale fairly easily, but only if appropriate and large training data sets exist.

Gray-box models use relatively coarse-grained physical models (typically, lumped models) with just a few parameters. Although these models seem to combine the advantages of the other two model categories (i.e., physics-based predictive ability of the white-box models and the scalability of the black-box models), they are inherently coarse so that the estimated building parameters may not precisely match the actual physical building parameters.

In this paper, we develop a simple gray-box model of a single-family home with a thermostat-controlled HVAC unit. We show that under certain limiting conditions, the model can be applied to interval electricity- or gas-consumption household data to relate HVAC runtime to the outside temperature and estimate the lumped yet physics-based model parameters that can quantitatively characterize retrofit opportunities and savings.

The paper is organized as follows. In the next section, we present the lumped model along with its underlying mathematical equations and derive a closed-form solution for the state variable, indoor temperature. In section three, we show how this solution can be applied to correlating runtime HVAC data with the lumped building parameters (e.g., overall R-value or HVAC efficiency) under certain limiting assumptions. In principle, the feasibility of the proposed method for remote audits can be tested using interval data from a broad range of homes that were energy-audited, so that the qualitative audit-based "ground truth" on the retrofit opportunities could be used for verification. In this initial paper, however, we apply the method to estimate overall U-value from interval gas data of a limited sample of 84 homes, all of which were characterized as "typical" in terms of their wall insulation. Section 4 explains implementation of our method to this sample including an approach to estimate HVAC runtime from hourly gas data. Although the estimated U-values generally lie within the "typical" range for the US homes, we also estimate the home U-values using PRISM for additional verification. The overall results are summarized in section 5.

2. Lumped Model of Home Thermal Response

Among various "physics-based" models of building thermal response, a model of Constantopoulos et al [1991] is arguably very popular. It is given by the following recursive equation:

$$T_{i+1} = \varepsilon T_i + (1-\varepsilon)\left[T_a - \frac{\eta q_i}{B}\right], \varepsilon = \exp(-\frac{\tau B}{m_c}) \quad (1)$$

where T_i is the temperature in the residence at time t_i, ε is the system inertia depending upon the insulation B, the thermal mass m_c, and the time span τ between the two time points t_i and t_{i+1}, q_i and η are the energy consumption and efficiency of the HVAC system, and T_a is ambient temperature.

Although Eq. (1) seems to be "physically reasonable," its actual origin is in the electrical circuit analogy. It can be shown, however, that this equation is equivalent to the following "thermal conductance" equation

$$\frac{dT}{dt} = \frac{B}{m_c}(T_a - T) - \frac{\eta q}{m_c}. \tag{1a}$$

A steady-state version of Eq. (1a) underlies the well-known PRISM model [Fels, 1986]. On the other hand, this model (of Constantopoulos et al. [1991]) misses such essential physical element as the thermal capacity of walls[1].

Since the model needs to be tailored to a major retrofit opportunity, it needs to incorporate characteristics of building insulation and air tightness. Accordingly, we devised a minimal lumped model that comprises two capacitances (indoor air and lumped wall) and two resistances (lumped wall and convection infiltration), i.e., it is an R2C2 gray box model [Berthou et al, 2014]. Following Tashtoush et al [2005], we use balance differential equations to model this system mathematically

$$C_r \frac{dT_r}{dt} = U_w A_w (T_w - T_r) + L(T_a - T_r) + \eta q^* + q_{\text{int}} \tag{2}$$

$$C_w \frac{dT_w}{dt} = U_w A_w (T_r - T_w) + U_w A_w (T_a - T_w) + q_{ext} \tag{3}$$

In Eqs. (2)-(3), C_r is overall thermal capacitance of the indoor air, U_w and A_w are overall heat transfer coefficient and area of the walls (i.e., building envelope), C_w is overall thermal capacitance of the walls, q^* is HVAC's energy consumption rate (positive for heating, negative for cooling and zero for off-state), η – thermal efficiency of the heating/cooling system, q_{int} – internal household heat gains/losses (e.g., solar gains, non-HVAC appliances, window openings), q_{ext} – external wall heat gains/losses (mostly, solar gains), L – the convective heat resistance that models air infiltration, and T_r, T_w and T_a are the temperatures of residence air, wall and ambient outdoor air. Note that these equations differ from those used in Tashtoush et al [2005].

