Estimation of Physical Buildings Parameters Using Interval Thermostat Data

Amine Lazrak
Fraunhofer USA, Inc.
5 Channel Center Street, Boston, MA 02210
+1 617-575-7251
alazrak@cse.fraunhofer.org

Michael Zeifman
Fraunhofer USA, Inc.
5 Channel Center Street, Boston, MA 02210
+1 617-714-6510
mzeifman@cse.fraunhofer.org

ABSTRACT
Significant energy savings can be achieved by retrofitting the enclosures and HVAC systems of existing residential buildings. Identification of building retrofit opportunities currently requires on-site energy assessments that are inconvenient to homeowners, expensive, and are of variable accuracy, making it challenging to deliver cost-effective retrofit opportunities at scale. Massive deployment of communicating thermostats provides a possibility for remote energy assessment by analyzing the associated interval indoor temperature and heating system run-time data. In this paper, we present a methodology to estimate the overall building insulation level, HVAC system efficiency, and building airtightness from the communicating thermostat data. The methodology uses a grey-box model of a residential building and includes identification of basic model parameters, followed by estimation and non-parametric modeling of generally variable external and internal heat gains/losses. In this way, it is also possible to predict indoor temperature and energy consumption of the building under various retrofit scenarios and user behaviors. Preliminary results demonstrate the feasibility of the proposed method.

CCS Concepts
• CCS → Applied computing → Physical sciences and engineering → Engineering. • CCS → Computing methodologies → Machine learning.

Keywords
Building energy; Parameter identification; Retrofit; Data analytics; Grey box modeling.

1. INTRODUCTION
The residential building sector has a large energy savings potential. Around 20 to 25 percent of U.S. homes have poor or no insulation, and a similar portion of Massachusetts homes have older – and thus likely inefficient – heating systems [1]. Such buildings may have higher heating energy consumption, and those with higher air leakage levels [2] tend to consume even more energy.

BuildSys '17, November 8–9, 2017, Delft, Netherlands

ACM ISBN 978-1-4503-5544-5/17/11…$15.00
https://doi.org/10.1145/3137133.3137161

Upgrading wall and/or attic insulation, air sealing, and HVAC systems in existing buildings can significantly reduce space heating and cooling energy consumption.

According to [1], savings of up to $4-5 billion per year can be achieved nationwide through basic retrofits of existing homes to reduce space heating energy consumption.

However, the identification of building retrofit opportunities currently requires on-site energy assessments that are inconvenient to homeowners, expensive, and are of variable accuracy, making it challenging to deliver cost effective retrofit opportunities at scale. For example, in the nation's top-ranked residential energy efficiency (EE) programs (Massachusetts), under 1% of households implement insulation, air sealing, or heating system retrofits per year, and a Northeast Utility estimates that about 30 to 35% of on-site audits ultimately result in major retrofits [3]. Finally, customers rarely get feedback on realized savings from energy conservation measures (ECM) beyond energy bills, and utility EE programs do not learn of potential large-scale field problems with ECMs until after completing costly EM&V studies, years after ECM implementation.

There is a need, therefore, for a scalable tool for remote energy assessment (REA) of homes. The massive deployment of communicating thermostats (CTs) provides a possibility for REA by analyzing the associated interval data on indoor temperature and HVAC runtime. Several organizations have studied this opportunity since early 2010-s as evidenced by patents (rather than peer-reviewed publications), e.g., [4-6]. However, most are geared towards relatively small energy efficiency improvements rather than identification and characterization of the three major retrofit categories. Only few references are devoted to evaluating residential building thermal performance, e.g., [7-8], but their underlying approach is based on correlating the slope of the room temperature curves obtained during thermostat setbacks with the outdoor temperature. This slope is actually an integral parameter depending on the building envelope U-value, external wall surface area, air infiltration and transient wall temperature. Accordingly, it is not optimal for evaluating building thermal performance and/or identification of specific ECMs.

The three major retrofit categories can be expressed as parameters of a mathematical model of building thermal response. The well-known grey-box models (GBMs), approximating an actual building by lumped elements, can provide a good compromise between model accuracy and the number of parameters, and there is a large body of research on designing and identification of grey-box models for buildings, e.g., [9-14] and references therein. However, several aspects of the problem at hand make a straightforward application of a GBM to REA challenging.

