A data-driven model for the analysis of energy consumption in buildings

Nicola Borgato[1*], *Enrico* Prataviera[1], *Sara* Bordignon[1], *Roberto* Garay-Martinez[2], and *Angelo* Zarrella[1]

[1]University of Padova, Department of Industrial Engineering, 35131 via Venezia 1, Italy
[2]Universidad de Deusto, Instituto Tecnológico Deusto, Facultad de Ingeniería, 48007 Avda. Universidades 24, Bilbao, Spain

Abstract. Data-driven models are gaining traction in Building Energy Simulation, driven by the increasing role of smart metering and control in buildings. This paper aims to enhance the knowledge in this sector by introducing a practical method to analyse heating consumption. The methodology involves the analysis of hourly total heating demand and outdoor temperature measurements to create and calibrate Energy Signature Curves. Importantly, the building Energy Signature Curve is calibrated independently for each daily hour, resulting in a subset of 24 data-driven models. After calibration, a disaggregation algorithm is proposed to distinguish space heating from domestic hot water usage. The method also evaluates the building's thermal inertia, examining the correlation between the hourly global energy consumption and the outdoor air temperature moving average. It also presents a methodology for improving the DHW heat consumption model. The methodology is applied to a case study of 51 buildings in Tartu, Estonia, with complete yearly demand measurements from the district heating operator. Thanks to the hourly calibration approach, R2 is 0.05 higher on average than the yearly Energy Signature Curve approach. The difference between estimated and measured annual energy consumption is 8% on average, demonstrating the practicality and effectiveness of the proposed method.

1 Introduction

Buildings are primary contributors to energy consumption and greenhouse gas emissions, constituting 40% of the EU's total energy usage [1]. Within buildings, space heating (SH) and domestic hot water (DHW) systems collectively account for over 27% of final energy usage and 20% of greenhouse gas emissions in the EU, with SH contributing approximately 85% and DHW around 15% to heat demand [2].

Scientific literature shows a wide variety of methodologies for analyzing building energy consumption. The evolution of simulation tools for predicting energy consumption in buildings has seen various models developed over the years, each with unique features and intended applications. This study focuses on analyzing a single building's energy

* Corresponding author: nicola.borgato.1@studenti.unipd.it

consumption, with particular attention to models that can deal with DHW and SH heat consumption.

Single building models can be categorized into engineering approach (white box models), data-driven approach (black box models), and hybrid approach (grey box models) [3]. This work focuses on data-driven methods, which can be used in scenarios where building information is lacking but consumption data are available for calibration [4]. In particular, the Energy Signature Curve (ESC) [5] represents a robust statistical approach for determining a building's energy demand, especially regarding space heating demand during the winter season, highlighting its significant dependence on the outdoor temperature [5]. In addition, a slight correlation between DHW heat load and outdoor temperature has also been demonstrated [6], [7]. Notably, Ivanko et al.'s study in Norwegian hotels underscores that DHW energy demand is minimally affected by outdoor temperature, with guest count being the primary variable [8].

Pedersen's work [9] significantly contributed to understanding ESC construction methodologies by developing a mathematical procedure for identifying Change Point Temperature (CPT). Typically, the ESC delineates two segments for buildings provided with heating systems and no cooling systems. The CPT marks the division between the two segments, representing the threshold outdoor temperature for the beginning and end of the heating period. The segment with outdoor temperature below CPT indicates the relation between energy consumption and outdoor air temperature during winter when the SH system significantly contributes to heat use. Instead, the segment with a temperature above CPT represents the energy consumption related to the warm season, where SH is unnecessary, and only DHW is produced. It shows minimal dependence on outdoor air temperature [8].
Some studies separate temperature-dependent and temperature-independent segments by demand instead of based on outdoor air temperature [10]. This choice seems more convenient in models involving multiple weather parameters, such as outdoor temperature, global solar irradiance, wind speed, and wind direction.

The main issue concerns the difficulty in predicting the consumptions related to DHW production. Existing methods, relying on statistical models, clustering algorithms, and building simulations, have been explored in the literature, each with specific advantages and limitations [10], [11]. The main struggles are the availability of detailed information about occupants [13], excessive generalization, and lack of information on specific building use [12]. An additional issue derives from the availability of a model to simulate SH heat load but not for the DHW part. While ESC was used to represent SH, DHW heat consumption was obtained by subtracting the modelled SH from the measured energy values, creating high inaccuracies in this load estimation [8].

Various studies emphasized the importance of meticulous attention to activities and appliances for high-resolution DHW profiles [13]. However, challenges arise in non-residential settings due to the complexity and cost of gathering detailed occupant behaviour and equipment operation data. Comparisons with standard estimation models, such as those from ASHRAE, reveal significant discrepancies in DHW consumption profiles [14]. The differences manifest in uneven water usage distribution, higher peaks, lower drops, and distinct usage patterns compared to standard profiles. In response to these problems, [15] proposed an innovative approach for analyzing DHW heat consumption in nursing homes and highlighted the necessity of relying on actual experimental data rather than solely depending on standards. However, the observed enhanced oscillations in DHW demand during winter compared to summer, contrary to the expected steady consumption pattern, pose an issue in the accuracy of DHW and SH splitting models. A refined model is proposed in this paper to address this. It aims to mitigate these fluctuations by examining and standardizing DHW consumption profiles during summer.

