

A Data Analysis Approach for Diagnosing Malfunctioning in Domestic Space Heating

Seyed Amin Tabatabaei

Behavioural Informatics Group
Computer Science Department, VU University Amsterdam
Amsterdam, The Netherlands
s.tabatabaei@vu.nl

Abstract— Around one third of worldwide energy usage is for the residential section and 60% of the energy consumption in this domestic area is for space heating. Therefore, monitoring and controlling this part of energy usage can have a major effect on the overall energy consumption and also on the emission of green house gases. Smart thermostats can play an important role in achieving more economic energy usage in the residential section. This paper shows an innovative way to use data collected by a relatively simple thermostat to estimate the thermal properties of a building. Moreover, it will be shown that in case of malfunctioning and high usage, this method can be used to firstly find out that there is a problem, and also to identify the cause of the problem. This information can be used as feedback to the resident to help him or her to discover the problem, and also point at the cause of it. Applying this idea in a smart thermostat or a mobile application with access to the data gathered by a simple thermostat can encourage residents to limit their energy usage for space heating. And in case of a problem that leads to an increase in energy usage, it can understand its occurrence and diagnose the cause.

Index Terms— Smart homes, Smart thermostat, Space heating, Energy usage, Energy saving, Degree day.

I. INTRODUCTION

Residential energy usage represents about one third of global energy use, and around 60 percent of energy in our houses is used for water and space heating. Therefore it plays a key role in energy-related environmental problems. More specifically, in European countries, more than 40% of total energy usage is used in buildings [1]. Around 43% of the total energy usage in the European Union in 2006 was spent on heat related needs [2]. It is predicted that increasing demand for building services and high thermal comfort levels, together with the amount of time spent indoors, will increase the energy demand in buildings in the future [3].

Smart thermostats are used more and more to get insight in domestic energy usage and can play an important role in achieving more economic energy consumption. They can control the heating system to reduce the usage, for instance by decreasing the indoor temperature during sleep and absence periods. On the other hand, they can play an important role to increase the inhabitants' awareness about their usage, and en-

courage them to change their behavior, or their house, or both into a more sustainable situation.

Technological progress has led to an increased adoption of energy monitoring systems within households. There are different ways to use these monitoring technologies to encourage inhabitants to reduce their usage, e.g., by changing their energy consumption behavior or by increasing the insulation level of their houses. Many projects try to encourage people to have more green behavior by comparing their energy usage to others [4], [5]. However, if someone finds out that his or her usage is much higher than average, he or she may find some excuses like this: “my house is larger than average, so it is reasonable to use more energy to heat it up”; and this may be correct or it may not. Therefore, it is important to have a fair comparison. To do that, it is better to compare each house to houses with comparable thermal properties. The authors of [6] proposed a technique to estimate the characteristics of the house, by analyzing the data collected by a thermostat. These estimations can be used to cluster the buildings according to their characteristics, and compare the consumption of each house with similar ones. In [7], it is suggested to use the historical data of a thermostat to compare the usage of a house in two different time periods, with almost the same outdoor temperature. The approach proposed in [6] does not have such a constraint, and can be used for temporal comparison between any two periods with different outdoor temperature.

Temporal comparison can be used to discover the unusually high consumption in a house. In such a case, a smart thermostat can send feedback and warnings to the customer, e.g. “Energy usage is significantly higher than expected”. However, it would be difficult for households to discover the source of this over usage. It may happen because of different reasons, like changes in the behavior of residents in opening windows, or malfunctioning of heating system devices.

The current paper proposes a data analysis approach that not only can find the existence of over usage, but also is able to diagnose its cause. To achieve that, the data collected by a relatively simple thermostat are used to estimate the thermodynamic characteristics of the building (like in [6]), and by performing temporal comparison for these properties, both the existence of a problem and also its cause are diagnosed.

The proposed approach can be applied to a smart thermostat with communication interface. Or, it is also possible to apply it to a mobile or web application with access to the data gathered by a relatively simple thermostat.

Fig. 1 shows an overview of the different parts of the proposed approach. As depicted, this method is based on the analysis of sensory data, gathered by a simple thermostat (indoor temperature, outdoor temperature and gas usage). Firstly, the thermal characters of the house are extracted, based on the provided data in different periods (Section IV). Then, by temporal comparison of the extracted features, occurrences of malfunctioning and their possible reasons will be examined (Section V).

This paper is structured as follows: next section explains some basic theory of thermodynamics. Section III is about the simulation environment and the simulated model that is used to validate the proposed techniques. Section IV describes the approach to estimate the thermal characteristics of a house based on the collected data. Section V explains the approach for discovering the malfunctioning and its cause, by performing temporal comparison. Finally, the last section provided a discussion.

II. THEORETICAL BASIS

Heat transfer between a building and the outside cause changes in its temperature. In this section, some thermodynamics of this heat transfer are discussed.

A. Gaining energy

Buildings gain heat energy through several ways (e.g., residents' bodies, sunshine radiation, heating system).

