PREDICTION OF ELECTRICAL CONSUMPTION OF DOMESTIC HOT WATER BASED ON STATISTICAL MODEL

Joana GIRAUD-BIT. Tutor: Christian GHIAUS, Ibrahim ALI KACHALLA

INSA Lyon.

PIRD 22/01/2024

Abstract

The objective of the project is to determine the required tank capacity to store energy from solar panels. Solar panel energy is intermittent, and its availability varies. To store this energy, we plan to use it for water heating and store it in a tank. The analysis focuses on identifying consumption patterns in selected apartments based on a two-year data collection. The goal is to assess how apartments can collectively act as a unit and compare their patterns. This analysis aims to identify groups of apartments that exhibit similar behaviors. In the second part of our analysis, we seek to identify patterns within individual apartments to understand if they display self-correlation over time. Additionally, we explore the correlation of average consumption for the apartments with itself. It is suspected that finding correlations within a group of apartments using the average may be more achievable. These hypotheses will be explored in this paper. Ultimately, the objective is to uncover recurring patterns that can be used for predicting hot water consumption and subsequently determine the app tank size for storing the energy.

Keywords: Electrical building prediction; solar panel energy consumption; domestic hot water.

I- Introduction

In the European context, the building industry is responsible for approximately 40% of primary energy consumption. It is also a significant contributor to CO_2 emissions, around 30% according to ADEME. These emissions can be categorized as direct and indirect, with direct emissions encompassing aspects like furniture and their usage, and indirect emissions due to the production of electricity and heat required within a building. The main objective of this research is to mitigate CO_2 emissions and minimize the interaction with the grid, particularly those resulting from the consumption of fossil fuels which tends to release a higher volume of CO_2 into the atmosphere. To this end, there has been a massive shift towards Renewable Energy Sources over recent years. The decreasing availability of fossil fuels leaves no choice but to develop sustainable solutions, and reliance on Renewable Energy Sources serves as a viable response to this issue.

However, the utilization of renewable energy presents some challenges such as their intermittent availability and less efficient energy storage capabilities when compared to fossil fuel. For example, solar panels produce electricity exclusively when exposed to sunlight and are ineffective at night or under cloudy conditions. Consequently, solar panels find more application in regions with a lot of sunlight. Subsequently we need to gather the energy generated from the solar panel, store it, and make it accessible whenever wanted.

The main goal is to store renewable energy and manage renewable energy to precisely match the daily energy consumption of buildings. Nevertheless, it needs a comprehensive understanding of a building's energy consumption which can often be a difficult task. This highlights the importance of accurately predicting building energy consumption to align it with energy production. Correct predictions offer the opportunity to rely on renewable energy to produce the exact amount of energy required. It also enables to anticipate the way to produce electricity and to consume the energy locally available. It is then important to note that failing to utilize renewable energy when it is available can lead to missed opportunities for clean energy usage, showing the need to predict a framework.

In this research, the focus centers on predicting the electrical energy required for heating domestic hot water. According to the International Energy agency, heating domestic hot water consumes between 7.5 and 40% of energy. Reliable prediction in this domain provides the capacity to rely on renewable energy production and then, reduce the environmental impact. It is worth noting that the consumption of electrical energy for domestic hot water is directly correlated with the building's hot water demand. Consequently, the initial step involves several questions: How do we effectively store the energy generated by the solar panel, and what quantity should be stored to ensure access to hot water whenever required? For the purpose of storing energy for heating domestic hot water, a solution is to use a tank. This tank can receive the hot water produced by the solar panel, preserving it for use when necessary. However, a subsidiary question arises: How many tanks do we need? The consideration here is whether a single tank is sufficient for one apartment or if it should serve a group of apartments. In this paper, we will explore different approaches to address this question, recognizing that predicting the consumption of a group of apartments may be easier than for a single apartment.

Additionally, the question arises regarding the quantity of water that needs to be stored to ensure a consistent supply of hot water. To answer this question, we must work out the mean consumption of the apartments. This mean consumption serves as a reliable indicator of the accumulated amount required to obtain sufficient water when needed. Since the energy generated by solar panel is intermittent, only accessible when there is sunlight, it cannot provide energy on demand. Determining the required accumulation allows us to store enough water over a specific period when solar energy is available. The objective is to identify a timeframe when consumption patterns become consistent, enabling the sizing of tank associated with the required water volume.

Then, the next step involves predicting the domestic hot water consumption in a building, which then gives information about the required electrical energy production for water heating.

II- Literature

Ibrahim Ali Kachalla et al. [1] researching the prediction of domestic hot water (DHW) and its electrical consumption, highlights multiple factors influencing these consumptions. These factors encompass climate, environment, building characteristics, time of use, user behavior, technical factor and system design, control design and strategy, sources of data and measurement technique and intraday energy market.

Analyzing climate conditions and seasonal variations involves real-time and logistic projections, considering local meteorological conditions, seasonal fluctuations, habits, environmental concerns, working days or weekend and socio-economic level to predict consumption trends.

Illustrating the importance of these factors, the study conducted by Kaiser Ahmed et al. [2] focused on creating an hourly profile of individual DHW consumption, incorporating seasonal variations and occupant numbers. The study, based on 182 apartments with 379 inhabitants in Finland over a two-year period, revealed an annual average domestic water consumption of approximately 43L /day. Importantly, they observed maximum consumption in November and minimum in July, with variations of +15.3% in colder periods and -17.4% in warmer period. Furthermore, they described that the consumption was higher from November to February and lower from May to July. The study emphasized the importance of considering seasonality, prompting efforts to derive correction factors based on the month for an accurate consumption profile.

The main focus of Kaiser Ahmed's research was the creation of an hourly profile, a task that aligns with our project goals. If we identify correlations and establish profiles successfully, we can proceed to predict water consumption and subsequently, electrical consumption. However, it is crucial to identify correlations. The success of our hypotheses would enable us to target the prediction of Domestic Hot Water (DHW). Some researchers have already conducted studies on this topic.

For example, the research conducted by Lukas G. Swan et al. in 2011 [3] tried to predict the DHW consumption. Their analysis centered on a bottom-up approach for annual predictions in Canada. Although this differs from our current research, their findings are noteworthy. They employed an artificial neural network as a statistical method to estimate DHW consumption, suggesting that consumption is influenced by household demographics including dwelling type, the population density, system energy factor, soil temperature, heating and cooling degree day, storage tank, dishwasher use, clothes washer use… Notably, they identified a correlation between DHW consumption and average soil temperature, which tends to be greater in areas with colder soil temperatures.

Ibrahim Ali Kachalla et al. [1] also accentuated the significance of time to use in influencing consumption. Peak-hour needs are impacted by user activities, and understanding patterns can be highly complex. In their research Kaiser Ahmed et al. also attempted to explore these aspects. They showed the distinctions between weekdays and weekends, noting that the morning peak consumption was generally lower than the late evening peak as mentioned earlier.

Moreover, understanding people's behavior can provide valuable insights for predicting DHW consumption. The study conducted by Cao et al. [4] focused on the influence of occupant behavior on consumption, considering the shower habits. Utilizing Support Vector Machine (SVM), they aimed to predict occupants' shower profiles, achieving results with an average accuracy and recall of 0.3 and 0.7 each. After predicting the shower profiles, they estimated the future hot water demand by summing individual predictions for each occupant. Impressively, the study demonstrated that predicting the shower habits of inhabitants could lead to accurate predictions. To prove it they worded out the root mean square error which was equal to 77.83% and the hot water supply assurance which was equal to 99.01%. This study declared that occupant's behavior can have a huge impact on the consumption.

Furthermore, Ibhrahim Ali Kachalla et al. [1] also discussed the impact of control design and strategy on energy consumption, highlighting examples such as rescheduling control based on occupant consumption patterns to reduce wastage. Drawing on the findings of Kaiser Ahmed et al.'s [2] study which aimed to make consumption profile, it is obvious that such profile analysis can contribute to controlling consumption over time.

Additionally, he compiled a list of methods for predicting consumption. Primarily, there are time series forecasting models designated to anticipate future water consumption by analyzing past trends and historical data. These models aim to identify patterns caused by daily routines or specific events. Among the most utilized models are Auto-Regressive Integrated Moving Average (ARIMA), Seasonal Auto-Regressive Integrated Moving Average (SARIMA), Exponential Smoothing Model, and machine learning methods such as Recurrent Neural Networks (RNN) and Long-Short Time Memory (LSTM) networks.