1.1 Novelty

This paper aims to analyze existing buildings in-depth and develop methodologies to evaluate their behavior and consumption patterns using easily accessible data. First, the methodology exploits the ESC method to calibrate 24 independent models (one for each hour of the day) for each analyzed building. This method is applied to 51 buildings in Tartu, Estonia, where total heating consumption is available for an entire year.

Secondly, the proposed methodology refines the existing models for separating SH and DHW heat usage by introducing a novel approach for DHW profile estimation. One of the key issues with the existing approach is the absence of a dedicated model for DHW heat load. This work uses a clusterization technique to find similarities among different days regarding DHW consumption patterns during summer. This information is used to build standard profiles of DHW demand, which will be extended to the winter season. Such models are applied to existing case studies mentioned above.

2 Methodology

The presented methodology includes several steps aiming at processing the hourly total heating consumption demand of each analyzed building.

2.1 Hourly data segmentation

Ensuring data quality and pertinence for modeling is crucial [16]. Therefore, the initial phase involves pre-processing and selecting data inputs necessary for the analysis. Data cleaning involved meticulous cleaning of the dataset, deleting repetitions, fixing missing values and potential errors, and monitoring discrepancies. The second preliminary step involves input selection. The main advantage of using an ESC method relies on the necessity of a small amount of accessible and easily obtainable monitored variables. Only two input datasets are needed: yearly outdoor air temperature (Ti) and yearly global heating consumption (Ei), i.e., the total SH and DHW consumption of a building, with hourly resolution.

After these preliminary steps, the data are reshaped, subdividing the 8760 global energy demand values into 24 distinct vectors, each representing one hour of the day. Each vector, spanning from 0:00 am to 11:00 pm, comprises 365 values representing the global energy consumption for the corresponding hour across all days of the year (see Equation 1).

$$E_{0am} = [E_{0\ am,\ 1\ Jan},\ E_{0\ am,\ 2\ Jan},\ E_{0\ am,\ 3\ Jan}\ \ldots\ E_{0\ am,\ 31\ Dec},] \quad (1)$$

The primary objective of this subdivision is to independently calibrate the ESC for each vector, creating 24 different ESCs for each building instead of one. This process helps identify unusual energy consumption patterns or exceptional loads within specific day hours [17], [18]. This process exploits the analogy of consumption patterns of the same hour from different days. The precision in energy consumption representation through 24 different ESCs is higher because the data used to create the regression are much more similar among each other since they represent the same hour of the day but for a whole year. Thus, the variation caused by changes in setpoints or users' behavior is minimized, highlighting the dependence on the outdoor temperature.

2.2 Energy Signature Curve (ESC)

Typically, the ESC is a linear data-driven model, which represents the linear dependency between the heating demand of a building and the outdoor air temperature. Usually, ESC is

split into two segments by a critical outdoor temperature, the CPT, see Figure 1. When the temperature is below the CPT, the model captures data where both DHW and SH heat consumptions are present, while points above the CPT are located where only the DHW load exists. Typically, the former, i.e., the *winter season*, is characterized by a strong dependency between consumption and outdoor temperature, while the latter, i.e., the *summer season*, has a very weak dependency. The ESC can be represented using a piecewise function, as shown in Equations 2 and 3 [19].

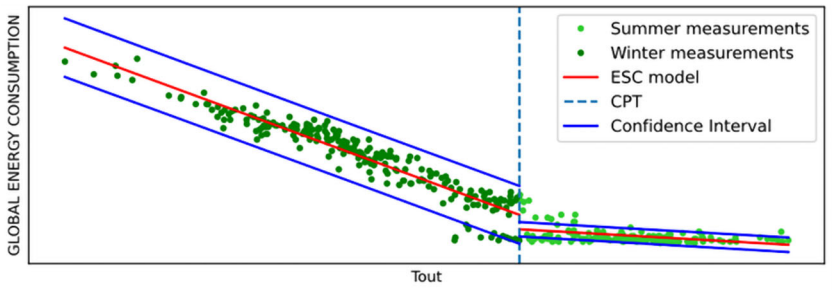

Fig. 1. Example of Energy Signature Curve.

$$Y = \alpha_1 + \beta_1 \bullet T_i \quad if\ T_i < CPT \tag{2}$$

$$Y = \alpha_2 + \beta_2 \bullet T_i \quad if\ T_i \geq CPT \tag{3}$$

α_1, α_2 (kWh) and β_1, β_2 (kWh/°C) represent coefficients derived from least-square linear regression of global energy demand (Ei) with outdoor temperature (Ti) in their respective zones. Y is the predicted energy demand obtained from the linear regression model. Visual continuity at CPT is achieved by equating Equations 2 and 3 at Ti = CPT [19]. As previously stated, this method is independently applied to each hour of the day (see Equation 1), providing 24 hourly-based ESCs.