- **Bodies of residents** are almost always warmer than the house, since the temperature of a regular house is usually less than 37° C. As a result, there is a continuous heat transfer from the bodies of people who live there, to the indoor air. However, since this is a relatively negligible part of heat gained by a house, it is ignored in this work.
- **Sunshine radiation** can be a noticeable part of the gained heat of a building during a sunny day. However, for the sake of simplicity, this source of heat is ignored in this paper. As a result, this work is more applicable for houses that do not receive much radiation from the sun during winter days.
- For most of the houses in the cold areas, a **heating system** is the main source of heating energy during winter time. Different buildings use different kinds of heating systems. For the regular gas-based or electrical heating systems, one form of energy (chemical, electricity) is changed into thermal energy. The performance of a system is an important parameter that directly affects the provided amount of thermal energy.

$$\text{Provided Energy} = \rho \times \text{Input Energy} \quad (1)$$

where ρ stands for the performance of the heating system and the input energy shows the amount of energy

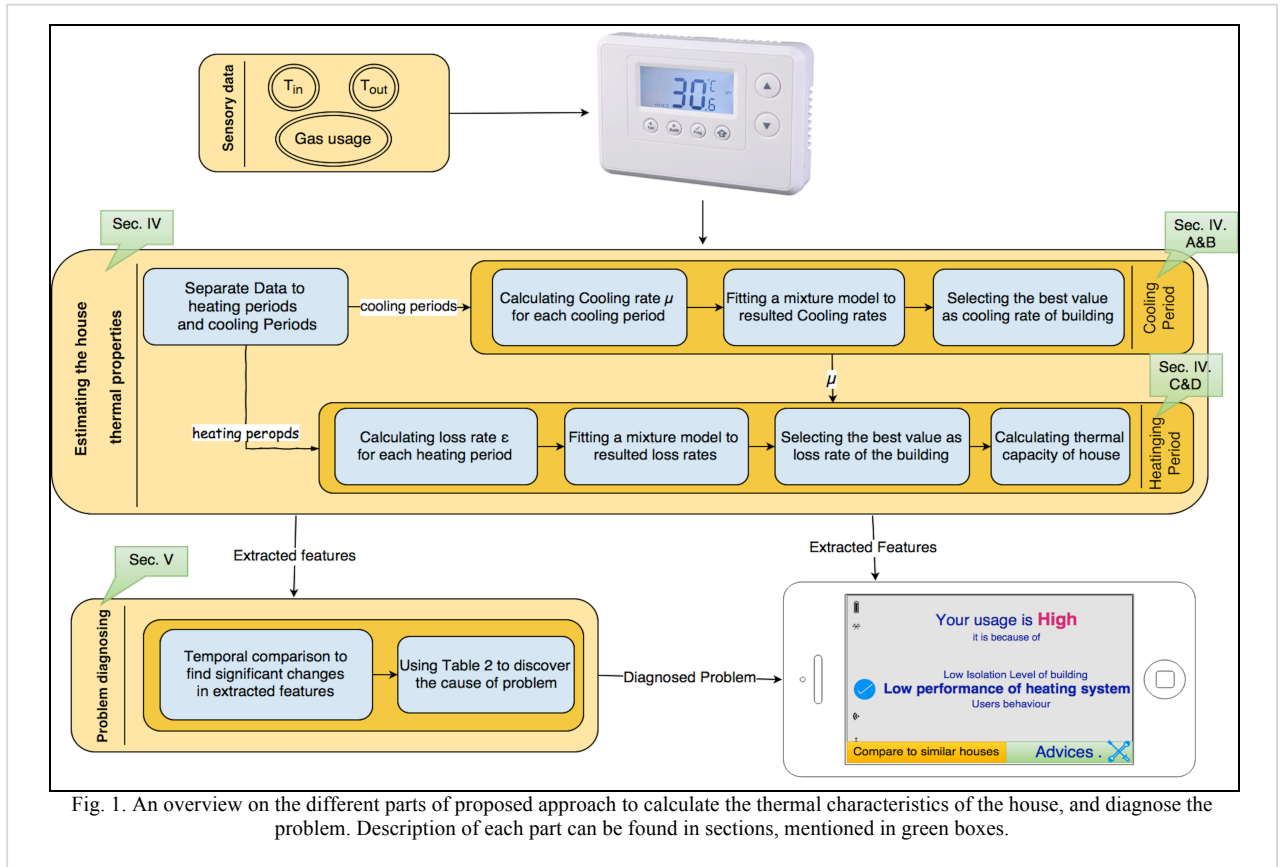


Fig. 1. An overview on the different parts of proposed approach to calculate the thermal characteristics of the house, and diagnose the problem. Description of each part can be found in sections, mentioned in green boxes.

that is provided to heating system in another form. For most of the existing thermostats, the amount of input gas is one of the sensed variables. It is easy to calculate the available energy in a particular volume of natural gas: the energy content of one cubic meter of natural gas is about 10 kWh.