Mohammad Ebrahim Banihabib et al., [5] conducted research in Tehran, located in northern Iran, focusing on the forecasting methods of ARIMA and Non-Linear Auto-Regressive Exogenous (NARX). Their objective was to assess the accuracy of both methods with and without additional forecaster factors such as climatic variables. Their aim was to evaluate the impact of each factor on determining the most effective model. The evaluation of their methods involved criterions like Root Mean Square Error (RMSE), Mean Absolute Relative Error (MARE), Maximal Root Error (MAXRE), Mean Bias Error (MBE) and coefficient of determination (R^2). Their findings indicated that the ARIMA method yielded an improved forecast for DHW consumption, particularly when considering sunny hours. In the case of the NARX method, incorporating such factors such as sunny hours and population growth enhanced the prediction model for DHW consumption. The relevance is attributed to the notion that increased sunshine hours can lead to more evaporation, while population growth directly impacts water consumption.

Another conventional method of time series research is the grey Markov model, as employed in the study conducted by Wang Zhaocai et al. [6] for the annual prediction of DHW. The researchers observed that this model provides a better fit of the original data series due to its lack of after-effects. The accuracy was further improved through Markov correction applied to the initial grey model. It's noteworthy that this method is not well-suited for long-term domestic water consumption prediction; it aligns more closely with our study's focus on short-term predictions.

For consumption prediction, data-driven, machine learning methods and Artificial Neural Networks can also be employed, considering data features such as outdoor weather, indoor environment conditions, time occupancy type, occupant energy usage behavior or historical energy consumption and degree of cloudiness. Lazzari Florencia et al., [7] employed this approach to predict hourly consumption one day-ahead of electrical consumption using smart meters and information on the local weather and the influence of the stochastic part of the user's behavior. The study was conducted in the southeastern region of Spain and fiscal smart meters were used to make the research applicable to real scenarios where personal data gathering is not feasible. To predict the overall electricity consumption, the researchers employed Gaussian mixture clustering to identify behavior clusters. Then they utilized the Extreme Gradient Boosting classification model, a tree method, to predict the day-ahead behavior pattern. This prediction was integrated into an Artificial Neural Network to enhance the analysis of user behavior and predict the electrical hourly consumption for the following day. Finally, the classification model was evaluated using the Euclidean Distance-based Accuracy.

Ibrahim Ali Kachalla et al. [1] also highlighted the Support Vector Machines (SVM) methods that are less prone to overfitting and can yield robust models. This methodology was applied in a study conducted by Jae Yong Lee et al., [8], where they aimed to forecast the demand for DHW using SVM. SVM, as a structural analysis method, considers the relationship between water consumption and various influencing factors. The study focused on 918 households over two months in both summer and winter in the Republic of Korea. Like Ahmed et al., they underlined the importance of seasonal behavior, noting that winter flow demand was 4.7 times higher than in summer, with distinct morning and evening peak consumptions during weekdays from 7 to 8 in the morning and 18 to 19 in the late afternoon. The researchers also explored the demand predicting based on the preceding outdoor temperatures. They discovered that the consumption might depend on the outdoor temperatures several days before the prediction day. To forecast DHW demand, they employed SVM, utilizing outdoor temperature and DHW energy demand. The study included k-fold cross-validation and the Nash Sutcliffe model efficiency coefficient to identify the most effective model. Their results indicated that the predictions were most accurate in winter when considering a four-day ahead outdoor temperature, while a five-day ahead temperature was optimal for summer.

Finally, we can introduce another approach that depends on the specifics of the study's country. For instance, Wojciech Rzeznik et al., [9] conducted a study comparing actual and forecasted domestic consumption and heat power demand in buildings in Poland. Their research included a comparison of various methods to determine the most effective one. Data for this research were gotten from the annual monitoring of DHW and heat power demand in eight buildings situated in Central Poland during 2021. The compared prediction methods included Sander's, Recknagel's, the standard method and the method according to Polish regulation 2018 and 2015. Their evaluation focused on the relative forecasting error. It is noteworthy that the study revealed that the optimal method for predicting DHW consumption differed from that for forecasting heat power energy for DHW. Specifically, for DHW consumption, the best method was identified as the standard regulation of 2015 in Poland, whereas for heat power, the standard method proved to be the most accurate. The 2015 regulation method is grounded in factors such as apartment area, average heat power, average efficiency of the DHW system, usable energy demand for DHW and annual water consumption. On the other hand, the standard method relies on the average and maximum hourly DHW demand, utilizing the hourly coefficient of irregularity.

III- Methodology

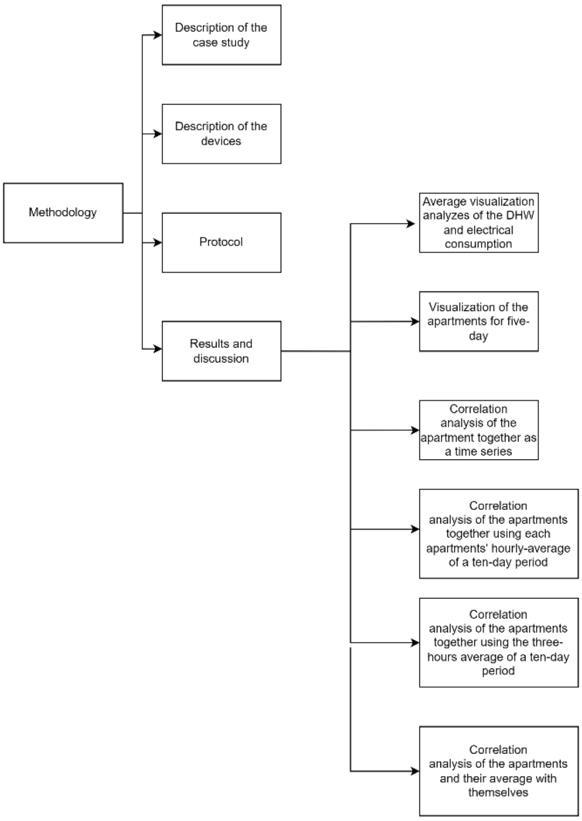

Figure 1: Framework

a. Description of the case study

The case study involves a two-years data collection period from 2020 to 2021 in the City of Villefontaine, situated near Lyon in France. The measurements were conducted in a building, which is a R+5 building. The apartments of the building are respectively identified as 33, 34, 35, 36, 63, 64, 65, 67. It is noteworthy that apartment 33 remained unoccupied during the monitoring period. Therefore, its data was excluded from the analysis.

The case study also involves the city of Rive de Gier in the Loir department in central France. One of the buildings we are studying is a R+8 building, and the other is a R+6 building. The measurements were made on three apartments of the R+8 building and two apartments of the R+6 buildings, they are identified as 25, 26 and 27 in the R+8 building and as 8 and 9 in the R+6 building. However, we won't take into consideration the apartment 8 because we don't have enough measurements about the electricity consumed by the DHW.

b. Description of the devices

To collect all the necessary data, the main monitoring device utilized was the Eco-Touch device, supplied by the company OGGA, and served as an energy supplier working with a 230V power source upstream of a circuit breaker. This instrument was able to measure electricity consumption directly and could facilitate other types of measurements such as indirect ones.

Direct measurements are those acquired by the Eco-Touch device itself, which was connected to the main electrical board of the house. It had four toroids, each one was monitoring a specific electrical consumption parameter. One toroid was dedicated to capturing the overall electrical meter reading, another focused on the circuit breaker for outlets, a third was dedicated to the additional electrical DHW circuit, and the fourth was designated for the electrical radiator circuit and, when present, the air conditioning circuit.

Indirect measurements, on the other hand, were accomplished through a set of sensors and transmitters. These included an indoor temperature and humidity sensor that could be placed in various indoor spaces where monitoring was required. Temperature readings were taken every 100 seconds and sent to the Eco-Touch whenever a change of 0.5°C or a change of 2% of humidity was occurring.

Additionally, an outdoor temperature and humidity sensor was used, which sent data when there was a variation of 0.6°C in temperature and 2% in humidity. If the values remained stables, periodic status updates were sent every 100 to 3000 seconds.

There was also an illuminance sensor, provided by the company Eltako which was equipped with a small solar cell and a battery. This sensor communicated with the Eco-Touch when the illuminance level was above 300 Lux and whenever there was a notable change of brightness, typically above 500 lux within about 100 seconds.

Furthermore, the monitoring system included a window opening detector and a volumetric meter for DHW consumption. The measurement of hot water was accomplished using three instruments working together: the volumetric meter Aquadis + mark ITRON, the Cyble Impluse Sensor mark ITRON and the OGGA Impulse Counter mark OGGA.