Since GBMs are mainly used for control applications, a majority of studies are concerned with individual commercial (office) buildings. For such buildings, it is relatively easy to obtain temperature measurements from building control systems and supplemental sensors during extended periods of inoccupancy (e.g., over weekends [11]). While the temperature measurements facilitate model design and reduces the number of parameters to be identified, measurements during unoccupied periods minimizes the problem of generally unknown internal heat gains [11]. For residential homes in REA applications, no systematic periods of inoccupancy exist, and typical available information is limited to building envelope dimensions (no characteristics of thermal *zones*, however), major construction material (e.g., wood), and approximate location (e.g., zip code). The control applications concentrate on prediction of state variables (temperature and HVAC status) rather than on identification of parameters per se. The inverse problem of parameter estimation in thermal models is essentially an ill-posed problem, which indirectly implies that large variations of parameter values can result in small changes of the state variables. In other words, the identified model can work well for temperature prediction even with parameters that are significantly different from the ground truth values.

Further, whereas air leakage effect is somewhat less important in commercial buildings, in US residential buildings infiltration accounts for 28% of total heat losses, which is larger than or comparable to wall- (19%), windows- (26%), or foundation- (19%) heat losses [15]. Moreover, infiltration models used in GBMs seldom include both wind and stack effects (which are equally important [16]), e.g., ref. [14] only includes the wind effect.

Lastly, the external and internal heat gains (or losses) that can be conventionally modeled (external - [9]) or estimated (internal - [11]) for commercial buildings are difficult to model for residential buildings in REA. Solar gain modeling is particularly challenging, due to the lack of both true local solar irradiation data and relevant building data (orientation, windows).

In this paper, we present a methodology to estimate the building insulation level, heating system efficiency, and building airtightness from CT data for homes with single thermostat and gas furnace. This methodology development is part of a larger project in which both CT and onsite energy assessment data for about 300 US homes will be obtained and analyzed. The proposed methodology is based on a GBM that yields a set of differential equations for lumped building elements. Our approach addresses the above challenges as follows:

1) We use the limited CT data to characterize *both* a zone and a home. Whereas a second-order GBM is generally considered to be accurate for the zone description [12], the lack of zone information leads to partial estimates of the insulation level/air tightness and makes HVAC efficiency estimation impossible. Assuming an even distribution of furnace-based heat supply throughout a home with single thermostat, we still apply a second-order GBM for a home. To compensate for the GBM coarseness and associated overfitting problem in this case, we propose to restrict the parameter search space by using analytical correlations obtained for a home over the entire heating season.

2) We model both wind and stack effects in air leakage.

3) We use data collected at nighttime, when external and internal heat gains/losses are minimal, to estimate the basic model parameters. Whereas we then estimate these gains using overall data somewhat similarly to Ref. [11], we address the model-caused uncertainty of these gains by modeling them using a statistical approach with conventional predictors.

The proposed approach is explained in the remainder of this paper.

2. APPROACH AND RESULTS

Figure 1 presents the main six steps of the proposed method.

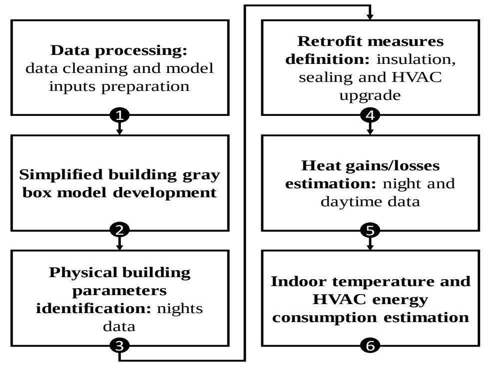

Figure 1. The six steps of the proposed methodology

2.1 Interval thermostat data

Available CT data usually comprise indoor temperatures and relative humidity, HVAC runtime, time stamp, wind speed and outdoor temperature. The two latter variables are usually not measured on site but pulled from the nearest weather station. Depending on CT manufacturer, the data can be, e.g., recorded at 5 minute intervals, with temperature resolution of about 0.1°F and time resolution for HVAC runtime of 1 second.

Two buildings with different energy performance characteristics were considered in this study. The CT data were obtained over 2016/2017 heating season. The available ground truth was partial: building 1 had a higher insulation level than building 2. Data cleaning included filling the missing values (by inter-/extrapolation) and reconstructing the binary on/off HVAC signal at 15 s resolution.

