Disaggregation of Heating and Cooling Energy Consumption via Maximum a Posteriori Estimation

Antoine Tavant
Cédric Simi
Eco CO2
Lyon, France
e-mail: antoine.tavant@m4x.org

Pierre-Alexis Chevreuil
Pierre Costini
Eco CO2 Nice, France
e-mail: pierre.costini@centraliens.net
e-mail: pierrealexis.chevreuil@gmail.com

Joris Costes
Eco CO2 and *Univ. Paris-Saclay*,
ENS Paris-Saclay, CNRS, Centre Borelli,
Gif-sur-Yvette, France
e-mail: joris.costes@ens-paris-saclay.fr

***Abstract*—Estimating energy use in heating and air-conditioning systems is crucial for effective building energy management. This article introduces a new method combining the use of degree-days with the maximum a posteriori estimation statistical method to disaggregate heating and cooling energy consumption from other uses. Degree-days provide a reliable measure of the demand for energy needed to heat or cool a building, while a posteriori estimation offers a robust statistical approach to refine these estimates based on available data. A significant challenge addressed by this method is the need to accurately estimate the parameters of the model, which is achieved here by leveraging a comprehensive database. The method's efficacy is demonstrated through a case study of a building with one year of collected data, illustrating its practical application. Our findings underscore the method's potential to enhance energy management practices and guide future research in heating and cooling energy estimation.**

Keywords-Smart meters; Non-Intrusive Load Monitoring; NILM; Thermosensitivity.

I. INTRODUCTION

With the increasing deployment of smart meters in buildings, it is now possible to access aggregated energy usage data with unprecedented ease and accuracy [1]. This wealth of data opens up new opportunities for advanced analytical methods to improve energy estimation practices.

Smart meters usually collect the total building energy consumption data at a very low temporal resolution. For instance, in France the daily consumption is the only quantity provided by default by communicating meters, with finer sampling rate mainly depending on the supply contract and meter type. Individual users have the possibility to opt in to a finer 30 minutes sampling rate, but it is rarely selected. These low frequencies increase the difficulty to detect appliances [2]. Usual disaggregation techniques discard explanatory variables, such as weather and date, to focus on pattern and signature detection [2].

The management of energy use in Heating, Ventilation, and Air Conditioning (HVAC) systems is a critical aspect of building operations, influencing both cost efficiency and environmental impact. HVAC represent 38% of buildings consumption worldwide, and up to 60% in Europe [1]. Hence, accurate estimation of heating and cooling energy consumption is essential for optimizing energy use, reducing costs, and meeting sustainability goals. There are two families of methods to model building energy consumption [3]:

- **The forward methods** use the building description to develop a theoretical energy signature model,
- **The reverse methods** use collected consumption data to fit a consumption model.

Forward models are usually used when few or no data is available, such as during design and during the first few months of use. However, it has been shown that they show large discrepancies when compared to actual consumption, as the occupants' behavior can have a significant impact on the consumption [4]. This article focuses on reverse methods, available when a significant amount of data has been gathered.

In France, 37% of households use electricity for heating [5]. On top of that, the air conditioning energy use is expected to increase in the future [6]. Hence, the assessment of HVAC energy uses with smart-meters requires disaggregating both the cooling and the heating energy uses from the total consumption measured.

Figure 1 introduces the typical data used in this study. The data have been collected in one household near Lyon, France, with electrical heating and no cooling system. The energy consumption is gathered using the Linky smart-meter of Enedis [7]. The OpenWeatherMap API was used for the outdoor temperature [8]. The data is aggregated on a weekly frequency, from Monday to Sunday, to reduce the impact of the day of the week on energy consumption, considering the differences between workdays and weekends. The top panel of Figure 1 shows the temporal evolution of the total weekly electric energy consumed by the household. The bottom panel of Figure 1 presents the same data as the left panel, showing the outside temperature to energy consumption relation. One point in black corresponds to the week starting on Monday the 1st of January 2024, a week when occupants were not present. We can identify two regions:

- when the mean weekly outside temperature is above 16°C, the energy consumption seems independent of the temperature,
- when the mean weekly outside temperature is below 16°C, the energy consumption seems negatively proportional to the temperate.