Overall, the case study has a huge monitoring approach, covering aspects such as window openings and closures, both internal and external temperature and humidity, external illuminance, volume of domestic hot water consumed, average electricity consumption for various systems including heating, DHW, household appliances, underground heating and the total average electricity usage.

On the figures below, there is the map of the sensor in the different buildings (

Figure 2, Figure 3, Figure 4 and Figure 5).

Figure 2: Plan of the second floor of the building.

Figure 3: Plan of the building.

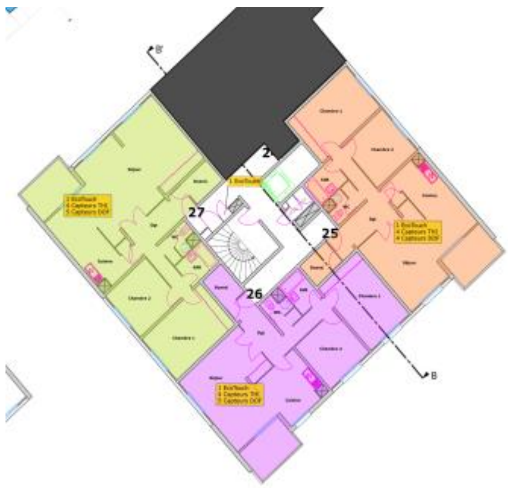

Figure 4: Plan of the R+8 of Rive de Gier

Figure 5: Plan of the R+6 building of Rive de Gier

c. Protocol

The initial phase of our study involves visualizing the data. The first analysis focuses on average values over the two-year period with hourly sampling to try to group our apartment together. Our second phase is to examine the data-patterns over a five-day period using hourly sampling to identify common trends of the apartment with themselves and together. This analyze will serve both of our hypothesis. Subsequently, these visualizations will serve as the basis for our future correlation analysis.

Our initial hypothesis involves grouping apartments together with similar patterns before predicting consumption. Predicting consumption for a group of apartments seems more manageable than for individual ones. Therefore, we start our study by searching a method to group our apartments together. Our first assumption is that they show a time-series connection, meaning some apartments are correlated over the days, with a correlation coefficient above 0.5. To test this hypothesis, we initially observe a five-day week time series from 07/12/2020 to 12/12/2020 for all apartments. Subsequently, we work out correlations among all apartments to identify potential groups. However, the patterns may change if the period is not from 07/12/2020 to 12/12/2020 so we extend the analysis by adding one more day and examining if the correlation coefficients remain constant or change. As a comprehensive approach, we also examine another five-day week from 14/12/2020 to 19/12/2020 to ascertain whether the identified groups, if any, persist or undergo modifications. In the case of changes, we look at the coefficients to understand their variations.

If our first hypothesis to group the apartments together using a time-series is unsuccessful, our second hypothesis to group the apartment together aims to identify correlation between apartments and patterns by smoothing the variations among apartments using average values over a ten-day period, creating a more accurate and refined feature than a single day's observation. This involves calculating the hourly average values, and then the three-hour hourly average over ten days for each apartment, along with observing the corresponding standard deviation. Then, we will analyze the correlation between the average values of each apartment to identify potential correlations among them. To validate the consistency of these correlations, we will replicate the study using another set of ten days, determining whether the observed correlations remain unchanged or undergo modifications.

If unsuccessful, an alternative approach involves investigating whether an apartment exhibits temporal self-correlation and determining the appropriate timestep for such correlation. This study begins with ten-minutes samples and progressively increases the timestep, moving to one-hour samples and further resampling for some days periods. The underlying hypothesis suggests that a sole apartment may have a longer timestep for self-correlation. If this proves true, the next step is to examine the average consumption of the apartments, hypothesizing that it might reduce the timestep when combining multiple apartments. The rationale is that predicting consumption for a group of apartments may be more manageable than for an individual one. The objective is to minimize the timestep based on our findings, as this will guide the determination of the tank size needed to accommodate the consistent consumption. If the required storage timeframe is too extensive, it could lead to an impractically large tank.

If successful in establishing distinct groups, our further goal is to predict the domestic hot water consumption.

d. Results and discussion

i. Average visualization analyzes of the DHW and electrical consumption

The preliminary analysis is divided into two main sections. The first section examines the daily average consumption for domestic hot water and electrical usage of domestic hot water. While the second section explores the average consumption of domestic hot water and its electrical usage over several days. Both analysis uses the codes called 'Average Domestic Hot Water (DHW) and electrical analysis per day per Rive de Gier' and 'Average Domestic Hot Water (DHW) and electrical analysis per day per Villefontaine'. Each code contains a Jupyter Notebook with the same name that explains every step of the code. This analysis is connected to the hypothesis that we can categorize or cluster our apartments together.

Our initial step includes looking to the time intervals associated with each apartment's data since the data files have varying time step. The average time interval for electrical consumption data falls in the range of 2-3 minutes, while for domestic hot water consumption, it falls between 5-10 minutes. To standardize our analysis, we perform a resampling of all the data with an interval of 10 minutes.

We initiate our study by focusing on the average domestic hot water consumption. We obtain this data from the devices described earlier. We are first resampling our data to get the hourly average values. Then we have the Figure 6 and Figure 7 that illustrate the hourly average consumption of hot water for one day.

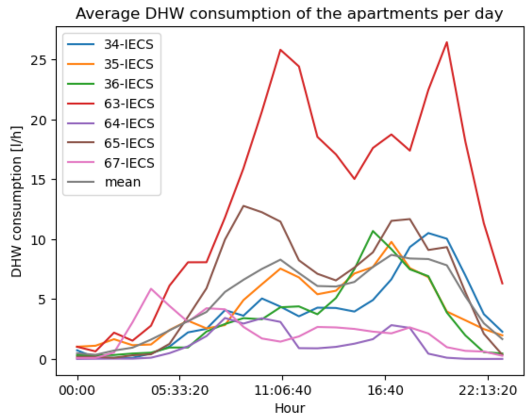

Figure 6: Graph presenting the average DHW consumption of each apartment per day of the city of Villefontaine

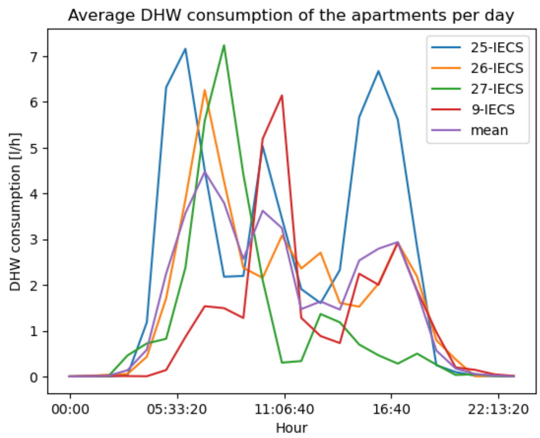

Figure 7: Graph presenting the average DHW consumption of each apartment per day of the city of Rive de Gier

To get these values, we worked out the average across the two-year dataset at our disposal. A pattern emerges from the data, with a peak of hot water consumption occurring between 5 and 11 AM and again in the late afternoon, present until 10 PM. It is noteworthy that most apartments apartments have higher consumption levels in the late afternoon in the city of Villefontaine while the peak is in the morning in the district of 'Maréchal Juin'. This suggests a recurring pattern where inhabitants tend to use domestic hot water primarily in the morning or late afternoon.

Further examination reveals that four apartments of the city of Villefontaine seem to share a similar consumption pattern. Their major consumption preak is in the late evening, and they have lower consumption levels in the morning. Nevertheless, it’s important to note that apartment 63 stands out with an extremely different consumption pattern. It has the highest daily consumption of domestic hot water over the day.

The examination of the apartment of the city Rive de Gier reveals that only one apartment shows a consumption pattern like the ones observed in the Villefontaine apartments. Specifically, this apartment demonstrates a peak consumption in both the morning and late afternoon. In contrast, the remaining apartments display different patterns. Notably, a common trend among these apartments is the occurrence of higher consumption peaks in the morning, although the timing may vary. The apartments 26 and 27 appear to share a similar consumption pattern throughout the day, while the apartment 9 also shows a comparable pattern, even though with a peak in late morning as opposed to the early morning peaks observed in the Villefontaine apartments.

For further information, we can also examine the histogram presented in Figure 8 and Figure 9.