B. Degree day

Degree day based energy analysis is a well known approach to quantify the relation between energy usage and the difference between outdoor and indoor temperature of a building (e.g.[8]–[11]). Through this way, it is possible to approximate the heating and cooling demand of a building (like [8], [12]). Even though the original definition of degree day is for a complete day (24h), but it is possible to define *degree day* for any period of time as well (like [12]):

$$DD_{\text{Period A}} = \int_{\text{Period A}} (T_{\text{in}} - T_{\text{out}}) dt \quad (2)$$

However, since in practical applications, the values of indoor and outdoor temperature (T_{in} and T_{out}) are not available continuously, this equation can be transformed into a discrete one:

$$DD_{\text{Period A}} = \sum_{\text{Period A}} (T_{\text{in}} - T_{\text{out}}) \Delta t \quad (3)$$

Here, the smaller Δt , the higher the accuracy.

C. Losing energy

During cold winter days, transfer of heat from houses to the outside takes place in a number of ways:

- **Conduction** refers to heat transfer that happens because of the adjacency of walls, roof, floor, etc. of a house and the outside air or soil. The material that is used in walls, roofs and floor has a big effect on conduction. The same holds for the area in m^2 of the walls, floor and roof: a more compact house, like with the shape of a cube or even a sphere, has lower conduction losses per m^3 volume than a less compact house, and the same applies to houses with multiple floors in comparison to one floor houses. The amount of energy lost in a time period through conduction has a linear relation with the difference between indoor and outdoor temperatures in that period (degree day of that period).
- **Infiltration** (air leakage) is unintentional introduction of outside air into a building, thereby replacing warmer inside air, typically through cracks in a building. The rate of this leakage of a building mostly depends on the state of the insulation of its walls, windows, roof, floor and it does not change dramatically from time to time.
- **Ventilation** means changing or replacing air in a building to decrease temperature or to replenish the fresh air. However, usually ventilation does not happen continuously, and it depends on the residents' behavior

(how frequently and for how long they keep the windows open).

According to these definitions, the rate of energy loss that happens due to the conduction and infiltration can be seen as a characteristics of the house. These are long-term characteristics, and do not change dramatically over time, except in case of improving or weakening the insulation level of the house. In contrast, ventilation is happening because of the behavior of residents in opening the windows and is a kind of short-term energy loss.

For all processes of conduction, infiltration and ventilation, the amount of energy lost in a period is assumed to be a linear function of the amount of degree days of that period. So, in general, for any period of time, we have:

$$\begin{aligned} \text{Energy Loss in Period A} &= \varepsilon_{\text{Period A}} \int_{\text{Period A}} (T_{\text{in}} - T_{\text{out}}) dt \\ &= \varepsilon_{\text{Period A}} \times DD_{\text{Period A}} \end{aligned} \quad (4)$$

Here ε is called the *loss rate*. For the periods of time that energy loss is just due to conduction and infiltration (not ventilation), energy loss depends on the long-term characteristics of the house. As a result, the value of the loss rate for such periods is practically constant, and it represents a characteristic of the house. In general, the better the insulation level of the house, the lower the value of the loss rate.

In contrast, the loss rates for the periods that ventilation occurs (in addition to conduction and infiltration) are not the same, as they depend on the rate of airflow through ventilation. In general, the loss rate for a period with ventilation is higher than the loss rate of a period without ventilation.

D. Thermal capacity C

The *thermal capacity* of a house, C , is a thermodynamic characteristic relating the heat added to (or removed from) it and the resulting temperature change. For a building, its thermal capacity has a direct relation with its size (in terms of volume in m^3). The larger the house, the higher the value of C .

The thermal capacity is defined as the ratio of the amount of heat energy transferred to an object and the resulting increase in temperature of the object. So, this concept shows how the temperature of house will change due to its net energy gain:

$$\text{Net Energy Gain} = C \times \Delta T_{\text{in}} \quad (5)$$

Here the net energy gain of house is the difference between the amount of energy provided to the house and the energy lost during a particular period of time. Moreover, ΔT_{in} shows the difference in temperature of the house at the beginning and the end of the considered period.

III. SIMULATED MODEL

To evaluate the proposed approach, because of lack of real datasets, a simulation program has been implemented. The system has to be analyzed in a dynamic state to monitor the trend of the house temperature and energy usage under different situations. The simulation of the system was done using TRNSYS

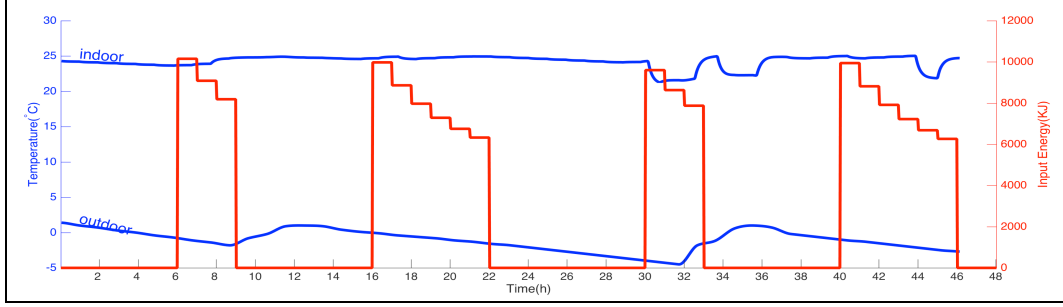


Fig. 2. Recorded values by thermostat for a sample period of 48 hours: values for indoor / outdoor temperature (per minute) and the input energy of the heater (per hour). The heating system is working twice a day, in the morning (from 6:00 to 9:00) and in the evening (from 16:00 to 21:00). Three significant drops in the indoor temperature are visible (around 30h, 34h and 44h); which are happening because of opening windows by residents (ventilation).