This article introduces an innovative method for estimating HVAC energy use that leverages Degree-Days (DD) computation combined with Maximum a Posteriori (MAP) estimation

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

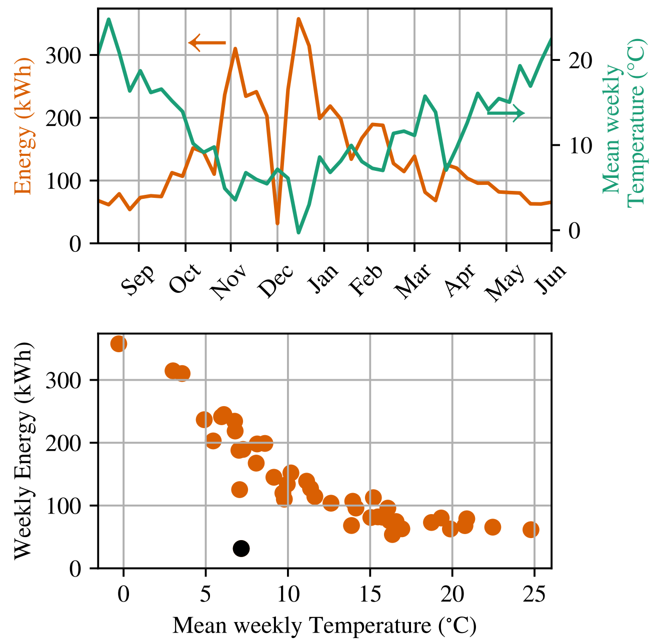

Figure 1. Data collected for one household: (top) the temporal evolution of the weekly energy (left axis) and the mean weekly outside temperature (right axis). (bottom) the weekly energy as a function of the mean weekly temperature. One point in black corresponds to an outlier.

which is more accurate than simpler approach such as a threshold when electricity is used for heating as well as other uses. DD is a well-established metric that quantifies the demand for energy to heat or cool a building, based on temperature deviations from a baseline [4]. MAP estimation, on the other hand, is a statistical technique that refines these estimates by incorporating prior knowledge and observed data to produce the most probable outcomes. By combining these two tools, we provide a way to estimate the energy usage of HVAC once enough data is collected.

To introduce the concept behind MAP estimation, consider the following analogy. Suppose a player rolls two dice and the goal is to determine the individual values of each dice. Without prior knowledge, this task is impossible. However, knowing that each dice can show a number between one and six, with a uniform probability distribution, the total number of possible outcomes is $\frac{6\times7}{2} = 21$, giving a 4.76% chance of guessing the correct pair. If the player provides the combined sum of the two dice, the estimation becomes significantly more manageable. For sums of 2, 3, 11, or 12, the possible combinations of the dice values are unique and known with certainty. Conversely, for sums of 6, 7, or 8, there are three possible pairs, resulting in a $\frac{1}{3}$ probability of guessing the correct pair. This analogy illustrates that by integrating prior knowledge (the uniform probability distributions of each dice) with an observation (the sum of the dice), the a posteriori estimation of each dice's value is substantially improved compared to the a priori estimation. In this study, the observation is the aggregated energy consumption recorded by a smart meter, which represents the sum of heating, cooling, and other energy uses. By applying MAP estimation, we can more accurately decompose this sum into its constituents, thereby enhancing the precision of our energy usage estimates.

A notable challenge in implementing this method is the need to estimate the prior distributions of the heating, cooling, and other uses needs, and whether they are correlated or not. This correlation is essential for precise estimation but can be difficult to determine directly. To overcome this limitation, we utilize a comprehensive database that provides the necessary correlation information, ensuring that the estimation method remains robust and reliable.

The following sections of this article detail the components and implementation of this new estimation method. We begin in Section II with an overview of DD computation and MAP estimation, followed by a discussion on addressing the correlation of random variables. Section III outlines the step-by-step process of integrating these components. Section IV aims to estimate a correlation coefficient from an extensive dataset. We then present in Section V a case study applying the method to the data shown in Figure 1, demonstrating its practical application and effectiveness. Finally, we discuss in Section VI the results, compare them with traditional methods, and offer insights into the implications and potential future developments.

II. Thermosensitivity Consumption Model

In this section, we introduce the thermosensitivity consumption model, which forms the foundation of our method for estimating HVAC energy use. The total energy consumed in a building over a given period, E, can be decomposed into three primary components:

$$E = E_{\mathrm{h}} + E_{\mathrm{c}} + E_{\mathrm{o}}, \tag{1}$$

where E_{h} is the energy required for heating, E_{c} is the energy required for cooling, and E_{o} represents all other energy uses not dependent on temperature variations, such as lighting, appliances, and other equipment.