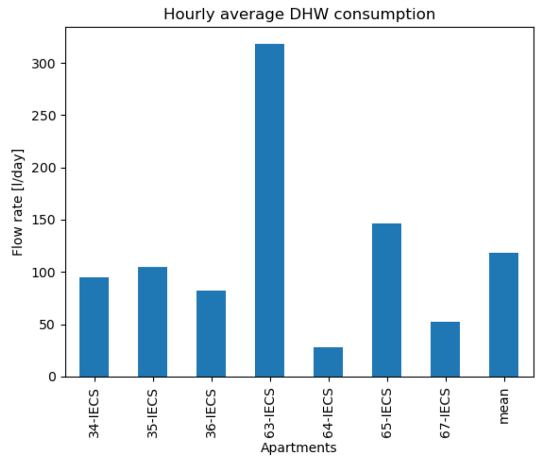

Figure 8: Histogram of the average DHW consumption per day of the city of Villefontaine

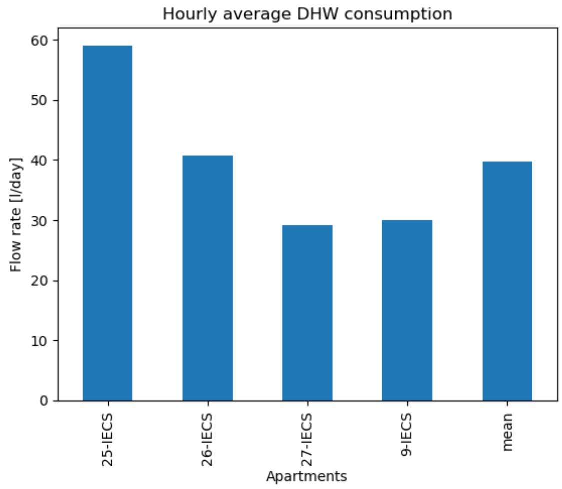

Figure 9: Histogram of the average DHW consumption per day of the city of Rive de Gier

As anticipated, the histogram of Villefontaine shows that the consumption patterns of apartments 34, 35, 36 and 65 are nearly identical. However, it is clear that the apartment 63 stands out with a significant higher consumption rate, almost three times that of the other apartments. This can be attributed to the fact that it accommodates more residents. On the other hand, the apartment 64 has notably lower consumption levels.

In contrast to the values observed in Villefontaine, it is evident that the apartment 25 shows a consumption pattern nearly identical to those observed in the group encompassing the apartment 34, 35, 36 and 65.

On the other hand, the apartments 9, 26 and 27 demonstrate lower consumption levels like the apartment 64.

Then, we will examine the electrical consumption associated with the demand of domestic hot water over one single day. In the Figure 10 and Figure 11, we observe a peak in electrical consumption for seven apartments at 10 PM. This is due to the boiler's operation during off-peak hours when there is minimal consumption. It presents a challenge when attempting to predict the electrical demand for hot water throughout the day, as it is based on the overall demand for a day rather than real-time demand. This is one of our motivations behind our goal to predict domestic hot water consumption as a precursor to accessing electrical consumption. The electrical consumption in the four other apartments is distributed throughout the day, attributed to the fact that the system remains active all day.

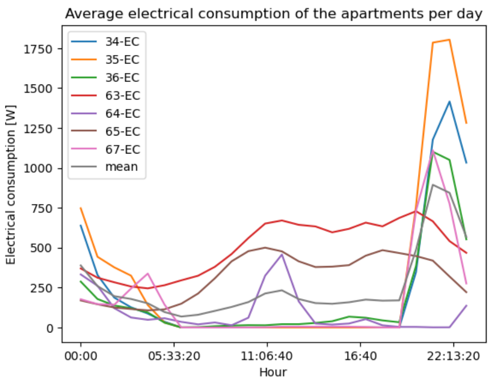

Figure 10 : Average electrical consumption of each apartment of the city of Villefontaine per day

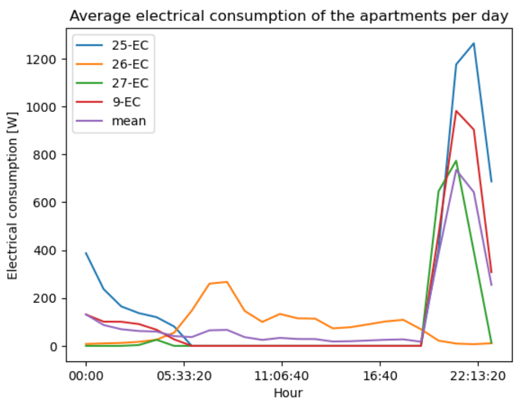

Figure 11 : Average electrical consumption of each apartment of Rive de Gier per day

Furthermore, examining the histogram in Figure 12 and Figure 13, we note that the apartment 63 still has higher consumption compared to the others, but the clear pattern observed earlier is not as pronounced. In this context, we can group the apartments 34, 35 and 65 together, while the consumption pattern of the apartment 36 is not as clear as previously.

In the case of Rive de Gier, it is notable that the apartment 25 shows a tendency to share similar values with the apartment 36. Similarly, the apartments 9, 26 and 27 appear to show values equal to those of the apartment 64, as observed previously for water consumption.

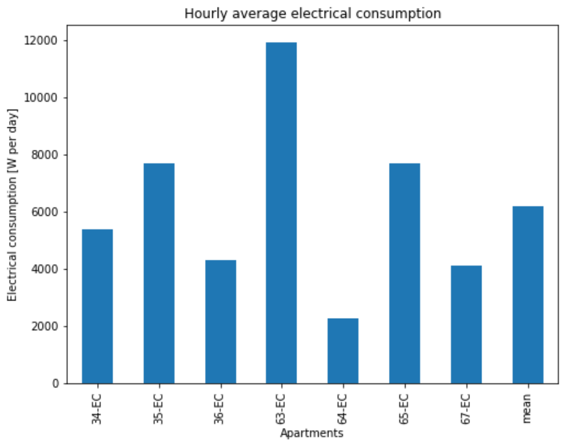

Figure 12 : Histogram of the electrical consumption per day of Villefontaine

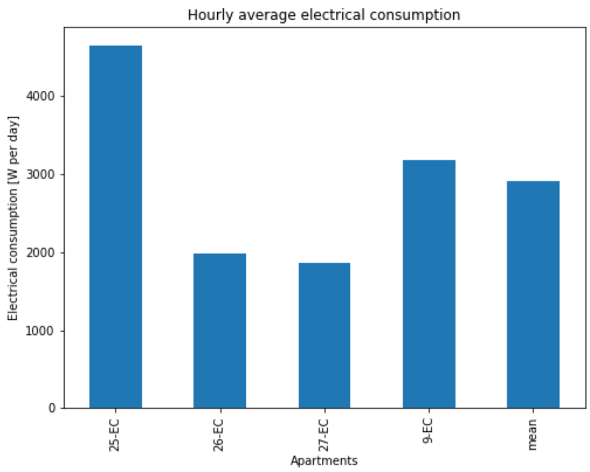

Figure 13 : Histogram of the average electrical consumption of the apartments per day of Rive de Gier

ii. Visualization of the apartments for five-day

Then, we follow the protocol's guidelines, which prescribe a data collection period of five weekdays, starting from December 7 to December 11, 2020, to visualize it using the code named 'Data Visualization for One Apartment'.

As previously mentioned, this section concerns both of our hypothesis. First of all, the visualization over a five-day period can help us to form group of apartments that share similar pattern. Secondly, following our second approach, it can help us into finding a self-correlation of the apartment. Before working out correlation, we want to visualize the data. There our analysis concerns a five-day consumption pattern.

Our study is based on hourly predictions, and for this phase of our study, we are evaluating the hourly average for each apartment over the course of the five days selected.

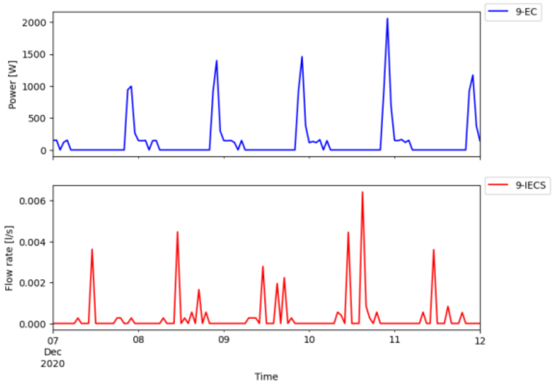

Figure 14 : DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 9

Figure 14 provides information regarding the consumption of DHW and the associated electricity usage for heating water in the apartment 9. The electrical analysis exhibits a gap during the day, with a peak occurring towards the end of the day. This pattern is due to the system operating during off-hours. Notably, there is a peak flow rate in the middle of the 7th day, while from the 8th to the 11th, the consumption appears to be distributed more evenly throughout the day. On December 10th, both DHW and electrical consumption reach their highest points, with this day showing the maximum electricity usage over the five-day period. Furthermore, the lowest electrical peak corresponds to the lowest consumption observed on December 7th. Interestingly, on December 9th, the DHW consumption is spread across the day, whereas on the 8th, there is a peak with some lower consumption. Despite these differences both days result in an equal electrical peak at the day's end. In general, it is evident that the peaks in domestic hot water consumption typically occur at the same time of day and exhibit a similar visual pattern.