This model, Eqs. (2)-(3) is mostly applicable to forced-air heating and cooling (in dry climates) systems. Unlike the forced-air systems that supply heated or cooled air and accordingly can be modeled as a source term (q^*) in Eq. (2), boiler-based HVAC systems need special modeling, as the heat generated by fuel is applied to water that heats up metal radiators that, in turn, heat the indoor space. At a first approximation, a gray-box model of a boiler-based heater can be a combination of an ideal heater and a lumped capacitance [Peeters et al, 2008]. Development of boiler model is beyond the scope of this paper.

A closed-form solution to Eqs. (2)-(3) has the following form:

[1] As indoor air is not actually subjected to the thermal difference between the outdoor and indoor temperatures, it exchanges energy with the internal surfaces of walls and partitions [Molina et al, 2003].

$$T_r(t) = a + b\exp(-s_1 t) + c\exp(-s_2 t)\,, \tag{4}$$

where parameters a, b, c, s_1 and s_2 are functionally related to the parameters of Eqs. (2)-(3) and to the initial conditions (i.e., initial room temperature and initial wall temperature).

In principle, the five parameters of Eq. (4) can be estimated by fitting this equation to experimental indoor temperature curves, and the physical parameters related to the retrofit opportunities, i.e., U_w, L, and η, can be calculated by using the functional relation between them (not shown in this paper, see Ref. [Zeifman and Roth, 2016]). However, because of the lack of indoor temperature information, whole-home energy consumption interval data cannot be directly used with Eq. (4). Moreover, even though the fitting to Eq. (4) seems to be straightforward for CT data that usually include interval indoor temperature data [Goldman et al, 2014], there are several challenges that complicate parameter estimation.

Among these challenges are

- **Large estimation uncertainty** A straightforward way to fit Eq. (4) to an experimentally recorded time-dependent (i.e., interval) indoor temperature is by selecting periods of time during which the HVAC status did not change (so that q^* was either approximately constant or zero) and the outdoor conditions were quasi-constant (so that the wall temperature was not subject to changing outdoor conditions). Accordingly, a single "on" or "off" portion of a HVAC cycle should be used. Since during a usually prolonged setback the outdoor conditions typically change significantly, the setback periods of time cannot be straightforwardly used for estimation. For the portion of a regular cycle at a fixed temperature setpoint, the indoor temperature is likely to change just within the thermostat deadband (typically, 1°F), whereas the recommended range of temperature for the parameter estimation of equations of type (3) in literature is tens of degrees [Bouache at al, 2015]. Accordingly, the numerical fitting procedure may not converge at all or produce unreliable parameter estimates.
- **Internal/external heat gains unknown** The exact value of internal/ external heat gains q_{int}/q_{ext} are not known. Moreover, these gains are likely to change with time along with occupant activities and weather.
- **Direct separation of insulation and infiltration is not possible** Although not obvious in Eq. (1), the parameters U_w and L cannot be readily separated since the initial wall temperature $T_w(0)$ is linearly proportional to the ambient temperature.

One way to alleviate these challenges and make Eq. (4) amenable to energy interval data is by

- **Using a Taylor expansion of the solution, Eq. (4).** This yields two important gains:
  - Eq. (4) becomes algebraic with easily estimable parameters
  - By considering consecutive periods of on and off, the internal gains can be cancelled. Moreover, instead of the interval temperature data, durations of time on and off can be used for parameter estimation.
- **Restricting data to periods of time with minimal internal heat gains/losses** Typically, during night time peoples' activities are minimal, and no solar gains exist (at least in the continental US) so that experimental data collected during the night time would have minimal or negligible values of q_{int}/q_{ext}.

- **Considering correlation between wind data and the infiltration** Unlike the insulation, air leakage is known to depend upon wind speed and direction [Gowri et al, 2009]. Using a mathematical model for such dependence along with the wind data, we can separate the insulation and infiltration.

In the next Section, we consider application of Eq. (4) to interval energy consumption data.

3. HVAC Runtime – Outside Temperature Correlation

Following the reasoning detailed in the previous Section, in what follows we restrict consideration to nighttime data only. Since the variable q^*, characterizing in Eq. (2) energy consumption rate of a thermostat-controlled HVAC system, is zero when the system is off and often approximately constant when the system is on, it is convenient to consider the basic equations (2)-(3) separately for the separate HVAC states, i.e., on or off. Assuming the duration of time on/off to be relatively short[2], we can take a two-term Taylor expansion of the solution, Eq. (4). Assuming also constant ambient conditions during such short periods of time, we get the following linear approximation for time on (or off) duration:

$$\frac{1}{t_{on/off}} \approx -\frac{bs_1 + cs_2}{\Delta T}, \tag{5}$$

where ΔT is the thermostat deadband.