CPT determination relies on several factors including user behaviour, climate, and location, as it determines the transition between the winter and the summer seasons. A mathematical iterative methodology has been applied to determine CPT [20]. The first regression starts with the two measured global heat consumption values with the highest outdoor temperature. The obtained slope coefficient (β) is stored with the lowest outdoor air temperature. The process is repeated considering for iteration n, the $n + 1$ group of points with the highest outdoor temperature, storing every time the slope coefficient β and the minimum temperature. When winter season points start to be included in the iteration, β slope starts to decrease due to the temperature dependency change, defining the CPT point. A higher absolute value of β means a higher dependence of energy consumption on outdoor air temperature.

2.2.1 Implementation of Confidence Intervals

After calibrating the ESC, data must be cleaned by outliers, which deviates from the linear regression. In this perspective, the method from [19] involving Confidence Intervals (CI) is applied to refine the Energy Signature Curve. CI calculation defines a confidence band outside which points are considered outliers. The estimated range is based on a multiple of the statistic's standard error (SE), determined based on a theoretical distribution. The t-Student distribution was used, applying a CI of 95%. To clarify the methodology presentation, Equation 4 reports the general law for calculating upper and lower values of CI. In Figure 1, it is possible to find a graphical example of CI.

$$C.I._i = Y_i \pm S_{res} \bullet t_{(1-a2)} \bullet \sqrt{(1 + 1/n + (x_i - x_{mean})/((n-1) \bullet S_x^2)} \quad (4)$$

Y_i is the predicted value of energy use, obtained from the linear regression model; S_{res} is the residual standard deviation concerning the regression line; $t_{(1-a2)}$ is the value from t-Student's criteria; n is the sample size; x_i is the actual values of the outdoor air temperature; x_{mean} is the mean value of the outdoor air temperature; S_x is the standard deviation of the outdoor air temperature. Confidence intervals are then applied iteratively to both the points below the CPT and points above, i.e., winter and summer seasons (Figure 1), thus identifying and removing outliers and recalibrating the ESC. The iterative process continues until all experimental values are contained within the confidence interval zone or the number of iterations reaches a predefined limit, preventing excessive exclusion of data points. This interruption control is crucial above the CPT part of the curve, where the confined CI area is small due to the line's horizontal nature.

2.2.2 Key Performance Indicators

In evaluating the accuracy of the ESC model, R^2 was chosen as the main Key Performance Indicator (KPI) [18]. R^2 quantifies the percentage of the variance in the data that the model's predictors can predict. The R^2 variable is computed using Equation 5.

$$R^2 = 1 - \Sigma (X_i - Y_i)^2 / \Sigma (X_i - X_{mean})^2 \quad (5)$$

Y_i is the value obtained from the prediction model, X_i is the measured data, and N is the number of observations. However, this KPI needs to correctly interpret points above CPT due to the horizontal behavior of the ESC in this part. For this reason, the yearly energy consumption deviation (YEC) was introduced as the second KPI, expressed as a percentage between the real consumption and the ESC prediction (Equation 6).

$$YEC = 100 \bullet |\Sigma X_i - \Sigma Y_i| / \Sigma X_i \quad (6)$$

2.3 Analysis of Temperature Lag

Analyzing buildings' thermal response through statistical methods like ESC is usually difficult due to the low physical interpretability of this type of model. However, exploring a building's temperature lag represents a fundamental perspective often neglected in energy consumption analyses. A possible solution for evaluating buildings' thermal inertia is to consider the dependence that heating consumption has on input variables from many previous time steps [21]. The method proposed by [19] is followed in the current work.

The methodology for analyzing temperature lag involves the calculation of an ESC considering as an independent variable a Temperature Moving Average (TMA) rather than the outdoor temperature of the same timestep. The TMA_{1-n} is calculated as the average temperature of the time window between the current timestep *1* and the *n-th* timestep before. Thus, 48 different ESCs are calibrated considering TMA_{1-2}, TMA_{1-3}, ..., TMA_{1-48}, providing 48 regression coefficients k_{ESC}. Plotting these k_{ESC} values against hour lag delineates the relationship between *TMA* and heating energy use, unveiling the optimal correlation between preceding outdoor temperatures and energy demand. The optimal correlation happens when k_{ESC} is minimum (maximum in absolute value).

2.4 Splitting of DHW and SH energy consumption

After calibrating and cleaning the hourly-based ESC for each building, a procedure is developed to split DHW and SH consumption from the global heating demand. The methodology adopted in this study starts from the ESC and follows the method proposed by [8]. The splitting operation was conducted separately for each hourly-based ESC. In particular, the minimum ESC value was identified during the summer season, i.e., within values with the outdoor air temperature above CPT [8], [19]. This minimum point is considered the DHW demand and subtracted from the whole global heating consumption, allowing the calculation of SH demand (Equation 7).

$$E_{SH} = Y(x) - min(Y(x)) \tag{7}$$

Where x represents the outdoor air temperature, Y the ESC points, and E_{SH} the values for space heating consumption. The difference between global heat demand and the calculated SH demand is the DHW demand [10].