A. Degree-Days

DD is a widely used metric to quantify the demand for energy needed to heat or cool a building [9]. Heating Degree-Days (HDD) and Cooling Degree-Days (CDD) are calculated based on deviations from a baseline temperature. The baseline temperatures depend on the building and occupants, and they need to be calibrated to maximize the performance of the model using historical data [9]. The duration of the integration period is at least one day, but can last longer. Practically, using a weekly sampling for the energy and the degree-days allows merging the differences related to the day of the week, such as weekends. HDD is used to estimate the energy needed to heat buildings. The integral formula to compute HDD is:

$$HDD = \int \left[T_{\mathrm{base,h}} - T(t)\right]^{+} \mathrm{d}t, \tag{2}$$

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

where $T_{\text{base,h}}$ is the heating baseline temperature, $T(t)$ is the outdoor temperature at time t, and $[\cdot]^+$ is the positive function

$$x^+ = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases} \tag{3}$$

Equation (2) assumes that the heating is switched on when the outside temperature is below $T_{\text{base,h}}$. CDD are the equivalent of HDD for cooling needs. Its integral formulation is:

$$CDD = \int [T(t) - T_{\text{base,c}}]^+ \, \mathrm{d}t, \tag{4}$$

where $T_{\text{base,c}}$ is the cooling baseline temperature. It's worth noting that other ways to compute HDD and CDD have been proposed in the literature. While we chose to use the integral formulation, other formulations may be used without changing the rest of the methodology or conclusions of the paper.

B. Degree-Days and Energy Consumption

Using the notation H for the Heaviside step function defined as

$$H(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases} \tag{5}$$

the energy required for heating and cooling can be modeled linearly with respect to HDD and CDD, respectively:

$$E_\text{h} = (\alpha_\text{h} \cdot HDD + \epsilon_\text{h}) \, H(HDD), \tag{6}$$

$$E_\text{c} = (\alpha_\text{c} \cdot CDD + \epsilon_\text{c}) \, H(CDD), \tag{7}$$

while other uses' energy consumption is modeled by

$$E_\text{o} = E_\text{baseline} + \epsilon_\text{o}, \tag{8}$$

where α_h and α_c are coefficients that represent the sensitivity of energy use to DD, and ϵ_h, ϵ_c and ϵ_o are random variables that capture the deviation from the model for heating, cooling, and other uses, respectively. The Heaviside step function ensures that ϵ_h and ϵ_c are added when the corresponding use is required.

III. Methodology for HVAC consumption disaggregation

This section describes the new proposed method to disaggregate the cooling and heating energy from the other uses.

A. Consumption model

The deviations ϵ_h, ϵ_c introduced in (7) and (6) account for various uncertainties and factors that affect energy consumption, but are not linearly correlated to the DD. These can include variations in building occupancy, operational schedules, sun illumination, and efficiency of the HVAC systems. Additionally, ϵ_o of (8) represents the variation in the energy use for non-HVAC purposes. Notably, they do not represent measurement error, which we assume to be nonexistent.

It is important to note that these variables can be correlated. For instance, an unusual higher occupancy of the building may increase the energy consumed by HVAC, as well as the energy consumed by other uses. On the other hand, a change in the AC temperature settings will increase the energy use for cooling, but not the one concerning other uses. We assume the deviations to the thermosensitivity model with normal distributions:

$$\left[\epsilon_\text{h}, \epsilon_\text{c}, \epsilon_\text{o}\right]^T = \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}\right), \tag{9}$$

with $\mathbf{\Sigma}$ the covariance matrix. We assume in the following that the sampling period is short enough so that heating and cooling cannot be present within the selected period:

$$\begin{cases} HDD = 0 & \text{if } CDD > 0, \\ CDD = 0 & \text{if } HDD > 0. \end{cases} \tag{10}$$

This hypothesis typically requires the sampling period to be shorter than a month. Hence, we can express the covariance matrix as

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_\text{h}^2 & 0 & \rho_\text{h}\sigma_\text{h}\sigma_\text{o} \\ 0 & \sigma_\text{c}^2 & \rho_\text{c}\sigma_\text{c}\sigma_\text{o} \\ \rho_\text{h}\sigma_\text{h}\sigma_\text{o} & \rho_\text{c}\sigma_\text{c}\sigma_\text{o} & \sigma_\text{o}^2 \end{pmatrix}, \tag{11}$$

where σ_x is the standard deviations of ϵ_x, with $x \in \{h, c, o\}$, and ρ_h (respectively ρ_c) represents the correlation coefficient between the ϵ_h (respectively ϵ_c) and ϵ_o. In the following section, we will detail the probabilistic inference method proposed to estimate the values of ϵ_h, ϵ_c and ϵ_o.