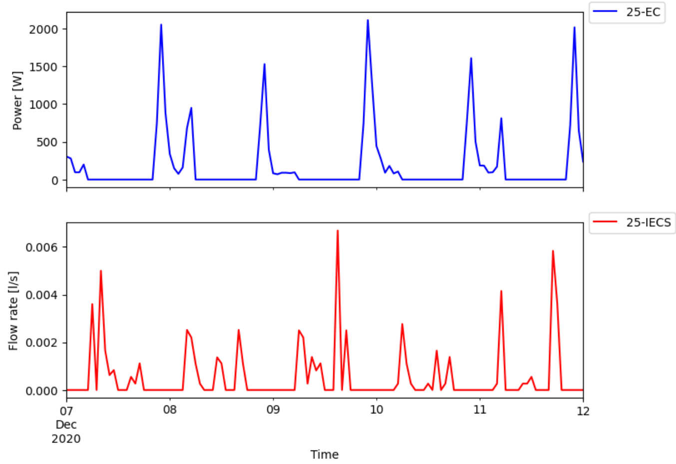

Figure 15 : DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 25

Figure 15 provides insights into the DHW and electrical consumption of the apartment 25. The electrical analysis indicates a lack of usage during the day, with a peak occurring in the late evening and early morning, even though the early morning peak is lower than the nighttime peak. When comparing this apartment to the apartment 9, it is observed that the consumption is higher on December 9th instead of December 10th. However, the electrical peak on that day is not the singular highest one. Interestingly, there are three peaks in electrical consumption, both around 2000 W, occurring on December 7th, 9th, and 11th.

Further examination reveals that the consumption pattern on December 7th appears to resemble that of December 9th but is more dispersed, with a peak in the morning instead of the afternoon. On the contrary, the consumption on December 11th differs significantly, featuring two peaks, one in the morning and another in the afternoon. Notably, the consumption on December 8th and 10th follows a similar pattern with spread consumption and an electrical peak around 1500 W.

In general, it is evident the peaks in domestic hot water consumption occur at the same time of day, but their values are very different.

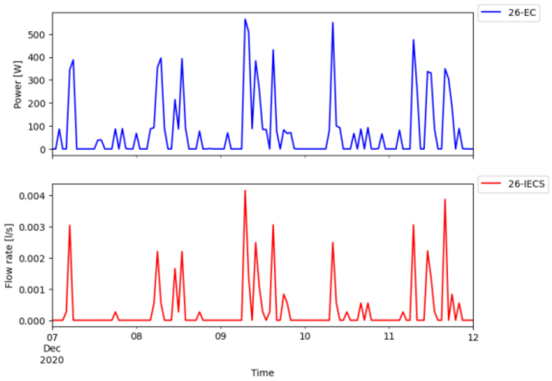

Figure 16 : DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 26

Figure 16 illustrates the DHW and electrical consumption patterns in the apartment 26. The electrical analysis reveals a consistent consumption throughout the day, meaning that the system operates without off-hours. Notably, the electrical consumption peaks never exceed 500 W. Interestingly, the peaks in electrical consumption align with the DHW peaks observed throughout the day.

For example, the single peak on December 7th corresponds to an electrical peak of approximately 400 W. Similarly, both DHW peaks on December 8th are repeated in the electrical consumption, manifesting as two distinct peaks. Furthermore, December 9th is the day with the highest electrical, featuring a morning peak and additional dispersed peaks throughout the day. This high consumption corresponds with high demand for DHW. Interestingly, the minor peak on December 10th results in a notable morning peak in electrical consumption.

In general, it is evident that the peaks in domestic hot water consumption occur at the same time of day. However, they do not share similar pattern as the peaks vary for each day.

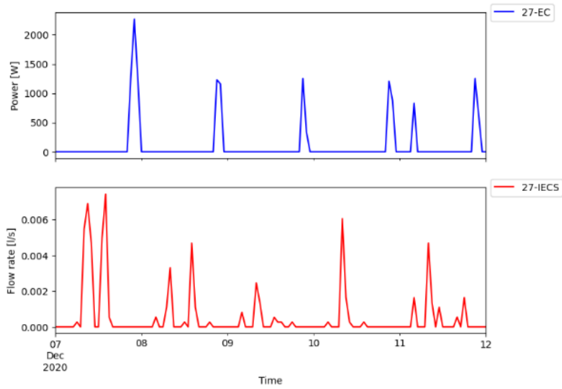

Figure 17: DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 27

Figure 17 represents the consumption patterns of the apartment 27. The electrical analysis reveals that the system operates during off-hours, exhibiting peaks in the late evening. The highest peak in electrical consumption aligns with the DHW peak consumption on December 7th, characterized by two morning peaks. Interestingly, the four subsequent peaks in electrical consumption remain consistently the same, even though the associated DHW consumptions do not exhibit identical patterns. For instance, on December 8th, two mornings DHW peaks are observed, while on both December 9th and 11th, the consumption is lower and dispersed throughout the day. Notably, December 10th stands out with a sole DHW peak in the morning, contrasting with the preceding days, despite having the same level of electricity consumption.

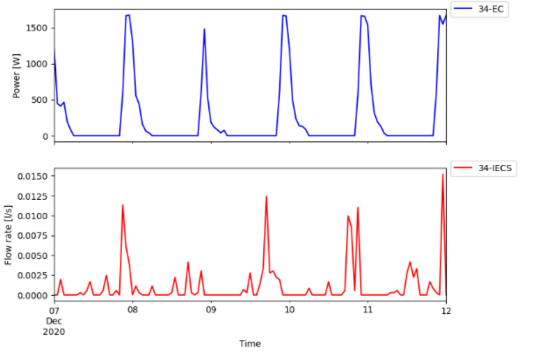

Figure 18: DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 34

Figure 18 displays the DWH and electrical consumption trends in the apartment 34. The electrical system operates during off-hours, with peaks occurring in the late evening. It is noteworthy that these peaks are consistently around 1500 W. However, it's worth noting that the DHW consumption do not really shared patterns across all days. Indeed, peak consumptions occur towards the end of the day but with different amplitude, except for December 9th, when the consumption is spread throughout the entire day.

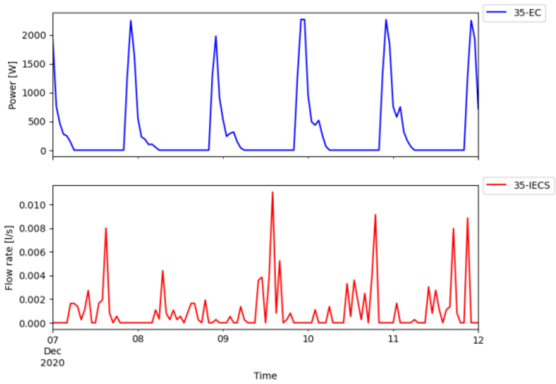

Figure 19: DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 35

Figure 19 represents the DHW and electrical consumption patterns of the apartment 35. The electrical consumption patterns reveals that the system is operating during off-hours. Like the apartment 34, the apartment 35 also displays electrical peaks consistently around 2000 W even though variations in DHW peak consumption. For instance, December 9th, 10th and 11th appear to exhibit the highest DHW consumption peak, whereas on December 8th, the consumption is dispersed throughout the day. December 7th represents a mix of patterns observed in other days, featuring both spread consumption and a peak in the late afternoon.

In general, the DHW consumption presents large variations throughout the day which does not share a similar pattern.