Energy losses in the DHW circuit (E_{loss}) must also be estimated [11]. First, an average consumption profile for each month in the summer season is built. Every profile contains 24 values, representing the average heating consumption per hour in the considered month. Afterward, the mean energy demand between 0:00 am and 4:00 am is considered energy losses in the water circuit, assuming negligible DHW use during sleeping hours. The beginning and the end of the summer season (when space heating is not provided to the buildings) depends on the specific location of the case study, as national standards often regulate it. However, it is acceptable for most buildings in mild and cold mid-latitude climates to consider July and August as part of the summer season. The SH and DHW thermal load separation was conducted using Equation 10, where ETH represents the measured values of global heating consumption.

$$E_{DHW} = \begin{cases} E_{TH} - E_{SH} + E_{loss}\,, & if \;\; E_{TH} > E_{SH} \\ E_{loss}, & if \;\; E_{TH} \leq E_{SH} \end{cases} \tag{8}$$

The obtained values for DHW and SH must be positive to be physically acceptable. For this reason, if the model calculates a negative DHW heat demand, the value of SH heat demand is recalculated and reduced, leading to a DHW heat demand equal to zero. This approach assures the positivity of both DHW and SH heat demand while maintaining their sum equal to the overall heat usage for every measured value of the initial dataset.

2.5 DHW heat load clusterization and standard profile construction

A dedicated methodology to cluster summer DHW heat demand profiles is employed in this study [15]. This approach comprehensively examines DHW consumption profiles during summer to identify similarities among days and delineate standardized consumption patterns. The process exploits statistical tests for identifying similar consumption patterns between different days of the week, followed by statistical grouping to recognize minimum, medium, and peak consumption hours within daily profiles. Student's t-test and Fisher's criterion were used to find similarities among days. As a result of this analysis, the days of the week were subdivided into two groups. Within each group, a mean consumption profile was calculated. The hours with minimum, medium, and peak consumption were identified by using a procedure based on Student's t-test. Each hour was assigned to a specific zone of consumption (minimum, medium, peak) [15].

Successively, a summer standard profile was generated. Based on the consumption zones identified in the previous step, each hour is assigned to one of three load periods: minimum, medium, or maximum. The consumption of each hour is then set to the corresponding threshold value of heat load for that period. For example, the minimum load hours have the same consumption as the threshold value between minimum and medium heat consumption; the peak load hours have the same consumption as the threshold value between medium and peak heat consumption. In contrast, medium load hours have the mean consumption between the two thresholds.

The next step is converting the summer standard profile into a winter standard one. The summer standard profile was scaled up by the average ratio of DHW heat load in winter and summer. DHW heat consumption in summer corresponds to the global heat consumption in this period, while the previous splitting operation obtained DHW heat consumption in winter. The winter standard profile was used to adjust the DHW and SH profiles. DHW values that present a large difference from the values of the standard profiles are corrected by raising or lowering them by a certain fixed percentage. The subtracted or added part of the load was added to the SH value of the corresponding hour to keep the total heat consumption matching the metered values.

3 Case study

The analysis involves 51 buildings in Tartu, Estonia. The approach adopted for Tartu's dataset follows the methodology delineated in Section 2.

3.1 Tartu, Estonia

The case study regards the district heating network in Tartu, Estonia, specifically the Tarkon-Tuglase area. This area encompasses 54 buildings connected to the district heating system, serving residential, commercial, educational, and office buildings. The network covers 5.34 km and delivers up to 4.3 MW of heat, totaling around 8.2 GWh of annual consumption for both SH and DHW heat production. Smart energy meters installed in each building provide hourly measurements of the primary circuit's supply temperature, return temperature, and mass flow rate. The energy meter ensures that the measuring error remains below 5% in all the variables read [18]. The district network operator provided no information about buildings' structures or intended uses to avoid privacy issues. The study integrates weather data from a weather station at the University of Tartu. Outdoor air temperature is the only weather variable used in this paper.

4 Results

In Tartu's cold climate, during the winter season, the heating demand pattern is consistent for all 24 hours, as the heating system is always on. As reported in Figure 2, the maximum observed global heat demand reaches about 150 kWh in the analyzed building.

4.1 Energy Signature Curve (ESC)

For example, Figure 2 shows the result of the ESC reconstruction operation for a nocturnal and a daily hour of the day. On the CPT's left is the SH dominant part, i.e., the winter season, while on the right is the summer season. During the winter season, the dependence of the heat consumption on outdoor air temperature is more significant than during the summer season, when only DHW needs to be provided to the building. In blue, the boundaries of CI

are highlighted. In this case, most of the measurements are inside confidence intervals. This energy consumption pattern is typical of residential buildings.

However, using the CPT method to split the winter and summer seasons might not be accurate for the points relatively close to the CPT. Indeed, points on the lower part of the plot, on the left of CPT, present a remarkably similar consumption to the summer points despite being included in the SH-driven part. A detailed analysis of these points revealed that they belong to middle season days, a transitory period of the year between winter-spring and summer-autumn. Further development of this model proposes a method to allocate such energy consumption more accurately.