2.2 Zone thermal model

For a building zone, we propose the following second-order GBM:

$$C_r \frac{dT_r}{dt} = Q_{HVAC} + Q_{fan} + q_{int} + U_w(T_w - T_r) + q_{inf} \qquad \textbf{Eq. 1}$$

$$C_w \frac{dT_w}{dt} = U_w(T_r - T_w) + U_w(T_a - T_w) + q_{ext} \qquad \textbf{Eq. 2}$$

$$q_{inf} = \rho_{air} c_{p,air} \left(C_1 V_{air}^{2.6} + C_2 |T_a - T_r|^{1.3}\right)^{0.5} (T_a - T_r) \qquad \textbf{Eq. 3}$$

$$Q_{HVAC} = RT * q_{hvac} \text{ and } Q_{fan} = RT_{fan} * q_{fan} \qquad \textbf{Eq. 4}$$

where variables T_r, T_w, T_a are respectively indoor, wall and outdoor temperatures, Q_{HVAC} is HVAC heat supply [Wt], Q_{fan} is HVAC fan residual heat supply, q_{int} internal heat gains/losses affecting directly T_r, q_{ext} is heat gains/losses affecting the walls externally and q_{inf} heat loss due to air infiltration, V_{air} is the wind speed and RT the HVAC binary on/off signal. RT_{fan} is the HVAC fan runtime on/off signal that was reconstructed based on the indoor temperature data. The parameters to be identified are C_r and C_W, which are respectively internal (air and equipment) zone and wall heat capacities, U_w overall enclosure heat transfer rate (assumed to be the same for inward and outward directions), C_1 and C_2, which are coefficients characterizing respectively the wind and stack effects, and q_{hvac} the heat supplied by the heating system.

2.3 Zone-based identification

To minimize the effect of unknown external and internal heat gains/losses, we consider nighttime (i.e., recorded from 12 am to 5 am) data for estimation of the basic parameters. This is similar to Ref. [11] that uses weekend data for an office building; however, multiple nights might be needed to compensate for the lack of excitation during a single night period.

The continuous-time differential equations describing the heat transfer between the building zone and its surroundings form a nonlinear state-space representation. Its parameters to be identified are presented in section 2.2. The state variables are T_r and T_w, the inputs are T_a, RT, RT_{fan}, $q_{int/ext}$ and V_{air} and the output is T_r.

The differential equations are discretized and an identification process is then applied to minimize the mean square error (MSE) between actual and predicted output using a MATLAB toolbox [17]. Note that we obtained a closed-form solution to a simpler second-order GBM and estimated its parameters by curve fitting earlier [18]. Although that approach does not involve discretization and accordingly is more accurate, it cannot be used with the nonlinear infiltration model, Eq. 3.

The initial lumped wall temperature was considered as a parameter to be identified too. Only identified parameters for which the fit, Eq. 5, is significant and MSE is low were considered for parameters selection.

$$fit = 100\left(1 - \frac{\sum(T_r - \hat{T}_r)^2}{\sum(T_r - mean(T_r))^2}\right) \qquad \textbf{Eq. 5}$$

$\hat{T}_r$ is the model output, estimation of T_r.

Once the parameters are identified, the whole data set (nighttime and daytime) is considered to identify the heat gains/losses for successive time windows (4 hours in this study). In this step, all the parameters in the differential equations (Eq. 1 and Eq. 2) were those estimated by nighttime data except for q_{int} and q_{ext}, that were identified for the considered time windows. Further, as a simple yet meaningful machine learning method, we used a linear regression model to predict heat gains/losses. The candidate predictors were: time window number, month, week of the year, day of the week (workday or holiday), and difference between indoor and outdoor temperatures and relative humidity.

The statistical characteristics of the identified parameters are given in Table 1 and Table 2. Only the most relevant parameters were reported, other parameters are usually considered to be known.

Due to noise in data and, potentially, lack of excitation [12], the identified parameters are different for different nights. However, the main building physical parameters seem to converge to specific values, as an analysis of the scatter plots of each parameter versus the outside temperature (averaged per identification period) reveals. Indeed, the underlying model assumptions are likeliest to be valid when the difference between the outdoor and indoor temperatures is large. These converged values were close to the median values. Note that, due to the lack of zone geometry information, the parameter U is actually a product of the lumped wall U-value and wall surface area. The estimated value of Q suggests a relatively small zone size (see Section 2.4 below).

The results suggest that the second building's thermal conductance is 22% higher than that of the first building, which agrees with qualitative assessments of insulation levels (i.e., building #2 (bui#2) rated as medium and building #1 (bui#1) as high), assuming similar zone surface areas. An example of the estimated indoor and lumped enclosure temperatures is presented in Figure 2. These temperatures were simulated using the selected identified parameters (median values). The temperatures estimated with the GBM follow the dynamic of the actual ones with a satisfactory degree of precision. Note that the lumped wall temperature is close to that at the steady state (i.e., the average between indoor and outdoor temperature).