B. Inference

The goal of this section is to outline the method for estimating E_h, E_c, and E_o from a single measurement of total energy consumption E, given the HDD, CDD, and the prior distributions (9). Combining (1) with (7) to (8) and (10), we have:

$$E = E_\text{baseline} + \epsilon_\text{o} + \begin{cases} \alpha_\text{c} \cdot CDD + \epsilon_\text{c} & \text{if } CDD > 0, \\ \alpha_\text{h} \cdot HDD + \epsilon_\text{h} & \text{if } HDD > 0. \end{cases} \tag{12}$$

Determining the distribution of energy consumption between heating, cooling and other uses means that we have to estimate the values of $\epsilon_\text{o}, \epsilon_\text{c}$ and ϵ_h. As we assumed previously that the sampling period was chosen short-enough to ensure that heating needs are null when there are cooling needs and vice versa, we can study either of the 2 cases, and the result will be translatable to the other case. Thus, we start with the first case $HDD > 0$. The exact same derivation can be followed for the other case $CDD > 0$.

The total deviation of the measured energy E from the thermosensitivity model is

$$res = E - E_\text{baseline} - \alpha_\text{h} \cdot HDD = \epsilon_\text{o} + \epsilon_\text{h} \tag{13}$$

Hence, res is a random variable following the distribution

$$res \sim \mathcal{N}(0, \sigma_\text{h+o}), \tag{14}$$

with

$$\sigma_\text{h+o}^2 = \sigma_\text{o}^2 + \sigma_\text{h}^2 + 2\rho_\text{h}\sigma_\text{h}\sigma_\text{o}. \tag{15}$$

The MAP estimation of $\epsilon_\text{h}, \epsilon_\text{o}$ is defined as

$$\hat{\epsilon}_\text{h}, \hat{\epsilon}_\text{o} = \underset{\epsilon_\text{h}, \epsilon_\text{o}}{\arg\max}\, P(\epsilon_\text{h}, \epsilon_\text{o} | res), \tag{16}$$

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

with $\hat{x}$ the estimation of x and P the probability density function. Equation (16) means that the MAP estimation of $\epsilon_\text{h}, \epsilon_\text{o}$ are the values that maximize the likelihood function of $(\epsilon_\text{h}, \epsilon_\text{o})$, knowing res. Using the Bayes theorem,

$$P(\epsilon_\text{h}, \epsilon_\text{o}|res) = \frac{P(res|\epsilon_\text{h}, \epsilon_\text{o})P(\epsilon_\text{h}, \epsilon_\text{o})}{P(res)}. \tag{17}$$

Hence, as $P(res)$ doesn't depend on $\epsilon_\text{h}, \epsilon_\text{o}$:

$$\hat{\epsilon}_\text{h}, \hat{\epsilon}_\text{o} = \underset{\epsilon_\text{h}, \epsilon_\text{o}}{\arg\max}\, P(res|\epsilon_\text{h}, \epsilon_\text{o})P(\epsilon_\text{h}, \epsilon_\text{o}). \tag{18}$$

Recalling (13), (18) becomes:

$$\hat{\epsilon}_\text{h}, \hat{\epsilon}_\text{o} = \underset{\epsilon_\text{h}, \epsilon_\text{o}}{\arg\max}\, P(\epsilon_\text{h}, \epsilon_\text{o}). \tag{19}$$

The join probability density function of the bi-variable distribution

$$\begin{bmatrix}\epsilon_\text{h}\\ \epsilon_\text{o}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_\text{h}^2 & \rho_\text{h}\sigma_\text{h}\sigma_\text{o}\\ \rho_\text{h}\sigma_\text{h}\sigma_\text{o} & \sigma_\text{o}^2\end{bmatrix}\right) \tag{20}$$

is of the form

$$P(\epsilon_\text{h}, \epsilon_\text{o}) \propto \exp\left(-\frac{1}{2(1-\rho_\text{h}^2)}\left(\frac{\epsilon_\text{h}^2}{\sigma_\text{h}^2} + \frac{\epsilon_\text{o}^2}{\sigma_\text{o}^2} - \frac{2\rho_\text{h}\epsilon_\text{h}\epsilon_\text{o}}{\sigma_\text{h}\sigma_\text{o}}\right)\right). \tag{21}$$