Eq. (5) along with the functional dependences of the parameters b, c, and $s_{1,2}$ on the parameters of Eqs. (2)-(3) (not shown in this paper, see Ref. [Zeifman and Roth, 2016]) suggest that the inverse of the time on (or time off) duration, Eq. (5), is linearly proportional to the ambient temperature. Assuming that the initial wall temperature for such short on-off cycles is close to the steady state temperature[3], we get for the slope of this linear dependence:

$$|\,Slope_{t_{on/off}\ \mathrm{vs}\ T_a}\,| \approx \frac{\alpha + 2\beta}{2\Delta T}, \tag{6}$$

where

$$\alpha = \frac{U_w A_w}{C_r} \qquad \beta = \frac{L}{C_r} \tag{7}$$

For the intercept of the tangent line, Eq. (5), with the temperature axis (corresponding to $t_{on/off} \to \infty$), we get

$$(T_r(0) - T_{a,\mathrm{inter}})\frac{\alpha + 2\beta}{2} - \eta q^* = 0, \tag{8}$$

---

[2] Our estimates suggest that time on/off need to be shorter than 30-50 minutes for most wood or masonry walls for the two-term Taylor expansion to be accurate within 20%. If the actual times are larger, a higher-order Taylor expansion can be used.

[3] In reality, the initial temperature of the lumped wall deviates from the steady state temperature, which is one of the reasons for significant scatter in inverse time on/off – outdoor temperature correlation; see Figure 2.

where $T_{a,inter}$ is the intercept point on the ambient temperature axis. Eq. (8) allows estimation of the setpoint temperature[4] from the time-off plot ($q* = 0$) and then, the efficiency η from the time-on plot, if the energy consumption rate is known.

As we discussed earlier (see Section 2), the air leakage expressed by variables L and β can be separated from the insulation resistance (α) using a functional dependence of the air infiltration on wind speed. Accordingly, without loss of generality, we can neglect β in Eqs. (6) and (8). In this way, we can estimate the retrofit-related parameters from the correlation between the inverse time on (off) and the ambient temperature.

For example, for the overall U-value we get

$$U_w = \frac{2\Delta T \cdot C_r \, |Slope|}{A_w} = \frac{2.58 \cdot 10^3 \Delta T \cdot V \, |Slope|}{A_w}, \tag{9}$$

since

$$C_r = V \cdot 1.29[kg/m^3] \cdot 1000[J/kg \cdot K], \tag{10}$$

where V is the building volume. Note that although Eq. (9) requires both overall wall area (A_w) and building volume (V), their ratio can be considered as a less fluctuating form factor.

For initial testing of the predicted correlation, Eqs. (5)-(9), we used experimental data on three test huts, wooden shed-like structures (footprint ~8'x12', ~8' high) that had three different levels of insulation but were otherwise identical. The huts were equipped with portable air conditioning units controlled by a regular thermostat, and electric power and outdoor temperature were measured at 1-minute intervals during summer in Albuquerque, NM. We calculated the durations of time on/off using these data and estimated the overall U-values for the three huts using Eq. (9). The estimated U-values ranged from 1.25 [Btu/ h·ft2·°F] for the original hut to 0.18 for the most insulated hut, whereas U-values calculated for walls with EnergyPlus ranged from 1.43 to 0.15 for the same huts, which indicated applicability of the proposed method to estimation of overall U-value in controlled test conditions.

In this work, we apply this correlation to interval data of single-family detached homes located in Holyoke, Massachusetts. The methodology and results are presented in next Section.

4. Case Study: Real Homes in Heating Season

The proposed method estimates lumped physical parameters corresponding to the three major retrofit opportunities. Conventional on-site audits, however, do not usually yield quantitative parametric values. Instead, the retrofit opportunities are typically characterized qualitatively or categorically in onsite audits. Accordingly, to demonstrate feasibility of our method, it is desirable to have a sample of on-site audited homes that were determined to have various degrees of retrofit opportunities.