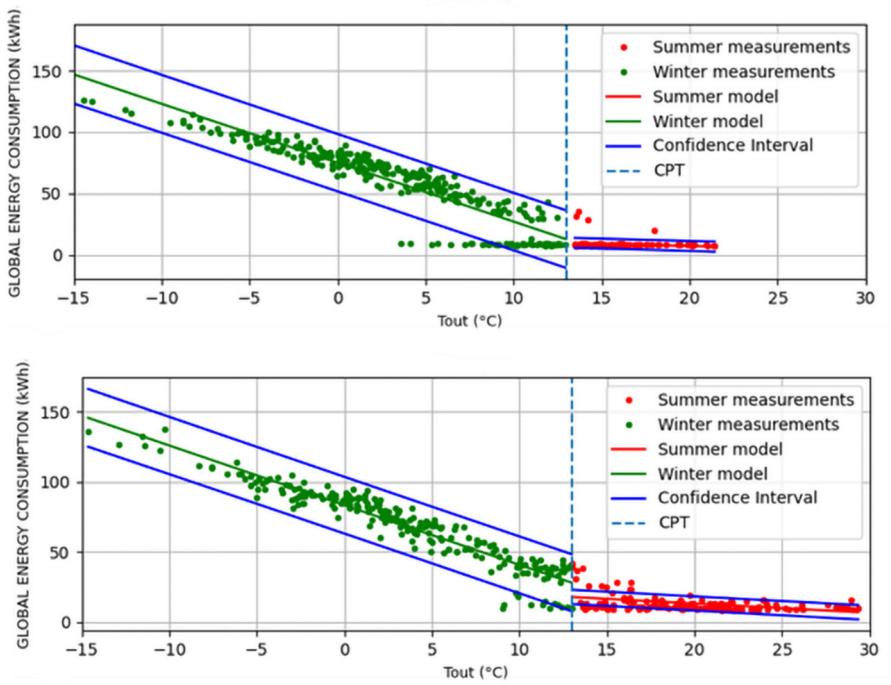

Fig. 2. ESC examples of diurnal (5.00 pm – 6.00 pm) and nocturnal (1.00 am – 2.00 am) cases.

CPT individuation operation was conducted for each hourly model, defining 24 different CPTs per building. These values are typically similar, proving that the duration of the heating season depends on building and users' behavior. Figure 3 shows the result of the mathematical procedure for determining CPT. The proposed example's CPT value is around 14°C, corresponding with the slope coefficient drop.

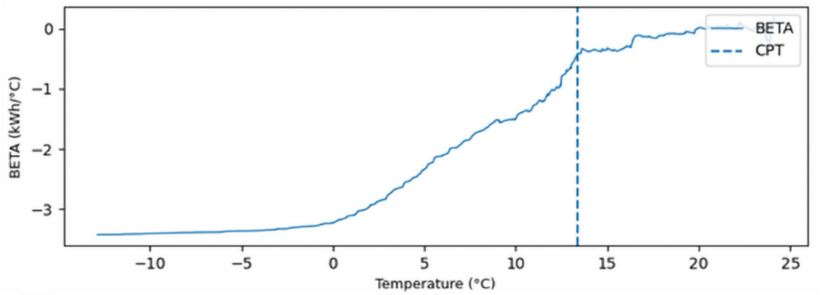

Fig. 3. Example of β variation by increasing temperature intervals.

Key performance indicators show the model's satisfactory performance. For the temperature-dependent part of the ESC, R^2 presents an average value of 0.80 among different day-hours and diverse buildings. R^2 reaches peaks of 0.90, which is equal to 0.50 in the worst cases. With the 24-hour modeling approach used in this analysis, performances are better than using a single model for the entire year. Indeed, the total yearly R^2 calculated using the combination of 24-hourly models shows an average increase of about 0.05 compared to the model applied to the complete year. This result proves that hourly modeling can provide more detailed information about the hourly heat load pattern and higher precision performances. Also, YEC shows satisfactory performance, approximating energy consumption with an error below 8%.

4.2 Total Heat Loss Coefficient (k_{ESC})

Some comparisons among different buildings in the dataset are provided, especially regarding the total heat loss coefficient (k_{ESC}) assessment. The slope of the SH-driven part (below CPT) of the ESC represents kESC [22]. In Figure 4, the variation of the kESC coefficient during different hours of the day is reported for some buildings in Tartu's dataset.

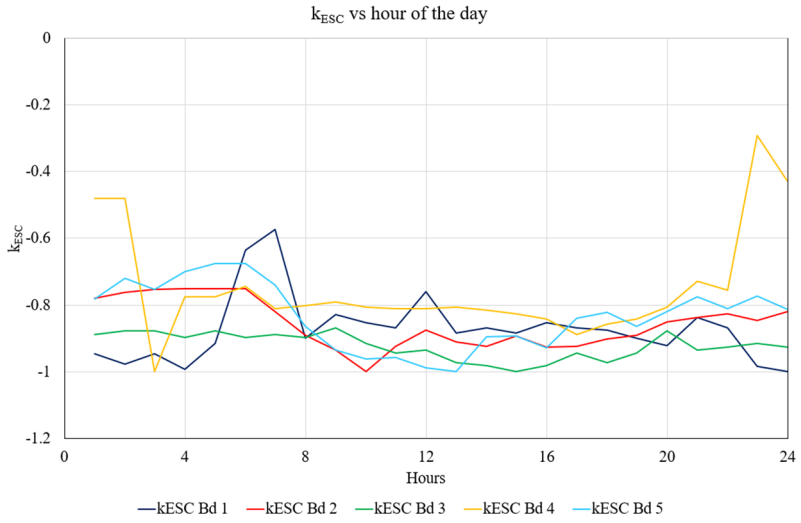

Fig. 4. k_{ESC} variation along hour of the day among different buildings.