Substituting (21) and (13) in (19), we obtain

$$\hat{\epsilon}_\text{h} = \underset{\epsilon_\text{h}}{\arg\min}\left(\frac{\epsilon_\text{h}^2}{\sigma_\text{h}^2} + \frac{(res-\epsilon_\text{h})^2}{\sigma_\text{o}^2} - \frac{2\rho_\text{h}\epsilon_\text{h}(res-\epsilon_\text{h})}{\sigma_\text{h}\sigma_\text{o}}\right), \tag{22}$$

$$\hat{\epsilon}_\text{o} = \underset{\epsilon_\text{o}}{\arg\min}\left(\frac{(res-\epsilon_\text{o})^2}{\sigma_\text{h}^2} + \frac{\epsilon_\text{o}^2}{\sigma_\text{o}^2} - \frac{2\rho_\text{h}(res-\epsilon_\text{o})\epsilon_\text{o}}{\sigma_\text{h}\sigma_\text{o}}\right). \tag{23}$$

That can be solved using the quadratic formula, to

$$\hat{\epsilon}_x = res\frac{\sigma_x^2 + \rho_\text{h}\sigma_\text{o}\sigma_\text{h}}{\sigma_\text{o}^2 + \sigma_\text{h}^2 + 2\rho_\text{h}\sigma_\text{o}\sigma_\text{h}}, \tag{24}$$

with x either h or o.

Thus, the MAP estimation of the heating energy $\hat{E}_\text{h}$ and of the other uses $\hat{E}_\text{o}$ is

$$\hat{E}_\text{o} = E_\text{baseline} + res\frac{\sigma_\text{o}^2 + \rho_\text{h}\sigma_\text{o}\sigma_\text{h}}{\sigma_\text{o}^2 + \sigma_\text{h}^2 + 2\rho_\text{h}\sigma_\text{o}\sigma_\text{h}}, \tag{25}$$

$$\hat{E}_\text{h} = \alpha_\text{h}\cdot HDD + res\frac{\sigma_\text{h}^2 + \rho_\text{h}\sigma_\text{o}\sigma_\text{h}}{\sigma_\text{o}^2 + \sigma_\text{h}^2 + 2\rho_\text{h}\sigma_\text{o}\sigma_\text{h}}. \tag{26}$$

A way to understand this estimation is to say that the deviation from the linear model, noted res, is split in two parts, one attributed to the heating energy usage, and the second to the other uses. The relative importance of the two parts depends on the relative value of the standard deviations σ_h and σ_o and the correlation coefficient ρ_h. The same process is followed to derive the estimation during a period with $CDD > 0$.

C. Estimation of the model parameters

In order to apply the inference proposed previously, the values of the different parameters are needed. They depend on the building details, HVAC system and the behavior of occupants. Hence, they need to be estimated from measured data following a reverse model approach [4].

a) Other uses: We can estimate E_baseline and σ_o from historical data when both $HDD = 0$ and $CDD = 0$:

$$\hat{E}_\text{baseline} = \langle E\rangle_{HDD=0,CDD=0} = \frac{1}{N}\sum_{i=1}^{N} E_i, \tag{27}$$

$$\hat{\sigma}_\text{o}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(E_i - \hat{E}_\text{baseline})^2. \tag{28}$$

with N the number of data points where $HDD = 0$ and $CDD = 0$.

b) Thermosensitivity models: We can estimate σ_h and α_h from the periods where $HDD > 0$. The estimation of α_c and σ_c follows the same logic with the periods where $CDD > 0$. A linear model is fitted such that

$$E - E_\text{baseline} \sim \alpha_\text{h}\cdot HDD, \tag{29}$$

to find $\hat{\alpha_\text{h}}$ the estimation of α_h. The standard deviation of the residuals of the linear model res is estimated with

$$\hat{\sigma}_\text{h+o}^2 = \frac{1}{N_h - 1}\sum_{i=0}^{N_h}\left(E_i - E_\text{baseline} - \alpha_\text{h}\cdot HDD_i\right)^2, \tag{30}$$

with N_h the number of points with $HDD > 0$. We can then use (15) to find σ_h.

However, the value of ρ_h is unobservable, as neither E_h nor E_o are measured. Two options are available:

- estimate σ_h with the extreme values $\rho_\text{h} = -1$, $\rho_\text{h} = 0$ and $\rho_\text{h} = 1$ to compute a lower and upper bounds,
- use a benchmark value estimated elsewhere, on a similar building, for instance.

The next section investigates the second possibility to obtain an insight on ρ_h.

IV. Estimation of residuals correlation

In this section, we use the extensive database EDRP [10] to estimate the value of ρ_h. This database includes more than 8000 households in the United Kingdoms with both electric and gas energy consumption measured hourly between 2007 and 2011. From this dataset, 2319 households have been selected on the following criteria

- electricity consumption independent with HDD (linear regression with $p - value > 0.1$);
- gas consumption showing no significant outliers;
- more than 52 weeks of data available.