Fraunhofer CSE has an agreement with Holyoke Gas & Electric (HG&E) to analyze interval data from approximately 15,000 homes of HG&E's customers located in Holyoke, Massachusetts. All of these homes are equipped with electric meters that provide 15-minute electricity interval data, while some (~5,000) are also equipped with advanced gas meters that provide 1-hour gas

[4] More accurately, it is the so-called balance point, as even at nighttime, there are additional to HVAC heat gains.

interval data. For this preliminary analysis, we received data from 84 coupled accounts for which some partial “ground truth” in terms of home insulation level[5] and HVAC heating type was available.

Fig. 1 shows electricity and gas consumptions over a cold and a mild nights for home #1 from this data set. For the cold night, the electricity consumption indicates presence of a cycling electric appliance with about 300 W power draw. Although this appliance could potentially be a fan of the gas furnace, no such appliance is observable for the mild night. Therefore, the cycling electric appliance cannot be the gas furnace fan; it could be, e.g., an auxiliary heating device. The gas consumption at both nights indicates presence of a large cycling gas appliance that could only be a space heating device with no forced air, i.e., a boiler[6].

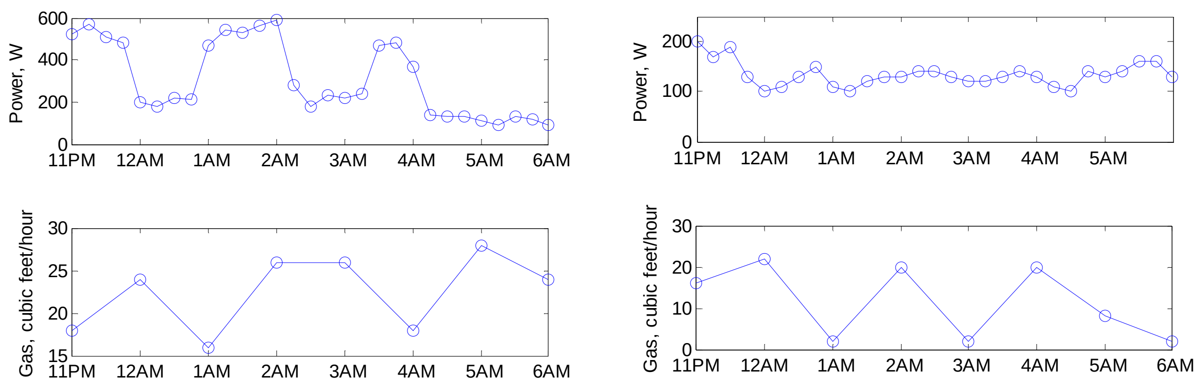

Figure 1. Electric and gas consumption recorded during night hours for home #1 from HG&E data set. Left: cold night of 20-21 February, 2015 (minimum temperature -18°F). Right: mild night of 24-25 December 2014 (minimum temperature 41°F).

The patterns of electricity and gas consumption shown in Fig. 1 are typical for the HG&E data set. The ground truth data on this data set also suggests that out of the 84 homes, only four used electricity and three used oil for space heating, with all the remaining homes using gas boilers, not furnaces. The main challenge of application of the proposed method to the data set is determination of HVAC runtime. Indeed, determination of time on/off of a 50-70 W device (i.e., boiler water pump) from 15-minunte resolution data is beyond the capability of the state-of-the-art disaggregation methods. The strongly non-constant gas consumption, however, indicates a possibility of using gas consumption data to determine characteristic runtime overnight.

4.A HVAC runtime estimation

A detailed description of the runtime estimation procedure is given elsewhere [Zeifman and Roth, 2016]. Briefly, the main idea is that the large fluctuations of gas consumption during nighttime as seen in Fig. 1, are caused by imposition of the hourly time window (recording time) on the cycling pattern of HVAC gas consumption (i.e., approximately constant gas consumption rate during time on and zero gas consumption during time off). By varying the potential durations of time on, time off and the time lag between the start of first cycle and the start of the recording

---

[5] HG&E assigns three insulation grades to homes according to audit results: good, typical and fair. All 84 homes were graded as typical.

[6] Electric water pumps of residential gas boilers usually draw about 50-70 W of power.

hour, and imposing some heuristic constraints to prevent meaningless results, it is possible to create estimated gas consumption profiles and select the one that best fits the actual gas consumption profile recorded over a particular night. Since there are only eight data points for overnight gas consumption, we assume constant durations of times on/off during a night, so that only three variables will need to be determined. In this case, we will only have a single pair of data points for the correlation, Eq. (5), per night.