The plot is normalized based on the maximum value for kESC of each building to highlight the variation over 24 hours. Hourly values of k_{ESC} were obtained through the hourly ESC model. Most considered buildings present a variation of up to 30% between the lower and the higher values. The increasing k_{ESC} happens for a few buildings at around 7:00 am. This variation is probably due to internal temperature setpoint variation, causing increased SH demand and a consequent modification of the SH plant operation.

4.3 Temperature Moving Average (TMA)

Figure 5 shows an example of the result obtained through the temperature moving average analysis. In this building, TMA analysis indicates a relatively lower thermal inertia when compared with the other buildings under examination. Figure 5 reveals a pronounced correlation between energy consumption and outdoor temperature at the 7-hour. In the case of implementing a model for the prediction of energy consumption, this information could

be useful to implement the correct dependence of the previous hour of the energy demand of the building. It must be said that this analysis only affects SH load, as DHW heat consumption is primarily dependent on sudden peaks in users' demand.

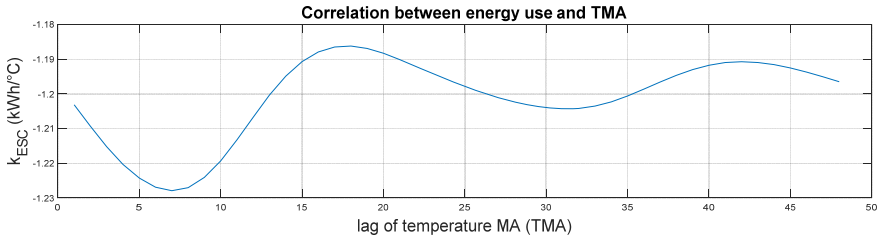

Fig. 5. Effect of temperature lag on energy use of a building.

4.4 SH and DHW splitting

Figure 6 and Figure 7 depict the results of the SH and DHW splitting operation. The energy demand is plotted versus outdoor air temperature (Figure 6) and the year's day (Figure 7). For these plots, the improvement given using standard profiles of DHW has not been implemented yet. A successful separation between SH and DHW was obtained. However, the DHW profile fluctuates significantly during the winter season compared to the summer season. These fluctuations are even larger in mid-season days, with low SH demand. In this period, almost the whole thermal load is attributed to DHW, creating peaks not representative of the real building's behavior.

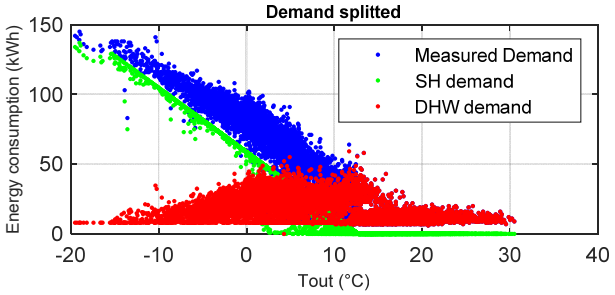

Fig. 6. Energy demand versus outdoor air temperature of a building in Tartu, Estonia.

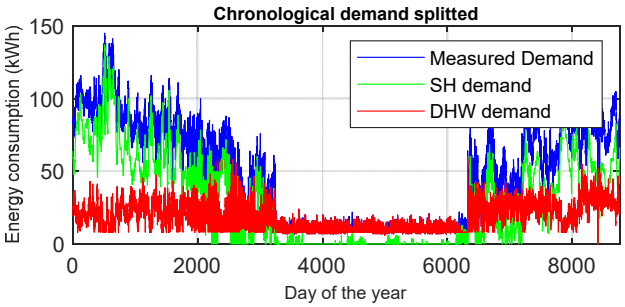

Fig. 7. Chronological energy demand of a building in Tartu, Estonia.

The standard profile correction explained in Section 2.5 is applied to improve the model in this perspective. Figure 8 reports an example of a winter standard profile representative of a group of similar days in terms of consumption within a specific building. For the analysis, two groups of similar days are created for every building. Hours of minimum, medium, and peak consumption can be identified. The clusterization process shows that, in most buildings, minimum consumption hours coincide with night-time, while peak consumption happens during lunchtime or evening. This behavior is typical of residential buildings. Some other buildings present high peaks in the morning, while minimum consumption hours are observed in the evening and at night. This peculiar pattern concedes with an extremely low energy demand during weekends, typical of commercial buildings, schools, and offices.

The comparison between the winter DHW demand obtained from the split operation and the winter standard profile reveals a mean difference typically ranging from 0% to 15%. However, if the difference between the value obtained from splitting and the corresponding one on the standard profile is too large (more than 30% of their average), they were adjusted by increasing or decreasing it by a certain percentage (it was used 50% as the adjustment factor). These percentages were decided through an analysis of many diverse buildings of the same dataset by considering a notable effect in terms of peaks flattening while maintaining the peculiar consumption pattern of the building. These percentages must be chosen according to the specificity of the dataset in the exam.