[13], a well-known simulation environment for dynamic building and equipments, including control strategies, occupants' behavior, and so on.

The house considered here refers to a traced one that has two neighbors from two sides. It has direct contact with outdoor air from the back and front side, and also from its roof. Moreover, its heating system is based on floor heating pipes connected to a gas-based heater and a pump that circulates water in a closed loop. A controller turns on the heater and the pump during two periods everyday, one in the morning (from 6:00 to 9:00) and one in evening (from 16:00 to 21:00). These periods are just for simulation, and it is not a constraint in the approach to have predefined periods for operating the heating system.

Since an important assumption of this approach is the low sun radiation during the winter, it is assumed that this house is located in Stockholm. Therefore, a dataset of the weather situation in Stockholm is used for this simulation.

The rate of air change due to infiltration is 10.8 m³/h and since it is related to the insulation level of the building, it does not change frequently. For ventilation, it is assumed that the residents may open a window in any minute with a uniform distribution, and will close it after a period of time with length from a normal distribution. During this period, the rate of air-flow is 108 m³/h. Table I shows the details of the simulated house in normal situation.

TABLE I. DETAILS OF THE SIMULATED HOUSE

Volume of the house	108	m ³
Area of the house	36	m ²
Height of roof	3	m
Infiltration	10.8	m ³ /h
Average temperature of neighbor 1	20	°C
Average temperature of neighbor 2	21	°C
Maximum power of heater	6.945	Kw/h
Maximum temperature of hot water	60	°C
Amount of ventilation, when window is open	108	m ³ /h
Simulation length	2	Months
Type of heating system	Gas-based floor heating	
Climate condition geography	Stockholm	
Probability of opening the window in a hour	0.04	
How long a window will stay open (from a normal distribution)	Mean = 2h Std = 1h	

It is assumed that the installed thermostat measures the inside and outside temperatures per minute, and the energy provided to the system is measured per hour (similar to the real dataset that is used in [6]). Fig. 2 shows the monitored values for a sample period of 48 hours.

IV. ESTIMATING THERMAL PROPERTIES OF THE HOUSE

In [6], a simple thermodynamic model of a house is introduced on the basis of capacity C and the loss rate ϵ of a house. In this section, an adapted version of this method is presented, and it is explained how to estimate the value of C and ϵ according to the collected data. To do that, the focus is on short periods of time (a few hours), in which the indoor temperature has a downward or upward trend.

A. Cooling Periods

Cooling period refers to a period of time that no heating energy is provided to the house via the heating system, and its temperature is decreasing. In this kind of period, the net energy gain of the house is equal to the lost energy (no gained energy). Therefore, Equation (5) can be rewritten in this form:

$$\text{Energy Loss} = -C \times \Delta T_{in} \quad (6)$$

By combining Equation (4) and Equation (6), for cooling periods we have:

$$C \times \Delta T_{in} = -\epsilon \times DD \quad (7)$$

In a cooling process the speed of the indoor temperature decrease is proportional to the difference between indoor and outdoor temperature. This proportion factor ϵ / C is called the cooling down rate, indicated by μ . So, in a cooling period

$$\mu = \frac{\epsilon}{C} = \frac{-\Delta T_{in}}{DD} \quad (8)$$

To calculate μ from Equation (8), we need to have data about the dynamics of T_{in} and T_{out} , so for any cooling period it

is possible to do that by using the data collected by the thermostat.

An important issue in selecting cooling periods is that no energy should be injected to the house during these periods. However, we know that in the common heating systems, even for the times that the heating system is off and no gas/electricity is used, some energy can still be transferred to the house from the hot water, which is still in radiators/pipes. As a result, for selecting proper periods as cooling periods, it is best to ignore the first couple of hours after the last gas usage of the heating system. It is assumed that the heat energy that is still in pipes after two hours is negligible. Therefore, there is one condition for selecting the cooling periods in the data: no energy provided by the heating system from two hours before starting time to the end of the cooling period. As an example, in Fig. 2, a cooling period is from 24h to 30h.

B. Selecting the best estimation of cooling rate μ

Given the number of time intervals in which cooling down takes place, for each house several estimated values for the cooling down rate are calculated. The reason for differences in calculated cooling rates for different cooling periods is the different situation of each period; for example, strong wind or no wind. Fig. 3.a shows the histogram of calculated values of the cooling rate (μ) for different cooling periods for the simulated house.