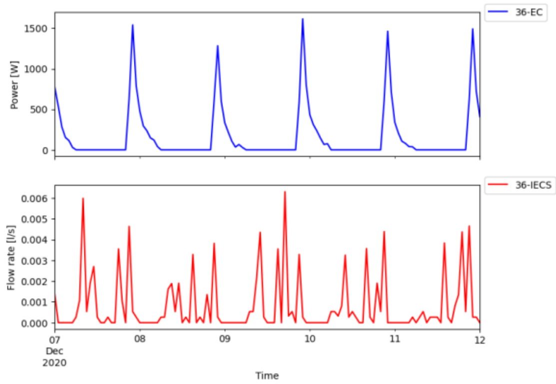

Figure 20: DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 36

Figure 20 illustrates the DHW and electrical consumption of the apartment 36. The electrical consumption pattern exhibits some peaks at the day's end meaning that the system is operating during off-hours. The consumption pattern exhibits a consistent use of DHW throughout the day. Peaks are spread across various times, with the highest peaks not consistently occurring simultaneously and varying from day to day. Furthermore, the days with the highest DHW consumption, such as December 7th and 9th, corresponds to the peak in electrical consumption. Nevertheless, it is noteworthy that even on days with lower DHW consumption, the electrical usage remains above 1000 W.

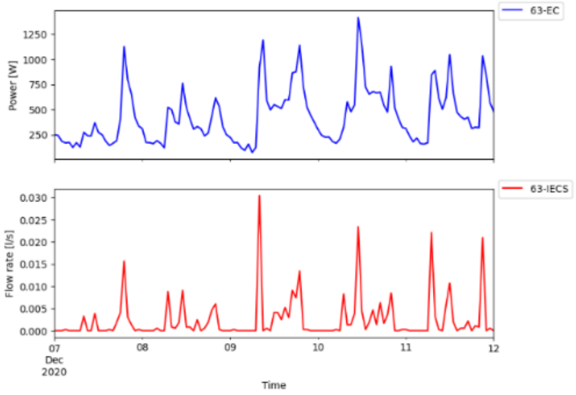

Figure 21: DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 63

Figure 21 displays the DHW and electrical consumption of the apartment 63. The electrical consumption is consistent throughout the day, we assume that the system is operating without off-hours. The peak in electrical consumption during the day aligns with the highest demand for DHW. Interestingly, even when the DHW demand is low, electrical consumption maintains a level around 250W throughout the day. On the contrary, for systems operating during off-hours, the consumption drops to 0 during the day. Furthermore, the DHW consumption appear not to follow a consistent pattern across the days.

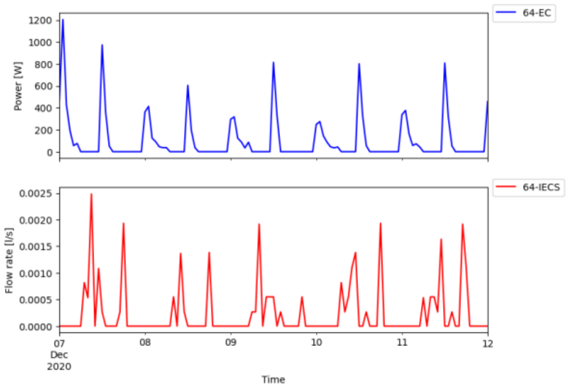

Figure 22: DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 64

Figure 22 provides insights into the DHW and electrical consumption of the apartment 64. The electrical analysis reveals two main peaks in the day: one in the midday and the other in the late afternoon. This suggests a potential dual operating of the system during off-hours. Comparing the DHW and electrical pattern, it can be assumed that the morning electrical peak corresponds to the DHW demand in the morning. For example, on December 7th, there is a higher peak at 12 in the morning, aligning with the high demand for morning DHW, and a lower peak in the afternoon, corresponding to the sole DHW peak in the afternoon. Conversely, the lowest electrical peak is linked to the lowest DHW consumption peak in the morning, as seen on December 8th. Notably, the electrical pattern remains consistent on December 9th, 10th and 11th, while the water consumption pattern differs. This suggests that the water consumption for these days is relatively similar in terms of electrical usage.

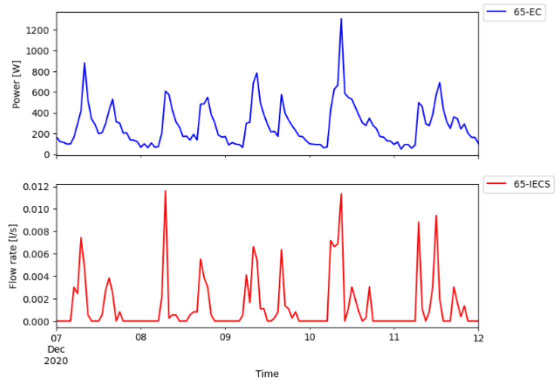

Figure 23: DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 65

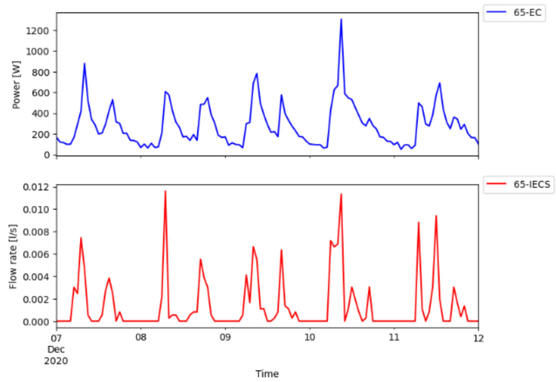

Figure 23 represents the DHW and electrical consumption of the apartment 65. The consistent electrical consumption pattern is a clue to deduce that the system is operating without off-hours. Once again, the electrical peak aligns with the demand for DHW. Interestingly, the highest peak on December 9th does not coincide with one of the highest electrical peaks. In fact, it corresponds to one of the lowest electricity consumptions over the five-days period. On December 12th, two main peaks are observed, yet the associated electrical consumption is around the weekly mean of approximately 600 W. On the contrary, the peak on December 10th results in the highest peak in the electrical analysis, highlighting it as the sole day with a discernible difference in electrical consumption compared to the other days. Furthermore, the DHW consumption appear not to follow a consistent pattern across the days.

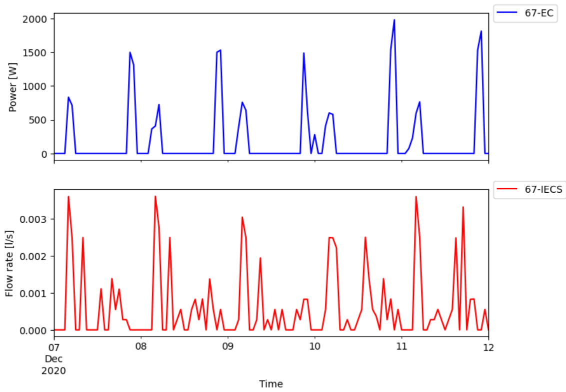

Figure 24: DHW (red) and electrical (blue) consumption between December 7th to December 12th, 2020 of the apartment 67

Figure 24 dwells the DHW and electrical consumption of the apartment 67. The electrical consumption, composed of peaks in the late afternoon, provides insights about the system used to get electricity, the system is operating during off-hours. Examining the DHW consumption patterns, there is a consistent spread of hot water usage throughout the day without recognizable pattern. Typically, one or two peaks are observed, followed by dispersed consumption through the day. Noting that the highest DHW consumption does not necessarily align with the highest electrical consumption. For example, the peak electrical consumption occurs on December 10th, while the DHW consumption on that day corresponds to one of the days with the most dispersed DHW demand. However, the overall DHW demand remains high throughout the day, distinguishing it from the other days with reduced demand.

The data presented in the figures above illustrates the hourly average values over a five-day period for each apartment. Notably, the apartments 34, 35 and 36 show similar consumption profiles, with their higher peak occurring nearly simultaneously. However, the apartment 34 show a lower consumption outside of these peaks compared to the apartment 35 and 36, which have higher consumption levels throughout the day.

Then the apartments 64 and 65 also appear to share comparable consumption profiles over the five days, with consumption values varying within a range of 10^(-2) for each apartment and the same peak. It is noteworthy that, the apartment 64 shows a domestic hot water consumption higher than the apartment 65.

We can assume that the apartments 25, 26 and 27 share a similar pattern as well. Their peaks occur at nearly the same time, although some apartments have higher peak than others. Furthermore, their higher consumption is around 0.006 l/s for each apartment.

The observed patterns in this second visualization are clearly distinct from those in the initial one. This difference will influence our correlation analysis. Indeed, we will determine whether it is more advantageous to identify patterns based on mean values or time-series correlations, if feasible.

Furthermore, there is an observed consistency in consumption patterns within certain apartments over a five-day period, supporting our hypothesis of correlating apartments with themselves.

iii. Correlation analysis of the apartment together as a time series

To implement this analysis, you can refer to the code titled 'Correlation analysis between the apartments'.

Our initial time-series correlation analysis spans from December 7th to December 12th, 2020, covering a five-day period as described in the methodology. The correlation values were worked out for all apartments across these five days. The Figure 25 illustrates that the correlation values consistently remain below the 0.5 threshold, indicating a lack of correlation among the apartments.