Once the basic physical building parameters of the zone are identified using the nighttime data, the internal heat gains of the zone can be estimated using the overall data. For the zone, we can assume the estimates to have no systematic deviations from the true values [11]. In this case, we can use the estimated heat gain/loss values to calculate the potential energy reduction of the building once it is retrofitted.

Table 1. Identification results for bui#1

Statistics	U_w [W/K]	Q_{HVAC} [W]
Max	49	1556
Mean	23	830
Median	23	807
Min	6	574

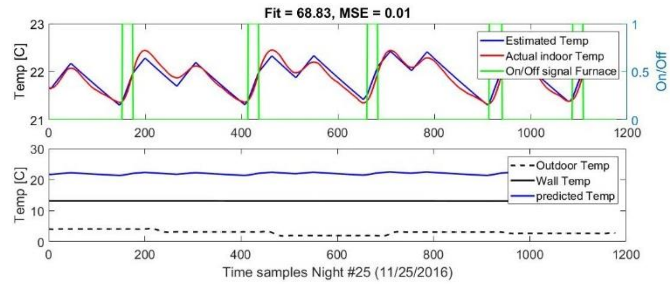

Figure 2. Example of indoor temperature evolutions, comparison between predicted and actual one, bui#1

Table 2. Identification results for bui#2

Bui#2	U_w [W/K]	Q_{HVAC} [W]
Max	61	887
Mean	30	564
Median	30	556
Min	14	357

2.4 Identification of overall home parameters

A single-thermostat home does not necessarily comprise a single thermal zone. Accordingly, the second-order GBM, Eqs. (1)-(2) with just five basic parameters can be too coarse to describe the thermal behavior of actual residential buildings.

On the other hand, using more equations and parameters may not be practical when external and internal heat gains/losses data exists and we need to consider relatively short periods of nighttime to compensate for this lack of data.

A solution to this problem can be a *restricted* second-order GBM, if we assume an even distribution of furnace heat over the interior floorspace of the building and also assume that the dynamics of the zone temperature follows that of the "average" indoor temperature. The main idea is that the experimental indoor temperature curve is no longer considered to be the "best" solution to which a GBM solution is conventionally fitted for parameter identification. Rather, the parameters of Eqs. (1)-(2) are estimated by fitting the GBM to an unknown yet "best" second-order solution, i.e., a hypothetical curve that may differ from the experimental ones. Although such a curve is unknown, we can assess some parameters that define this curve using overall approximated correlations. Such correlations, in turn, can yield confidence intervals for these parameters, that we propose to use to restrict the search space in the conventional GBM identification.

To illustrate this approach, consider periods of time when the heating system is "on." In US residences, gas furnaces are sized to enable quick temperature recovery, so that typical run times during non-recovery periods are short (up to 10-20 minutes). During these periods, the indoor temperature increases almost linearly in time (see also Figure 2). Since the selected window for nighttime (12 am to 5 am) typically excludes temperature recovery periods from nighttime setback, the durations of time "on" can be assumed to be short. If we neglect the infiltration and fan effects for simplicity, we will get the following two approximate correlations [19]:

$$C_{build}\frac{\Delta T}{t_{on}} = U_w A_w (T_a - T_r) + Q_{HVAC} \quad \textbf{Eq. 6}$$

$$0 = U_w A_w (T_a - T_r) + \frac{\sum_{t_{night}} t_{on}}{\tau_{night}} Q_{HVAC} \quad \textbf{Eq. 7}$$

where C_{build} is the thermal capacitance of the building air plus furniture/carpets, A_w is the overall external surface area (walls+roof+basement), U_w is the overall U-value (internal and external), ΔT is the thermostat deadband (typically 1-2 °F, can be easily deduced from CT data) and τ_{night} is the duration of nighttime excluding setback period. Eqs. (6) and (7) predict a linear dependence between the inverse run time (or total run time overnight) and corresponding indoor-outdoor temperature difference and the furnace heat supply. Accordingly, a linear regression model can be used for the corresponding parameter correlations and their confidence intervals.

Figure 3 shows the total time "on" overnight versus the temperature difference calculated for the two buildings considered earlier. The linear correlations are visible; the estimated slopes together with the overall external areas and typical U-values for medium and high insulation yield the following estimates for Q_{HVAC}: 23.2 kW for building 2 and 6.2 kW for building 1. Note that these estimates are much higher than the zone-based estimates (see Tables 1-2).