The first criteria mean that we can assume that the gas energy is mostly consumed for heating: E_h, and the electric energy consumption is related to the other uses: E_o. The two other criteria are used to improve the quality of the analysis. Figure 2 illustrates the analysis of one household. Figure 2(**a**) shows the available data, with the gas and electricity weekly energy, and HDD. Figure 2(**b**) shows the thermosensitivity models, and Figure 2(**c**) the relation between the gas and electricity residuals. From the residuals obtained, and assuming the gas is only used for heating and the electricity is only used for cooling, this household shows a correlation $\rho_\text{h} = 0.27$. Figure 3 shows the distribution of the correlations between

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

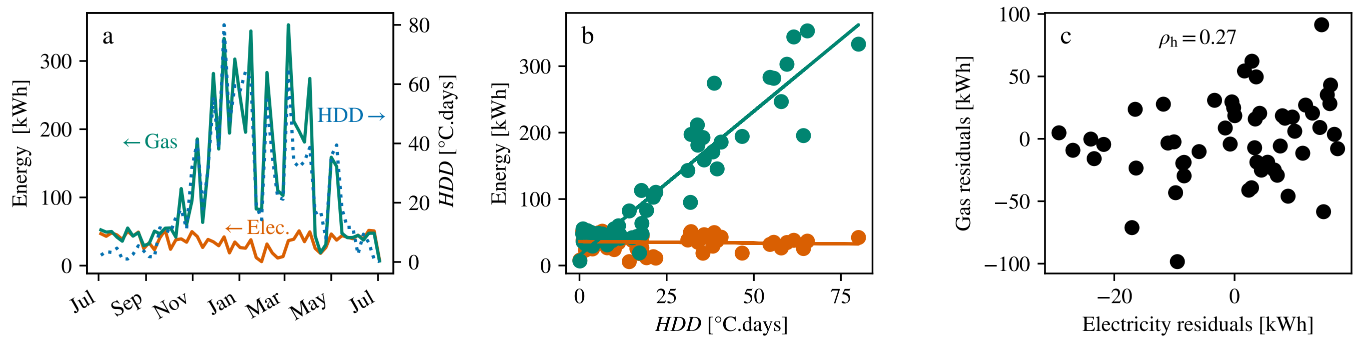

Figure 2. Illustration of one of the 2319 households analyzed. **a** : Time series of the gas (green) and electricity (red) weekly energy consumption (left axis), and HDD (blue dotted line, right axis). **b** : gas (green) and electricity (red) weekly energy consumption as function of HDD, markers represent the measured data showed in subplot **a**, and the solid lines represent the thermosensitivity models. **c**: scatter plot of the gas and electricity residuals to the models, meaning the difference between the markers and the lines of the subplot **b**. We compute for this household $\rho_{\rm h} = 0.27$.

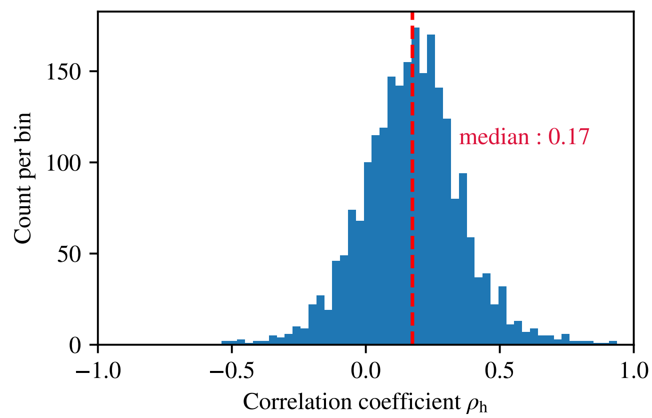

Figure 3. Distributions of the correlation between the electric and gas energy consumption residuals of the models, as illustrated in Figure 2, to estimate $\rho_{\rm h}$. A total of 2319 households has been analyzed.

the residuals for a selection of the 2319 selected households. As we can see, the values are spread over a large domain, with values of $\rho_{\rm h}$ ranging from -0.54 to 0.93. Fortunately, the distribution is relatively narrow, with most households exhibiting a weak positive correlation between $\epsilon_{\rm h}$ and $\epsilon_{\rm o}$. This means that using the median value $\rho_{\rm h} = 0.17$ is a relatively correct assumption for most cases.