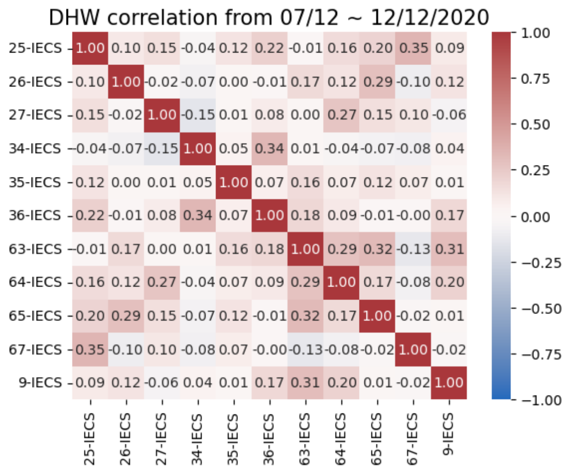

Figure 25: Correlation analysis

Consequently, our hypothesis of forming groups based on these specific days is not suitable. However, we aim to investigate whether this absence of correlation is exclusive to this particular time frame. To explore this, we will assess correlation between apartments on different days within December and then compare them with correlation on a different month, June. Our goal is to determine if correlation values persist for the same apartment within the same month and if they remain consistent across different months of the year.

Subsequently, to proceed, we aim to identify the highest correlation values and the corresponding apartments. Initially, the apartment 25 shows a correlation of 0.22 with the apartment 36 and 0.35 with the apartment 67. The apartment 26 has a correlation of 0.29 with the apartment 65. The apartment 27 exhibits a correlation of 0.27 with the apartment 64 but notably varies from the apartment 34, displaying a correlation of -0.15. The apartment 34, in turn, has one of the highest correlation values with the apartment 36 reaching 0.34. Lastly, the apartment 63 demonstrates notable correlation values with three other apartments: 0.29 with the apartment 64, 0.32 with the apartment 65 and 0.31 with the apartment 9. Our objective is to compare these correlation values with those obtained from the analysis covering one more day.

Continuing our analysis, we extend the time series correlation to encompass six days, from December 7th to December 13th, 2020. The correlation Figure 26 shows the relationship between all the apartments, revealing that none of them exhibit correlations surpassing the 0.5 threshold. Nevertheless, we aim to identify any differences in correlation coefficients between the previous analysis and this extended one.

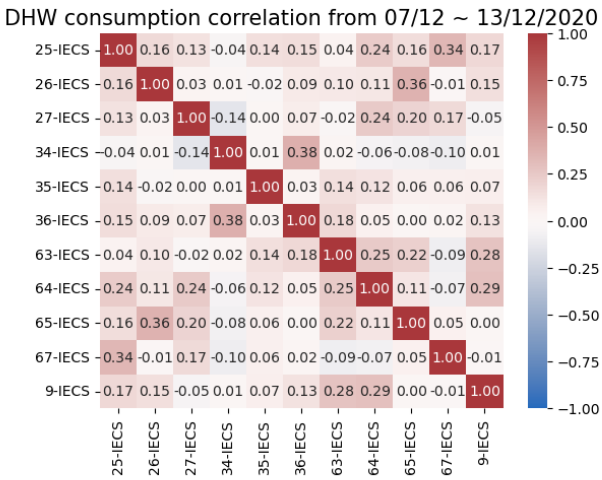

Figure 26: Correlation analysis

The apartment 25 maintains a high correlation with the apartment 67, although the correlation factor changes for the apartment 36 as it becomes lower. The observation suggests that the correlation values are subject to change over time. The apartment 26 appears to sustain a 'high' correlation with the apartment 65. The apartment 67 maintains a 'high' correlation with the apartment 64 and continues to differ significantly from the apartment 34. On the other hand, the apartment 34 shows the highest correlation with the apartment 36, reaching 0.38. Moreover, the apartment 63 maintains notable correlation values with the same three apartments as before: 0.25 with the apartment 64, 0.22 with the apartment 65 and 0.28 with the apartment 9. Lastly, the apartment 64 develops a higher correlation rate with the apartment 9, a change not previously observed.

In conclusion, the correlation values show variability over days, yet correlation values persist for most apartments. However, it is important to note that this analysis is an extension of the previous one, and we intend to compare the initial coefficients with those obtained from another week independent of December 7th to December 12th.

The analysis now shifts to the week from December 14th to December 19th, as showed in the Figure 27. As observed previously, the correlation coefficients once again fail to surpass the 0.5 threshold. Furthermore, comparing the coefficients between the week of December 7th and the week of December 14th reveals significant changes.

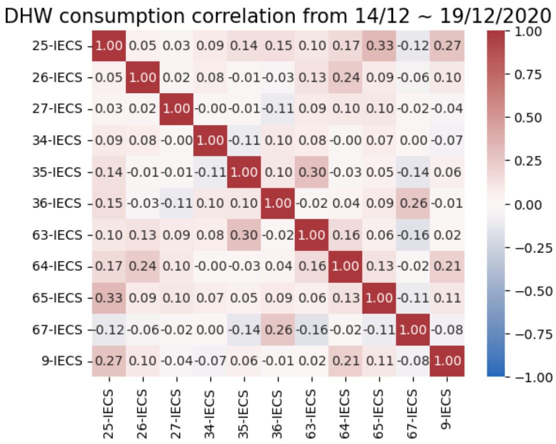

Figure 27: Correlation analysis

Indeed, the apartment 25 which previously exhibited a 'high' coefficient with the apartment 67 now shows a value of -0.12, indicating a considerable difference in consumption. Notably, the consumption is now more correlated with the apartment 65. The correlation between the apartment 26 and 65 is also lower than before. Additionally, the apartment 63 which previously had some of the highest coefficients with three other apartments, now demonstrates very low coefficients.

These observations underscore the dynamic nature of apartment consumptions, with correlation coefficients showing constant changes. Even with a one-week interval, both analyses yield results that differ significantly from one another.

Before adjusting our approach, we aim to explore the variations in correlation coefficients between two different months, specifically comparing December with June. The results are shown in the Figure 28.

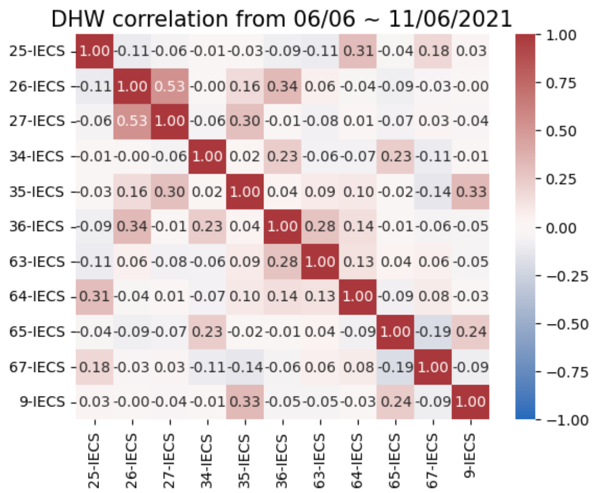

Figure 28: Correlation analysis

Surprisingly, one apartment, the 26, appears to surpass the 0.5 threshold when correlated with the apartment 27 in June, whereas the coefficient was near 0 for the same pair in December. This highlights the fact that the variation of correlation during different months makes it challenging to identify consistent patterns based on time series data.

Interestingly, the apartment 25 maintains a 'high' correlation coefficient with the apartment 64, similar to the December observation. However, the apartment 34 and 36, which displayed a 'high' correlation coefficient in the week of December 7^{th} and June, do not exhibit the same pattern in the week of December 14^{th}. This suggests that establishing a consistent correlation between these apartments is not possible.

In conclusion, identifying patterns between apartments based on time-series data proves to be difficult, and our initial hypothesis appears unsuitable.

iv. Correlation analysis of the apartments together using each apartments' hourly-average of a ten-day period

We now aim to explore if patterns emerge when examining the average consumption of apartments over ten days with a one-hour sampling interval.

To achieve this, we gather the data and resample it to an hourly frequency for each day. Consequently, for each apartment, we obtain ten values for each hour and the next step involves working out the hourly average of these ten values for every hour. This process provides us with the average consumption over a span of ten days for each hour. Our analysis consists of studying and comparing these average values for each apartment, seeking patterns between them.