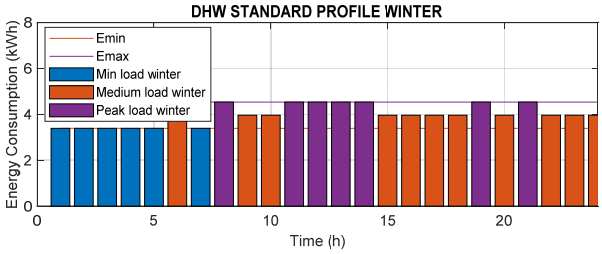

Fig. 8. Winter standard profile.

Thanks to the implementation of standard profiles, these peaks in DHW heat consumption have been significantly reduced. The subtracted heat load from the DHW profile is added to the SH one. Thus, SH appears even in moments with outdoor temperatures higher than CPT. This is a more accurate representation of what happens in real buildings. Indeed, the start and end of the heating season are usually defined by dates rather than a temperature value. The numerical determination of the performance of the standard profile approach is planned for further work. In addition, thanks to standard profile implementation, it is possible to solve the problem of summer days with SH assigned. It happens when low outdoor temperatures are registered during this season. With the standard profile method, the summer measured heat consumption is entirely attributed to DHW. An example of profiles after improvement with standard profiles is reported in Figures 9 and 10.

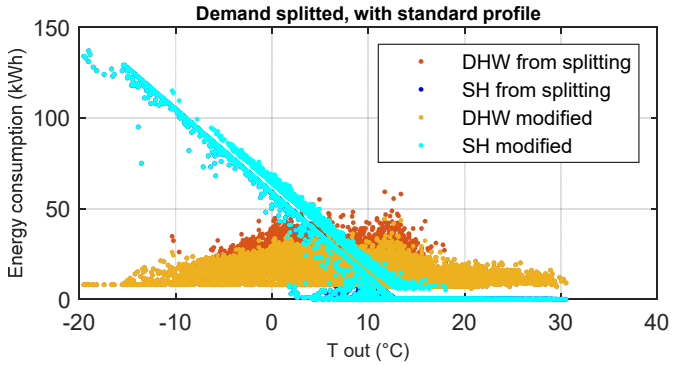

Fig. 9. Chronological energy demand of a building in Tartu, Estonia, after the standard profile correction implementation.

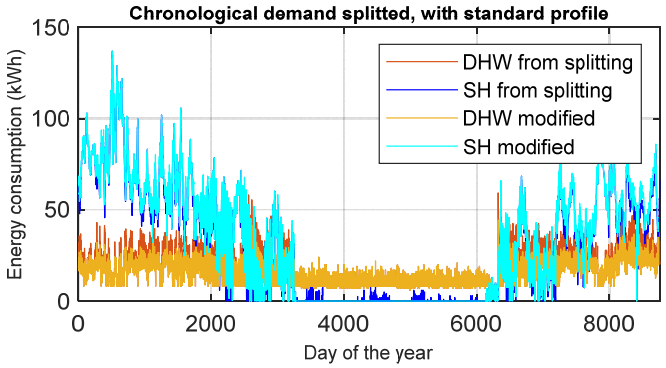

Fig. 10. Chronological energy demand of a building in Tartu, Estonia, after the standard profile correction implementation.

5 Conclusions

This paper presents a methodology using just global thermal energy consumption and outdoor temperature data as inputs, demonstrating its efficacy in analyzing building energy consumption. The procedure shows good results and is simple to implement for consumption analysis. In addition, the approach is suitable when data from building energy meters and weather stations are available, enabling wide applicability across various buildings. Outputs include detailed hourly consumption analysis, possibly individuating typical and extraordinary consumption patterns, and key building information like the Energy Signature Curve, $Qtot$, CPT, and DHW circuit thermal losses. Furthermore, global heat demand is split into SH and DHW, which offers insights into standard DHW summer profiles. This information is then used to create a dedicated model for DHW heat consumption to enhance the model's performance. The methodology can be applied to different buildings and is particularly suitable for the residential sector, generally characterized by regular heat load patterns. Accuracy in terms of R^2 of the linear regression model shows acceptable results, always above 0.70, exploiting separate models for each hour of the day. YEC, yearly energy consumption deviation, is also low, showing a good capability of the model in providing heating consumption results.

References

[1] EEA - European Environment Agency, "EEA greenhouse gases - data viewer," EEA greenhouse gases. Accessed: Dec. 22, 2023. [Online]. Available: https://www.eea.europa.eu/data-and-maps/data/data-viewers/greenhouse-gases-viewer

[2] A. Thonipara, P. Runst, C. Ochsner, and K. Bizer, "Energy efficiency of residential buildings in the European Union-An exploratory analysis of cross-country consumption patterns" (2019) doi: 10.1016/j.enpol.2019.03.003.

[3] V. S. K. V. Harish and A. Kumar, "A review on modeling and simulation of building energy systems," *Renewable and Sustainable Energy Reviews*, vol. **56**, pp. 1272–1292 (Apr. 2016) doi: 10.1016/J.RSER.2015.12.040.