Although this assumption is coarse, at the first approximation, it would result in increasing the dispersion in the inverse time on/off – outdoor temperature correlation, Eq. (5), and correspondingly increase the uncertainty in U-value estimation, Eq. (9), but should not lead to a bias in the estimated U-value.

A bias can be induced if a household implements thermostat setbacks during the nighttime. However, the gas consumption data plotted on Fig. 1 (or for randomly selected other homes in this data set) do not indicate a systematic reduction in gas consumption that would follow such a setback. On the other hand, a thermostat setback at night would either imply a different thermostat setpoint during most of the nighttime in case the actual room temperature, as sensed by the thermostat, quickly reached the new setpoint, or a larger thermostat deadband ΔT otherwise. Both possibilities can be mathematically modeled.

4.B Inverse runtime – outdoor temperature correlations and U-value estimations

Using the method explained in the previous Section, we estimated durations of time on and off for each night over the heating season 2014-2015, for each home in our data set. For the outside temperature, we used the Weather Underground hour-resolution historical data for Holyoke, Massachusetts, and calculated a median temperature for each night to be matched with the inverse runtime, Eq. (5). Since the gas-heated homes consume much more gas during the heating season, the seven homes that use electric- or oil heating were clearly identifiable on the overall gas consumption plots (not shown in this paper), and the data from these homes were excluded from further consideration.

Fig. 2 shows the experimental correlations for a randomly selected home. The predicted correlations, Eq. (5) are clearly seen for relatively small times on and off. For larger times (approximately those exceeding 30 minutes), the correlation is much weaker, most likely because the two-term Taylor expansion becomes a too coarse an approximation (see Section 3).

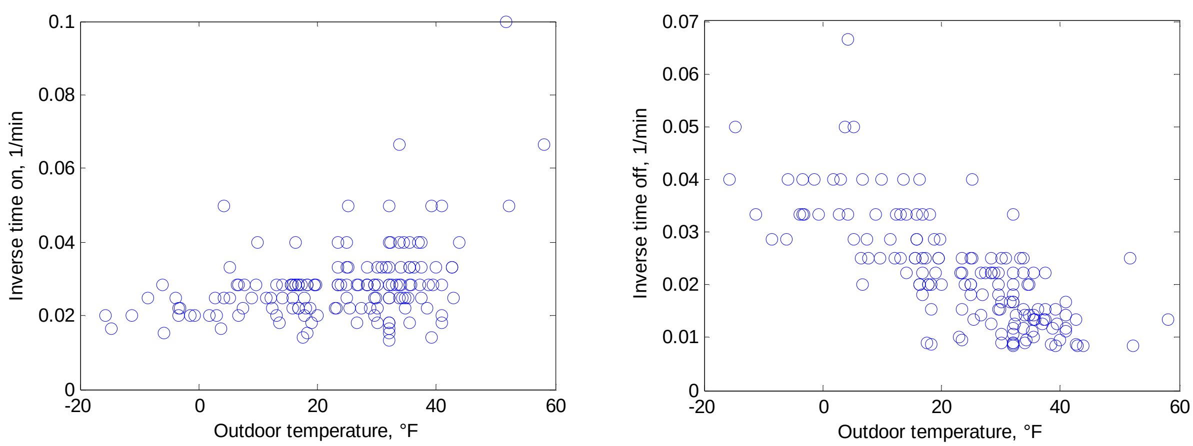

Figure 2. HVAC runtime vs outdoor temperature estimated for nighttime over heating season for home #56 from HG&E data set. For each night, a single value for outdoor temperature and estimated time on and time off was used.

To estimate the overall lumped U-value by Eq. (9), in this initial work, we simply selected data pairs with the temperature exceeding 20°F for the time on correlations, estimated the linear regression slope, and assumed the ratio V/A_w = 2.5 m and thermostat deadband ΔT = 1 °F for all the homes. Also, to convert the U-value calculated from the lumped differential equations, Eqs. (2)-(3) to the conventional U-value, we need to use a correction factor of 2 [Zeifman and Roth, 2016].

Fig. 3a presents the U-value estimation results. Since we do not have a ground truth for the U-value, we arbitrary translate the "typical" grade of the home insulations obtained by HG&E into 0.08 [BTU/ft2·°F·h] [7] as the recommended R-value for walls in US climate zone 5 [Insulation Chart, 2016] is 13 [ft2·°F·h/BTU]. It is seen in the Figure that the estimated U-values range from ~50% to ~ 200% of this "typical" value, which we believe is still within the "typical" category.