However, the cooling rate that is caused by conduction and infiltration (not ventilation) is assumed to be a fixed characteristic of a building. So, we have to choose the best answer among the values that came from different cooling periods. To do that, we have to diagnose the cooling rates that come from a period with or without ventilation. Then, the average of cooling rates for periods without ventilation can be chosen as a characteristic of the house.

It is clear that ventilation leads to much faster cooling

down, higher loss rate ϵ , and subsequently higher values of cooling rate μ . So, to be able to distinguish between cooling rates that are come from periods with and without ventilation, we assumed that these values come from a mixture model of two normal distributions. And, we used the expectation maximization algorithm [14] to find the two distributions which fit best to the data. Since the value of the loss rate for cooling periods without ventilation is lower than for periods with ventilation, the distribution with a lower value of the mean is selected as estimated cooling rates for the periods without ventilation (red distribution in Fig. 3.a.). The mean value of this distribution is chosen as the estimation for the cooling rate of the house.

C. Heating Periods

For periods that heating energy is provided and indoor temperature increases (for instance, in Fig. 2 from 6h to 9h), the provided energy is used for both replacing the lost energy and also to increase the temperature. We call these periods heating periods. For these heating periods, it holds that:

Provided Energy

$$\begin{aligned} &= \text{Energy Loss} + \text{Energy for Increasing Temperature} \quad (9) \\ &= \epsilon \times DD + C \times \Delta T_{in} \\ &= \epsilon \times DD + \frac{\epsilon}{\mu} \times \Delta T_{in} \end{aligned}$$

Therefore, by combining Equation (8) with this (9):

$$\epsilon = \frac{\text{Provided Energy}}{DD + \frac{\Delta T_{in}}{\mu}} \quad (10)$$

So, for each heating period it is possible to calculate the value of the loss rate ϵ according to the data gathered by the thermostat during that period, and the value of μ that resulted from the considered cooling periods.

D. Selecting best estimation of loss rate ϵ

Fig. 3.b shows the histogram of the calculated loss rates ϵ for different heating periods of the simulated house. For the same reasons as for the cooling rate μ , it is assumed that the values for ϵ come from a mixture model of two normal distributions. The average of the distribution with the lower value of mean (red distribution in Fig. 3.b) is chosen as the energy loss rate of the house, which is caused by conduction and infiltration, not ventilation. Consequently, the value of the thermal capacity C results from Equation 8.

V. DIAGNOSING THE CAUSE OF HIGH USAGE

There are some main reasons for high energy usage in a house. The first possibility is a significant decrease of performance of the heating system. It means that, even though a large amount of gas or electricity is given to the heating system, just a low fraction of it is transferred to heating energy for the house. The other possible causes for a (temporary) high usage can be high infiltration and ventilation. As it is described, in case of high infiltration, a continuous and long-term leakage of

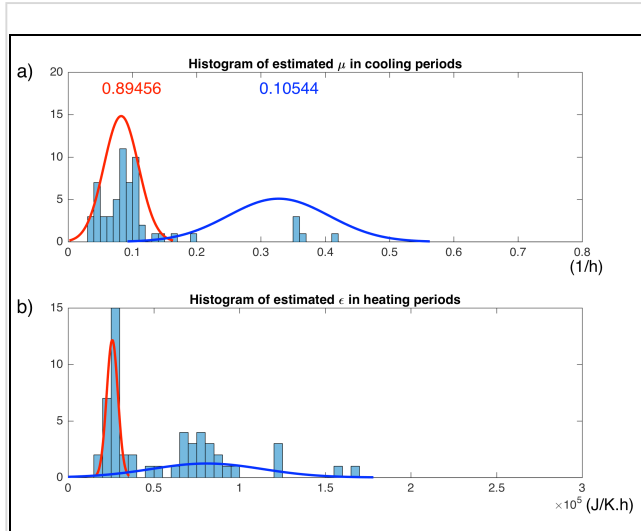


Fig. 4. Histogram of estimated values for μ and ϵ for a simulated house, in the normal situation. A mixture model of two normal distributions is fitted to them. Results: $\mu=0.0826$, $\epsilon=25784$, $C=311950$

- a) Calculated values for μ , resulted from cooling periods.
- b) Calculated values for ϵ , resulted from heating periods

warm air occurs. A high level of ventilation is caused by the behavior of residents in opening the windows more frequently and keeping them open for a longer time.

In the rest of this section, we will see how to diagnose the cause of high energy usage by comparing the results of the explained analysis. So, by using temporal data it will be possible to diagnose the cause of a temporary high usage level by comparing the results of this analysis to the historical data.

In each of the following subsections, one kind of malfunctioning is addressed. Then this malfunctioning is applied to the simulated model, and described analyses are applied on the resulting data to extract the thermal properties. At last, it is shown how to diagnose each type of malfunctioning by comparing the results to the results of a normal situation (Fig. 3).