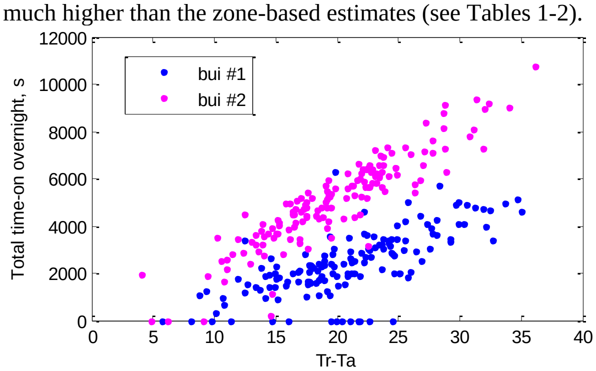

Figure 3. Correlations between total time-on overnight and T_r–T_a for two homes

With this coarse-grained approach, the external and internal heat gains/losses can no longer be estimated using the identified parameters due to the model error. However, using a statistical approach with such predictors as time of the day, day of the week and solar irradiation at weather station, we shall be able to separate the predictable portion of the estimated gains. The work on the restricted GBM and modeling of external/internal heat gains is currently in progress.

3. CONCLUSION AND PERSPECTIVES

In this paper, preliminary results of a research project that aims to develop, demonstrate and validate an energy performance assessment methodology to remotely evaluate and identify residential retrofit opportunities to reduce space heating energy consumption customized to individual homes are presented. Ongoing project work focuses on improving accuracy through machine learning techniques and physical model improvements, and on accurately estimating HVAC energy consumption. We will then validate the methodology on a large data set of different homes with different energy performance parameters.

4. ACKNOWLEDGEMENTS

This work was supported by DOE contract DE-EE0007571. The authors are grateful to Dr. Kurt Roth of Fraunhofer for valuable discussions and to Peter Klint of Eversource and Rich Wester of National Grid for providing anonymized CT and audit data.

5. REFERENCES

[1] K. Roth, "Using Communicating Thermostats to Automize, Customize, and Scale Home Energy Assessments," in *Better Buildings Summit. U.S. Department of Energy*, Washington, DC, 2017.

[2] "Building Energy Code Resource Guide: Air Leakage Guide," Building Energy Codes, 2011.

[3] "Massachusetts Energy Efficiency Advisory Council Annual Report".

[4] J. Steinberg and S. Hublou, "System and Method for Evaluating Changes in the Efficiency of an HVAC System". USA Patent 8,019,567 B2 .

[5] L. Spicer et al., "Thermal Time Constraints for Demand Response Applications". US Patent Patent 0016524 A1, 2012.

[6] A. Fadell et al., "HVAC filter monitoring". USA Patent 9,092,040 B2, 2015.

[7] E. Goldman et al., "Are thermostats the new energy audits?," Efficiency Vermont, 2014.

[8] G. Newsham et al., "Remote energy auditing: Energy efficiency through smart thermostat data and control," in *ECEEE Proceedings*, 2017.

[9] R. Kramer et al., "Simplified thermal and hygric building models: A literature review," *Frontiers of Architectural Research*, vol. 1, no. 4, pp. 318-325, 2012.

[10] M. Gouda et al., "Low-order model for the simulation of a building and its heating system," *Building Services Engineering Research and Technology*, vol. 21, no. 3, 2000 .

[11] Q. Hu et al., "Building model identification during regular operation - empirical results and challenges," in *American Control Conference*, 2016.

[12] Y. Lin, et al., "Issues in identification of control-oriented thermal models of zones in multi-zone buildings," in *51st IEEE Conference on Decision and Control*, Maui, Hawaii, USA, 2012.

[13] Harish, V.S.K.V and A. Kumar, "Reduced order modeling and parameter identification of a building energy system model through an optimization routine," *Applied Energy*, vol. 162 , p. 1010–1023, 2016.

[14] M. Siemann, Performance and Applications of Residential Building Energy Grey-Box Models, Ph.D. Thesis: University of Mariland, 2013.

[15] J. Huang et al., "Residential Heating and Cooling Loads Component Analysis," LBNL-44636, 1999.

[16] I. Walker, "Private Communication," 2017.

[17] L. Ljung, System Identification Toolbox™ User's Guide, Natick: The MathWorks, Inc., 2017.

[18] M. Zeifman and K. Roth, "Communicating Thermostats as a Tool for Home Energy Performance Assessment," in *IEEE ICCE*, 2016.

[19] M. Zeifman and K. Roth, "Residential Remote Audits: Estimation of Building Thermal Parameters Using Interval Energy Consumption Data," in *ACEEE Summer Study on Energy Efficiency in Buildings*, 2016.