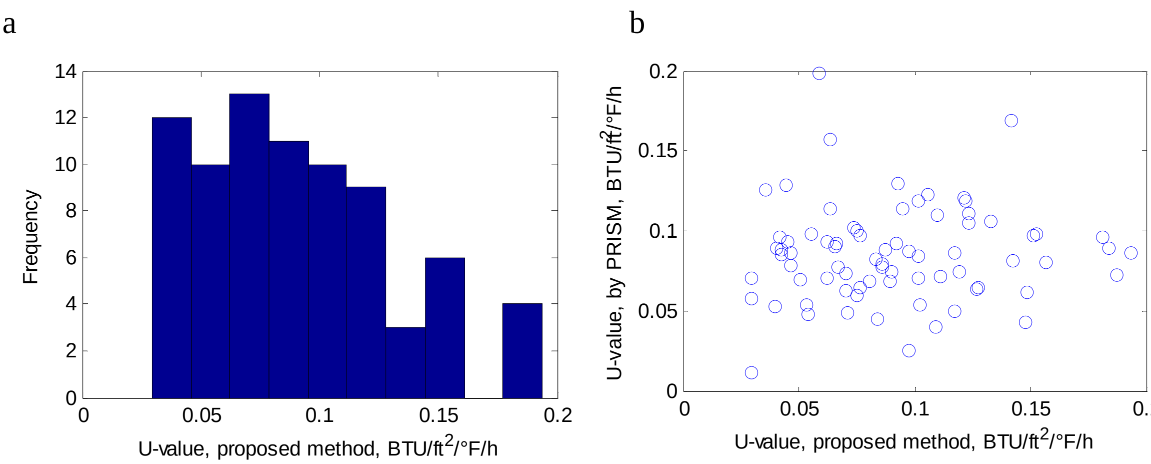

Figure 3. Overall U-values of HG&E homes estimated by the proposed method (a) and by the proposed method and PRISM (b).

To get a better sense of the feasibility of the proposed method, we used the advantage of availability of interval gas data. Accordingly, we converted the measured gas consumption flow rate into the rate of energy consumption using the combustion heat of the methane, and applied the PRISM method [Fels, 1986] to nighttime data to estimate the overall U-value, assuming 100% HVAC efficiency and the total external area of a home being 400 m^2. The PRISM-based estimates of U-values are plotted against the U-values estimated by the proposed method in Fig. 3b. It is seen in the Figure that the PRISM-based range of U-values is similar to that of the proposed method, which partially validates the latter. On the other hand, there is no correlation between the U-values calculated by either method. We attribute this lack of correlation to different sources of uncertainty underlying the two methods.

In the proposed method, the sources of uncertainty include model-based factors (general coarseness of lumped second-order gray-box models, limitation of two-term Taylor expansion and assumed in this work steady-state initial wall temperature), time estimation factors (indirect estimation of time on/off from hourly resolution data, arbitrary usage of 20°F as the boundary for correlations) and geometric factor (V/A_w). For the PRISM as applied to nighttime hourly data, the sources of uncertainty are also model-based (essentially coarseness of first-order lumped model coupled with the steady state and constant indoor temperate assumptions), the HVAC efficiency

[7] Since we neglected air leakage (modeled by β in Eqs. 7-9), the U-value is overestimated in this study even if the assumed form factor is correct.

that cannot be decoupled from U-value, and the geometric factor (A_w). Since these uncertainties are different and because the overall uncertainty apparently is larger than the actual variability of U-value among the homes, the estimated U-values are not correlated either.

5. Summary

In this paper, we developed a coarse yet physics-based lumped gray-box model. The proposed model was further simplified to work with HVAC runtime data that, under certain assumptions can be correlated with the outdoor weather conditions (temperature and wind speed) to yield three physical home characteristics that correspond to three major retrofit opportunities – building envelope insulation, air sealing improvements and HVAC efficiency upgrade. The model was applied to experimental data from about eighty homes and partially validated using the categorical "ground truth" on the home insulation levels and a conventional method for home insulation estimation.

Unlike the PRISM method, the proposed method does not require fuel consumption data for characterization of building insulation and air tightness. The proposed approach also does not require a steady state and can decouple the HVAC efficiency and building insulation level. Moreover, it operates with the building shape factor (volume to surface ratio), which is advantageous for estimation purposes over the surface area used in PRISM.