V. Applications on a Real Building

In this section, we present the application of the proposed estimation to provide the disaggregation of the energy consumption for the data introduced in Figure 1. The HDD are computed with a reference temperature $T_{base,heating} = 17$°C. From the data, and using the median correlation coefficient measured in Section IV $\rho_{\rm h} = 0.17$, we can estimate that over one year:

- $E_{\rm baseline} = 69$ kWh
- $\sigma_{\rm o} = 8.9$ kWh
- $\alpha_{\rm h} = 1.8$ kWh/°C.week
- $\sigma_{\rm h+o} = 75$ kWh hence $\sigma_{\rm h} = 73$ kWh

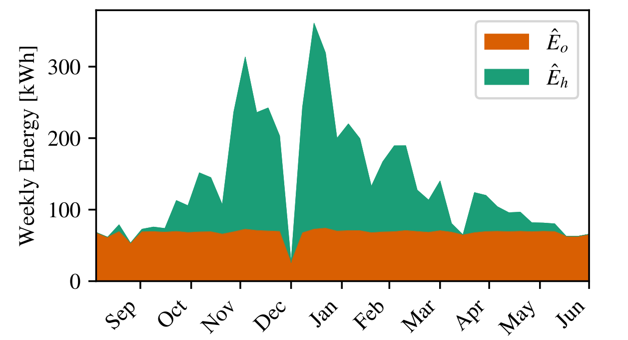

Figure 4. Disaggregation of the energy consumption of the household, presented in Figure 1: The orange area represents the Other uses, while the green area represent the heating energy. The value of $\rho_{\rm h} = 0.17$ is used.

Figure 4 presents the MAP estimation for the household whose data are shown in Figure 1 using the parameters estimated previously. We can see that $\hat{E}_{\rm o}$ remains relatively constant over the year, while $\hat{E}_{\rm h}$ increases in winter.

VI. Discussion

Several aspects of the method are discussed here-after.

A. Gains compared to alternative approaches

The usual approach of using a thermosensitivity model to estimate the heating and cooling energy needs is to discard the deviations to the model following the Maximum Likelihood Estimation (MLE) [11] :

$$\hat{E}_{{\rm o}MLE} = E_{\rm baseline} \tag{31}$$

$$\hat{E}_{{\rm h}MLE} = \alpha_{\rm h} \cdot HDD \tag{32}$$

$$\hat{E}_{{\rm c}MLE} = \alpha_{\rm c} \cdot CDD \tag{33}$$

This is an a priori estimation: the measurement of E is not used to update the estimates of the different components.

This approach has a Standard Error (SE) of $\sigma_{\rm o}, \sigma_{\rm h}$, and $\sigma_{\rm c}$ for $\hat{E}_{{\rm o}MLE}, \hat{E}_{{\rm h}MLE}$ and $\hat{E}_{{\rm c}MLE}$, respectively. One significant drawback is that $\hat{E}_{{\rm o}MLE} + \hat{E}_{{\rm h}MLE} + \hat{E}_{{\rm c}MLE} \neq E$, meaning that the disaggregation does not conserve energy.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

On the other hand, we can show that the SE of the MAP estimations are

$$SE^2_{\hat{E}_x} = \frac{\left(\sigma_x^2 + \rho_x \sigma_\mathrm{o} \sigma_x\right)^2}{\sigma_\mathrm{o}^2 + \sigma_x^2 + 2\rho_x \sigma_\mathrm{o} \sigma_x} < \sigma_\mathrm{o}^2 \quad (34)$$

$$SE^2_{\hat{E}_\mathrm{o}} = \frac{\left(\sigma_\mathrm{o}^2 + \rho_x \sigma_\mathrm{o} \sigma_x\right)^2}{\sigma_\mathrm{o}^2 + \sigma_x^2 + 2\rho_x \sigma_\mathrm{o} \sigma_x} < \sigma_x^2 \quad (35)$$

with x being h or c. In general, we have σ_o smaller than σ_x, meaning that the standard error of $\hat{E}_\mathrm{c}$ and $\hat{E}_\mathrm{h}$ are improved compared to the a priori estimates (32) and (33). On the other hand, $SE_{\hat{E}_\mathrm{o}}$ is bound by σ_x which is usually larger than σ_o, hence $\hat{E}_\mathrm{o}$ can potentially be worsened. This is due to the fact that the proposed a posteriori method validates (1), while the a priori estimation (31) does not, leading to the violation of the energy conservation.

B. Limitation due to the correlation coefficients

The values of ρ_h and ρ_c are difficult to obtain. They can vary wildly from one building to the other. In this article, we show that the correlation between the residuals of gas and electricity energy consumption vary from one household to the other, with a median value of $\rho_\mathrm{h} = 0.17$ obtained in the EDRP dataset.