A. Drop in the performance of heating system

When the heating system is not working properly, its performance will drop, and the provided energy through the heating system is lower than expected (dropping the value of ρ in Equation 1). Fig. 4 shows the results of the same simulation explained in Section 3, when the performance of the heating system is 75% of the original one.

Since no energy is provided by the heating system in the cooling periods, changes in the performance of the heating system do not have an effect on the cooling rate resulting from cooling periods. Therefore, by comparing Fig. 4.a and Fig. 3.a, the estimated cooling rates from cooling periods do not change significantly in case of malfunctioning of heating system.

In contrast, a drop in the performance of the heating system has a strong effect on the heating periods (Equation 10). In fact, because of the problem in the heating system, the “energy demand” will increase significantly; which leads to a significant increase in the estimation for both C and ϵ . In short, in this case, the values of both ϵ and C will increase dramatically, but the cooling rate (μ) will not. In the simulations, by comparing

Fig. 3 and Fig. 4, both ϵ and C are increased by 33%, and their ratio (cooling speed, μ) is not.

B. High infiltration by leakage

For all houses, some energy is lost through conduction and infiltration. Infiltration is usually caused by airflow through cracks in the windows frames; but it can be intensified via keeping a small window open for a long time (to change the indoor air by introducing more fresh air). Fig. 5 shows the results when the infiltration of the simulated house is increased from 10.8 to 27 m³/h.

By comparing Fig. 5 and Fig. 3, it is possible to discover the occurrence of high infiltration. This kind of malfunctioning changes the energy loss rate of house ϵ , while it does not have a significant effect on estimations of thermal capacity C . As depicted in Fig. 5, the selected value for thermal capacity has not changed significantly (8%), while the estimation of the loss rate has increased by 24%.

C. High Infiltration by leaving windows open

Another issue that may increase the thermal energy usage of a building is the behavior of residents in opening doors and windows frequently and leaving them open longer than normal. To assess this, the ventilation level of simulated model was changed; the average length of keeping a window open was increased to 5 hours. Fig. 6 shows the results.

However, this kind of energy loss does not occur for a long term, it does not relate to the insulation level of building, but it is a kind of ventilation. As a result, it does not have a significant effect on the estimations of characteristics of the building (ϵ and C), as is shown in the results in Fig. 6.

As described in Section 4.2, it is possible to distinguish between estimated leakage rate μ that occurs with and without ventilation. This is done by fitting a mixture of two normal distributions to the histogram of estimated leakage rates μ from

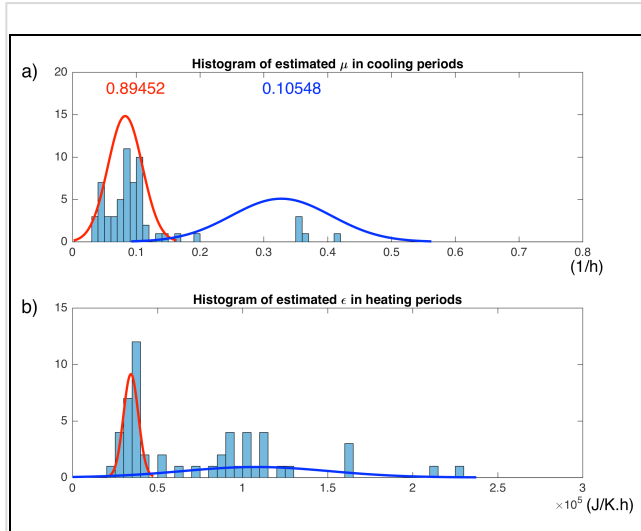


Fig. 5. Histogram of estimated values for μ and ϵ for simulated house, when the performance of the heating system is dropped for 75%. A mixture model of two normal distributions is fitted to them. Results: $\epsilon = 34375(+33\%)$, $C = 311950(+33\%)$

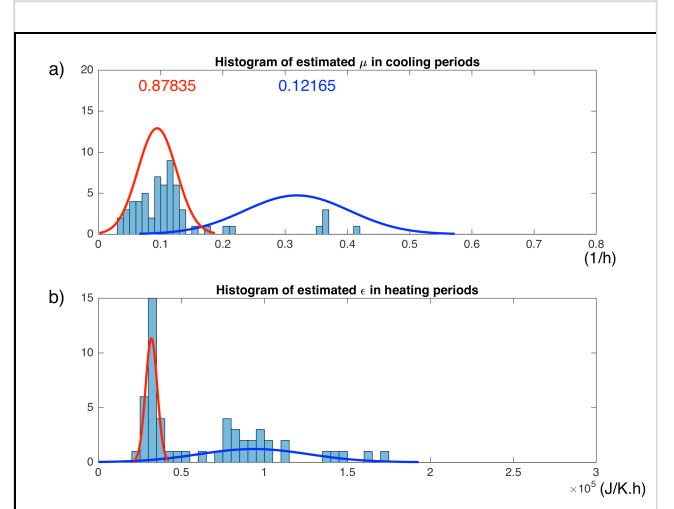


Fig. 6. Histogram of estimated values for μ and ϵ for simulated house, when more infiltration is happening, 27 m³/h. A mixture model of two normal distributions is fitted to them. Results: $\epsilon = 31961(+24\%)$, $C = 338405(+8\%)$