That the model designed for a "directly-controlled" HVAC system (like a gas furnace) worked well for the indirectly-controlled HVAC systems in the HG&E data set (with boilers, the room thermostat controls a water pump, but the boiler supply water temperature is controlled by an aquastat) indicates both robustness of the method and potential for further improvement. Such an improvement would involve a two-stage model for the boilers.

Other directions of improvement include development of more accurate methods for HVAC runtime estimation from interval electricity and/or gas data, estimation of building geometry using publicly-available data, and development of more accurate methods for computationally-efficient parameter estimation from the experimental data. The authors have also applied this method directly to interval data from communicating thermostats.

Acknowledgement

The authors are grateful to Sarah LaRose and Robert Letellier of Holyoke Gas and Electric for providing anonymous household data.

References

Afram, A. and F. Janabi-Sharifi. "Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods." *Energy and Buildings* 94, pp. 121-149 (2015)

Berthou, T. et al. "Development and validation of a gray box model to predict thermal behavior of occupied office buildings." *Energy and Buildings* 74, pp. 91-100 (2014)

Bouache, T., K. Limam, W. Bosschaerts. "New thermal parameters identification approach applied to the thermal renovation of buildings," *Energy and Buildings* 104, pp. 156–164 (2015)

Constantopoulos, P., Schweppe, F.C., Larson, R.C. "ESTIA: a real-time consumer control scheme for space conditioning usage under spot electricity pricing," *Computers and Operations Research* 18, pp. 751–765 (1991)

DOE 2011. Buildings Energy Data Book, tables 1.1.3 and 2.1.5 (2011)

EEAC. 2012. Massachusetts Energy Efficiency Advisory Council Annual Report, available at http://ma-eeac.org/results-reporting/annual-reports/ Assessed March 10, 2016

EEAC. 2013. Massachusetts Energy Efficiency Advisory Council Annual Report, available at http://ma-eeac.org/results-reporting/annual-reports/ Assessed March 10, 2016

Ecofactor. US Patent 8,019,567 B2 "System and Method for Evaluating Changes in the Efficiency of an HVAC System," by Steinberg, J.D., and S.D. Hublou (2011)

Fels, M.F. "PRISM: An Introduction." *Energy and Buildings* 9, pp. 5-18 (1986)

Gaasch, W.H., et al. "A Comparison of Methods for Early-Stage Retrofit Analyses," Proceedings of ACEEE Summer Study on Energy Efficiency in Buildings, pp. 11.149-161 (2014)

GE. 2012. US Patent Application 2012/0016524 A1. "Thermal Time Constraints for Demand Response Applications," by Spicer, L.B., et al.

Goldman, E., et al. "Are thermostats the new energy audits?" Efficiency Vermont (2014)

Google. 2015. US Patent 9,092,040 B2. "HVAC filter monitoring," by Fadell, A.M., et al.

Gowri, K., D. Winiarski, R. Jarnagin. "Infiltration Modeling Guidelines for Commercial Building Energy Analysis," PNNL report (2009)

Insulation Chart. "Insulation R Value Chart," Assessed March 10, 2016 http://www.greatdayimprovements.com/insulation-r-value-chart.aspx

Klint P. Private correspondence (2016)

Minehart, D.L. and Meier, A.K., "Using Simulations to Explore the Physical Meaning of PRISM's parameters," Proceedings of ACEEE Summer Study, pp. 4.145-155 (1992)

Molina, A., et al. "Implementation and assessment of physically based electrical load models: application to direct load control residential programmes," *IEE Proceedings - Generation, Transmission and Distribution* 150, pp. 61 – 66 (2003)

Peeters, L. et al. "Control of heating systems in residential buildings: Current practice," *Energy and Buildings* 40, pp. 1446–1455 (2008)

Tashtoush, B., M. Molhim, M. Al-Rousan. "Dynamic model of an HVAC system for control analysis," *Energy* 30, pp. 1729-1745 (2005)

Zeifman, M., B. Urban, K. Roth. "Enabling Virtual Energy Assessments for Homes: Methods for Quantifying Building Performance with Connected Thermostat Data," *Report to EverSource/ National Grid* (2016)

Zeifman M. and K. Roth. "Estimation of physical parameters of homes for remote audits using interval data" *Provisional Patent Application*, USPTO (2016)