[4] T. Ahmad, H. Chen, Y. Guo, and J. Wang, "A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review," *Energy Build*, vol. **165**, pp. 301–320 (Apr. 2018) doi: 10.1016/J.ENBUILD.2018.01.017.

[5] S. Schneider, P. Hollmuller, J. Chambers, and M. Patel, "A Heat Demand Load Curve Model of the Swiss National Territory," *IOP Conf Ser Earth Environ Sci*, vol. **290**, no. 1 (2019) doi: 10.1088/1755-1315/290/1/012107.

[6] A. Chmielewska, "Fluctuating temperature of the mains water throughout the year and its influence on the consumption of energy for the purposes of DHW preparation," *E3S Web of Conferences*, vol. **44** (2018) doi: 10.1051/e3sconf/20184400017.

[7] I. Meireles, V. Sousa, B. Bleys, and B. Poncelet, "Domestic hot water consumption pattern: Relation with total water consumption and air temperature," *Renewable and Sustainable Energy Reviews*, vol. **157**, p. 112035 (Apr. 2022) doi: 10.1016/J.RSER.2021.112035.

[8] D. Ivanko, Å. Lekang Sørensen, and N. Nord, "Splitting measurements of the total heat demand in a hotel into domestic hot water and space heating heat use," (2020) doi: 10.1016/j.energy.2020.119685.

[9] L. Pedersen, *Load Modelling of Buildings in Mixed Energy Distribution Systems*, no. February. 2007. [Online]. Available: http://ntnu.diva-portal.org/smash/record.jsf?pid=diva2:122458

[10] P. Bacher, P. Anton de Saint-Aubain, L. Engbo Christiansen, and H. Madsen, "Non-parametric method for separating domestic hot water heating spikes and space heating," *Energy Build*, vol. **130**, pp. 107–112 (2016) doi: 10.1016/j.enbuild.2016.08.037.

[11] A. Marszal-Pomianowska, C. Zhang, M. Pomianowski, K. Gram-Hanssen, and A. R. Hansen, "Simple methodology to estimate the mean hourly and the daily profiles of domestic hot water demand from hourly total heating readings," *Energy Build*, vol. **184**, pp. 53–64 (2019) doi: 10.1016/j.enbuild.2018.11.035.

[12] M. Aydinalp-Koksal and V. I. Ugursal, "Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector," (2007) doi: 10.1016/j.apenergy.2006.09.012.

[13] D. Fischer, T. Wolf, J. Scherer, and B. Wille-Haussmann, "A stochastic bottom-up model for space heating and domestic hot water load profiles for German households," *Energy Build*, vol. **124**, pp. 120–128 (2016) doi: 10.1016/j.enbuild.2016.04.069.

[14] J. Wiehagen and J. L. Sikora, "Domestic Hot Water System Modeling for the Design of Energy Efficient Systems," *National Renewable Energy Laboratory Report* (2002) [Online]. Available: www.nahbrc.org

[15] D. Ivanko, H. T. Walnum, and N. Nord, "Development and analysis of hourly DHW heat use profiles in nursing homes in Norway," *Energy Build*, vol. **222**, p. 110070 (2020) doi: 10.1016/j.enbuild.2020.110070.

[16] C. Deb and A. Schlueter, "Review of data-driven energy modelling techniques for building retrofit," *Renewable and Sustainable Energy Reviews*, vol. **144**, p. 110990 (Jul. 2021) doi: 10.1016/J.RSER.2021.110990.

[17] M. Lumbreras, G. Diarce, K. Martin, R. Garay-Martinez, and B. Arregi, "Unsupervised recognition and prediction of daily patterns in heating loads in buildings," *Journal of Building Engineering*, vol. **65**, p. 105732 (Apr. 2023) doi: 10.1016/J.JOBE.2022.105732.

[18] M. Lumbreras *et al.*, "Data driven model for heat load prediction in buildings connected to District Heating by using smart heat meters," *Energy*, vol. **239**, p. 122318 (Jan. 2022) doi: 10.1016/J.ENERGY.2021.122318.

[19] T. Tereshchenko, D. Ivanko, N. Nord, and I. Sartori, "Analysis of energy signatures and planning of heating and domestic hot water energy use in buildings in Norway," *E3S Web of Conferences*, vol. **111**, p. 06009 (Aug. 2019) doi: 10.1051/E3SCONF/201911106009.

[20] L. Pedersen and R. Ulseth, "Method for load modelling of heat and electricity demand Method for Load Modelling of Heat and Electricity Demand," *10th International Symposium on District Heating and Cooling*, vol. **Sektion 5 b**, no. Heat/cold generation, 2006.

[21] K. B. Lindberg and G. Doorman, "Hourly load modelling of non-residential building stock," *2013 IEEE Grenoble Conference PowerTech, POWERTECH 2013* (2013) doi: 10.1109/PTC.2013.6652495.

[22] M. Eriksson, J. Akander, and B. Moshfegh, "Development and validation of energy signature method – Case study on a multi-family building in Sweden before and after deep renovation," *Energy Build*, vol. **210**, p. 109756 (2020) doi: 10.1016/j.enbuild.2020.109756.