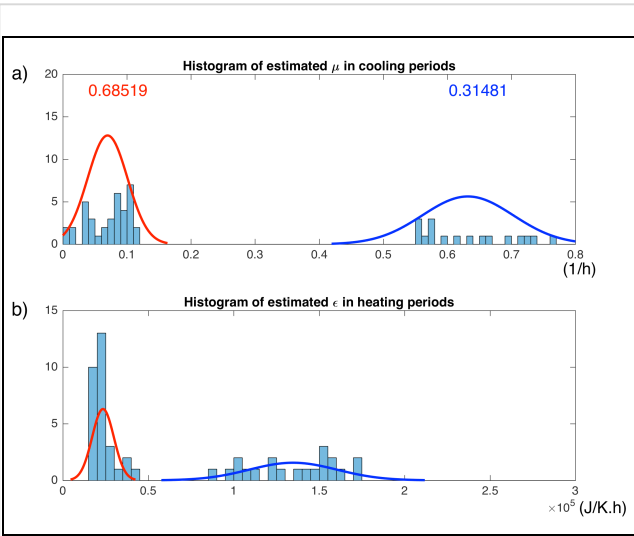


Fig. 7. Histogram of estimated values for μ and ϵ for simulated house, when more ventilation is happening. The average time of keeping the window open is increased to 5 hours.
A mixture model of two normal distributions is fitted to them.
Results: $\epsilon = 23494$, $C = 336552$

different cooling periods. So, more ventilation changes the situation of second normal distribution (blue distribution in Fig. 3.b). By comparing Fig. 6.a and Fig. 3.a, it can be seen that the mean of the blue normal distribution, and also the proportion of cooling rates that is covered by this distribution are changed, respectively from 0.32 to 0.63 (+96%) and from 0.105 to 0.314 (+199%).

To conclude this section, TABLE II summarizes the different causes of malfunctioning of heating process, in relation to the way to diagnose that through temporal comparison of the results by the proposed approach.

TABLE II. DIFFERENT CAUSES OF MALFUNCTIONING OF HEATING PROCESS, AND WAYS TO DIAGNOSE THEM BY TEMPORAL COMPARING OF THE RESULTS

Problem	Way to diagnose
Low performance of heating system	Both estimation of ϵ and C will increase significantly. But, the value of cooling ratio will not.
High infiltration	Los rate (ϵ) of house will increase significantly, but thermal capacity will not.
High Ventilation	No significant changes in ϵ and C . The mean value and the proportion of second normal distribution, which is fitted to the calculated cooling rates, will increase significantly.

VI. DISCUSSION AND FUTURE WORK

In an earlier paper [6], it has been addressed how a thermostat can be equipped with the means to estimate the basic thermal characteristics of a house over time: heat loss rate and heat capacity. It has been tested using a real data set including data of a variety of houses of different types.

A potential way to use the estimation about the thermal properties of a house is to apply them in a simple model, and use it to calculate the expectation of the model regarding to the energy usage of the house for a season. This expectation can be

compared to the real usage, and presented to the resident as feedback, for instance: “your usage is 10% higher than our expectation”. Or, the model can be used to generate if/then statements like this: “if you do not keep windows open for longer time periods, your consumption will reduce by up to 15%”.

In some papers (like [15]), it has been shown how it is possible to find the set point (goal temperature) of a house just by using its indoor temperature. It is used to measure the effect of lowering the set point of the house, and to estimate that to what extent a high set point affects the energy usage in a house. However, in the current work, we have focused on new elements. It is claimed that by using this approach, a thermostat or a mobile or web application that accesses the data gathered by a thermostat can discover the following kinds of malfunctioning in the heating process of a house:

- Reduced performance of the heating system
- High energy loss because of low insulation level of building
- High energy loss because of careless behavior of residents in opening windows frequently and for long time periods.

To validate this approach and do the analysis, a simulation environment of the house has been used. In a next step, some changes were done in the simulated house to obtain different kinds of malfunctioning that lead to higher energy usages. Finally, it was shown how it is possible to discover the cause of this malfunctioning by investigating the results in each case and compare them to the normal situation.

In general, it should be mentioned that this approach is based on some assumptions which do not apply to all houses and situations. Thus, its correctness depends on the accuracy of these assumptions:

- The temperature of the whole house is the same for any moment, which is an acceptable assumption just for small houses, like a studio. For large houses with several rooms, the temperature of different rooms may be different, and then only an average temperature can be used.
- The heating system is the only important source of heating for the building. It means that the heat transfer from the body of residents and by sun radiation to the house is ignored. This can be acceptable if just a few persons live in the house. Moreover, this approach is more applicable for areas that do not have much sun during the winter, like northern countries of Europe.
- The performance of the heating system does not change notably over time. This assumption is true for most of gas based and electric heating systems. However, in case of air source heat pumps, in which the performance is dependent on the outdoor temperature, this assumption is not true. However, by using some mathematical models (e.g. [16]), it is possible to estimate the performance for different moments, and calculate the characteristics of the house based on that.
- Energy provided by the heating system is just used for space heating. In some real cases, the provided energy

is used for both space and sanitation water heating. However, it may be possible to figure out in which periods the energy is also used for water heating. In [17], the summer data (when no space heating is done) were used to find a patterns for daily water usage of a house. Then, this pattern is used to have estimation about the fraction of energy that is used for water/space heating.