C. Lack of validation

To the best of our knowledge, there are no large open source datasets with labeled energy uses for heating or cooling. The dataset used in Section IV only provides indirect estimation of heating energy when assuming strong hypotheses, and no cooling energy is available. Hence, it was not possible to measure the performance of the proposed method.

D. Impact of occupancy on the consumption

The proposed method considered the outside temperature as the only explanatory variable. However, the occupancy and other factors such as price of electricity can also impact the building consumption. Such factors could be included in a more elaborate model, using direct values or proxies of these factors. For example, occupancy variation could be approximated using the day of the week, since most households have a different occupancy behavior during weekdays and weekends. It is also possible to estimate the occupancy from the energy consumption data itself [11].

VII. Conclusion and Future Work

In this article, we introduced a novel method for estimating the energy use of HVAC systems in buildings, leveraging degree-days computation and Maximum a Posteriori (MAP) estimation. By decomposing the total energy use into heating, cooling, and other components, and incorporating the impact of uncertainties through correlation coefficients, we provide a method to estimate the energy consumed by each component.

We outlined a systematic approach to implement this method, starting with the estimation of non-HVAC energy use from historical data points where there is no heating or cooling demand. We then demonstrated how to estimate the thermosensitivity coefficients and standard deviations of the heating and cooling energy components. Finally, we utilized correlation coefficients to derive MAP estimates for heating, cooling, and other energy uses, ensuring that the total energy consumption aligns with observed data.

The practical application of this method was illustrated through a case study, showcasing its potential for enhancing building energy management. The case study, based on the measurement of the heating correlation coefficient ρ_h from a dataset of more than two thousand households located in the UK, demonstrated the method's efficacy.

Future work will focus on validating the approach with a labeled dataset that includes both heating and cooling energy. Additionally, measuring the cooling correlation coefficient ρ_c and incorporating other explanatory variables, such as the day of the week to account for occupancy effects, are expected to further improve model performance.

References

[1] M. González-Torres, L. Pérez-Lombard, J. F. Coronel, I. R. Maestre, and D. Yan, "A review on buildings energy information: Trends, end-uses, fuels and drivers," en, *Energy Reports*, vol. 8, pp. 626–637, Nov. 2022, ISSN: 23524847. DOI: 10.1016/j.egyr.2021.11.280.

[2] A. Petralia, P. Charpentier, P. Boniol, and T. Palpanas, "Appliance Detection Using Very Low-Frequency Smart Meter Time Series," en, in *Proceedings of the 14th ACM International Conference on Future Energy Systems*, arXiv:2305.10352 [cs, eess], Jun. 2023, pp. 214–225. DOI: 10.1145/3575813.3595198.

[3] M. Mehregan, A. Naminezhad, S. Vakili, and M. Delpisheh, "Building energy model validation and estimation using heating and cooling degree days based on accurate base temperature," en, *Energy Science & Engineering*, vol. 10, no. 9, pp. 3638–3649, 2022, ISSN: 2050-0505. DOI: 10.1002/ese3.1246.

[4] G. Matheron, "Large-scale validation of energy performance diagnostics via consumption measured by connected meters," fr, *Hello Watt*,

[5] *Key figures of housing*. The Data and Statistical Studies Service (SDES), 2022.

[6] D. F. Birol, "The Future of Cooling," en, *IEA*, 2018.

[7] Deliberation 2022-64 of the Energy Regulatory Commission on the draft decision on the incentive regulation framework for the Enedis advanced-metering system in the LV voltage field lower or equal to 36 kVA (Linky) for the period 2022-2024, Feb. 2022.

[8] OpenWeatherMap, https://openweathermap.org, retrieved: January 2025.

[9] J. A. Azevedo, L. Chapman, and C. L. Muller, "Critique and suggested modifications of the degree days methodology to enable long-term electricity consumption assessments: A case study in Birmingham, UK," en, *Meteorological Applications*, vol. 22, no. 4, pp. 789–796, 2015, ISSN: 1469-8080. DOI: 10.1002/met.1525.

[10] AECOM Building Engineering, *Energy Demand Research Project: Early Smart Meter Trials, 2007-2010. [data collection]. UK Data Service. SN: 7591*, 2014. DOI: http://doi.org/10.5255/UKDA-SN-7591-1.

[11] F. Culière, L. Leduc, and A. Belikov, "Bayesian model of electrical heating disaggregation," in *Proceedings of the 5th International Workshop on NILM*, Nov. 2020, pp. 25–29. DOI: 10.1145/3427771.3427848.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org