As potential future work, it would be useful to validate this approach in real thermostatic laboratories or in real houses.

In this paper, it is assumed that the only available data are indoor temperature, outdoor temperature and input energy of heating system. However, by spread of more modern thermostats, it is becoming common to gather more data through thermostat (like the temperature of hot water, at the beginning and the end of a cycle, the presence of residents, etc.). So, an interesting future work can be looking for different data analysis techniques for different levels of available data.

ACKNOWLEDGMENT

The author is grateful to Dr. Wojtek Kowalczyk from Leiden University, who provided useful input on the data analysis.

REFERENCES

- [1] M. De Nadai and M. van Someren, "Short-term anomaly detection in gas consumption through ARIMA and Artificial Neural Network forecast," in *Environmental, Energy and Structural Monitoring Systems (EESMS), 2015 IEEE Workshop on*, 2015, pp. 250–255.
- [2] B. Sanner, K. Ria, A. Land, K. Mutka, P. Papon, G. Stry-Hipp, and W. Weiss, "2020-2030-2050 Common Vision for the Renewable Heating & Cooling sector in Europe," *Eur. Technol. Platf. Renew. Heat. Cool. Rep.*, 2011.
- [3] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy Build.*, vol. 40, no. 3, pp. 394–398, 2008.
- [4] P. Petkov, F. Köbler, M. Foth, R. Medland, and H. Krcmar, "Engaging energy saving through motivation-specific social comparison," in *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 2011, pp. 1945–1950.
- [5] J. Pierce and E. Paulos, "Beyond energy monitors: interaction, energy, and emerging energy systems," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 665–674.
- [6] W. van der Ham, M.C.A. Klein, S. A. Tabatabaei, D. J. Thilakarathne, and J. Treur, "Methods for a Smart Thermostat to Estimate the Characteristics of a House Based on Sensor Data with Varying Extent of Completeness," in *Elsevier Energy Procedia*, 2015.
- [7] P. Petkov, F. Köbler, M. Foth, and H. Krcmar, "Motivating domestic energy conservation through comparative, community-based feedback in mobile and social media," in *Proceedings of the 5th International Conference on Communities and Technologies*, 2011, pp. 21–30.
- [8] M. De Rosa, V. Bianco, F. Scarpa, and L. A. Tagliafico, "Heating and cooling building energy demand evaluation; a simplified model and a modified degree days approach," *Appl. Energy*, vol. 128, pp. 217–229, 2014.
- [9] R. G. Quayle and H. F. Diaz, "Heating degree day data applied to residential heating energy consumption," *J. Appl. Meteorol.*, vol. 19, no. 3, pp. 241–246, 1980.
- [10] M. S. Al-Homoud, "Computer-aided building energy analysis techniques," *Build. Environ.*, vol. 36, no. 4, pp. 421–433, 2001.
- [11] T. Ueno, F. Sano, O. Saeki, and K. Tsuji, "Effectiveness of an energy-consumption information system on energy savings in residential houses based on monitored data," *Appl. Energy*, vol. 83, no. 2, pp. 166–183, 2006.
- [12] A. Durmayaz, M. Kadioğlu, and Z. Şen, "An application of the degree-hours method to estimate the residential heating energy requirement and fuel consumption in Istanbul," *Energy*, vol. 25, no. 12, pp. 1245–1256, 2000.
- [13] S. A. Klein, W. A. Beckman, J. W. Mitchell, J. A. Duffie, N. A. Duffie, T. L. Freeman, J. C. Mitchell, J. E. Braun, B. L. Evans, and J. P. Kummer, others, "TRNSYS 16—A TRaNsient system simulation program, user manual," *Sol. Energy Lab. Madison Univ. Wis.-Madison*, 2004.
- [14] G. McLachlan and D. Peel, *Finite mixture models*. John Wiley & Sons, 2004.
- [15] A. Rogers, S. Ghosh, R. Wilcock, and N. R. Jennings, "A scalable low-cost solution to provide personalised home heating advice to households," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, 2013, pp. 1–8.
- [16] S. A. Tabatabaei, J. Treur, and E. Waumans, "Comparative Evaluation of Different Computational Models for Performance of Air Source Heat Pumps Based on Real World Data," in *Elsevier Energy Procedia*, 2015.
- [17] S. A. Tabatabaei and J. Treur, "Analysis of Electricity Usage for Domestic Heating Based on an Air-to-Water Heat Pump in a Real World Context," in *2nd International Congress on Energy Efficiency and Energy Related Materials (ENEFM2014)*, 2015, pp. 587–596.