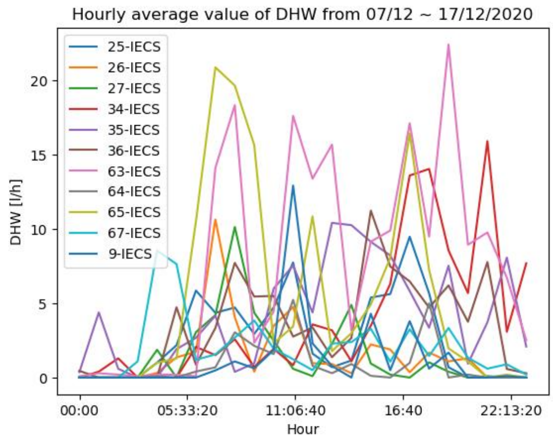

Figure 29 : Hourly average of the DHW consumption for each apartment

The initial results, presented in Figure 29 for the period from December 7^{th} to December 17^{th}, 2020, reveal that the consumption shows two peaks, one in the morning and another in the late afternoon. A notable pattern is observed between the apartments 65 and 63, where the peak values seem to align in time and are nearly equivalent. However, this observation is based on the average values of consumption. To enhance our understanding, we introduce the concept of band error values. Since each data point has its standard deviation, plotting these error bands can illustrate the range of values covered by all the apartments. This approach helps identify patterns and discern if certain apartments share similar consumption values. The average values along with their error bands are showed in Figure 30.

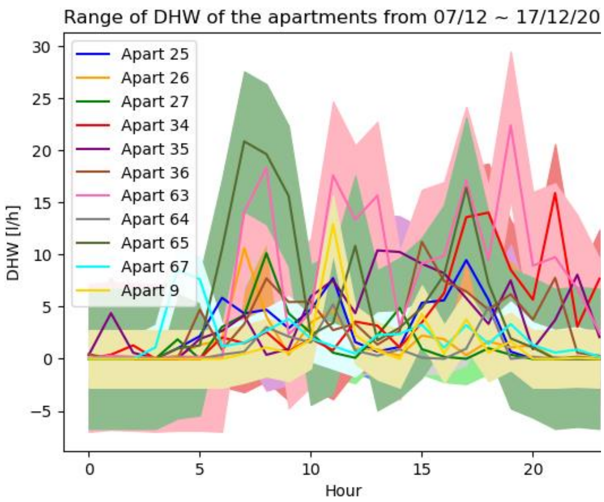

Figure 30 : Range of the DHW consumption with their band errors of the apartments

As observed previously, the apartments 65 and 63 show shared values, with the error band overlapping. Additionally, there is a resemblance between the apartment 9 and 67 in the late afternoon. However, due to the wide error bands of the apartment 63 and 65, identifying further patterns becomes challenging. To address this, the error bands for these apartments are removed in the Figure 31 and we only keep two of the error band.

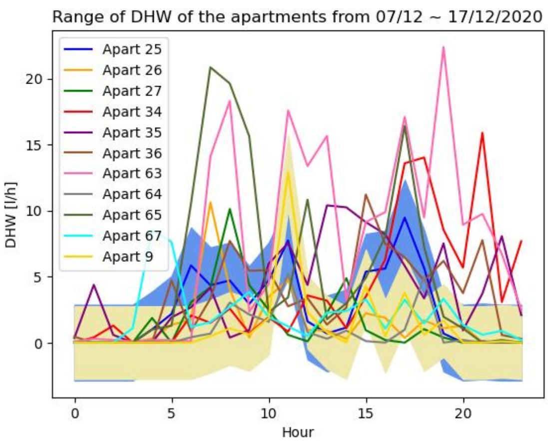

Figure 31 : Range of the DHW consumption with their band errors for the apartment 25 and 9

Further examination reveals that the apartment 9 (in yellow) and 25 (in blue) share common peaks at the same time. The blue line illustrating the apartment 25 remains mostly within the error band of the apartment 9.

We can also repeat this examination for all the apartment isolating the band error. However, we won't talk about it there and continue with our next analysis as these are the primary observations from the current visualization. The next step involves comparing these visual patterns with the correlation values to verify if there is indeed a correlation between the apartments.

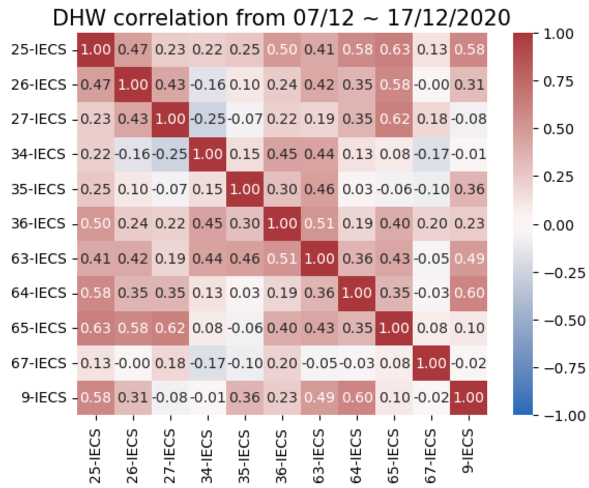

Figure 32 : Correlation analysis

In Figure 32, the correlation coefficients between each apartment are presented. Notably, using a threshold of 0.5, contrary to the previous time-series analysis, certain average values appear to be correlated. As anticipated, the apartments 25 and 9 exhibit a correlation of 0.58, surpassing the threshold. Moreover, the apartment 25 shows correlations with the apartments 36, 64 and 65. To form a group, the relationship between the apartments 9, 36, 64 and 65 is explored. While the apartment 36 is correlated with the apartment 63 and 25, it lacks correlation with the apartment 9, 64 and 65. However to constitute a group, all apartments within it should be correlated with each other. The apartment 64 correlates with the apartment 9, reaching a correlation 0.6. Lastly, the apartment 65 does not show correlation with either the apartment 9 or 64. From this analysis, it is deduced that a preliminary group of apartments sharing patterns can be formed, consisted of the apartment 9, 25 and 64.

Interestingly, the apartments 26, 27 and 63 are correlated with only one other apartment, suggesting the formation of pairs of apartments. The first pair includes the apartments 26 and 65, the second comprises the apartments 27 and 65 and the last pair is composed of the apartments 63 and 36. Meanwhile, the remaining apartments seem to operate on their own, lacking similar patterns with correlation coefficients consistently below 0.5.

Now the goal is to examine whether the groups formed earlier remain the same with a change of one day in the data. Figure 33 illustrates the correlation coefficients between the apartments from December 8th to December 18th, 2020.

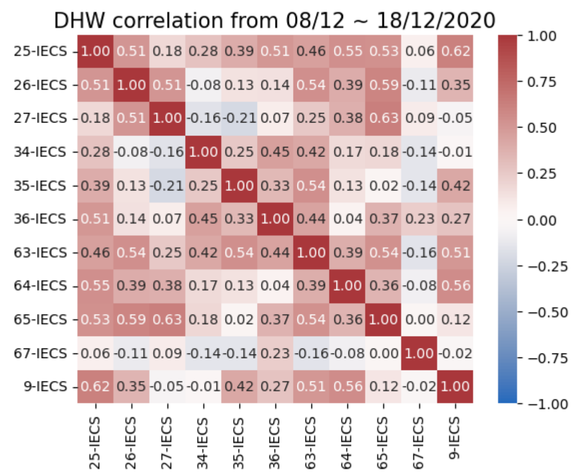

Figure 33: Correlation analysis

Notably, the apartment 25 still exhibits correlations with the apartments 36, 64, 65 and 9. However, the correlation coefficients, although similar, differ from the previous analysis. Additionally, this time, the apartment 25 appears to be correlated with the apartment 26 as well, a connection that wasn't present before. Considering the previous coefficient was near the threshold, this shift is not surprising, given the slight alteration in data from the earlier analysis.

Now let's compare the correlations among the other apartments to determine if the same groups persist in the analysis. The apartment 26 correlates with the apartments 25, 27, 63 and 65. In the previous analysis, it was only correlated with the apartment 65, but, as mentioned earlier, the coefficients were already close to the threshold. The apartment 64 maintains its correlation with the apartment 9, as seen before, and with the apartment 25. The apartment 65 correlates with the apartments 26, 27 and 63.

Independently, the apartment 9 shows correlation with the apartments 63, 64 and 25. The apartment 36 however, is only correlated with the apartment 25 which diverges from the previous analysis. These observations underscore that even with a one-day change, the coefficients can change, impacting the observed patterns.

From this analysis, three groups emerge. The first comprises the apartments 25, 64 and 9, which aligns with the findings from the previous analysis, which is reassuring. The second group consists of the apartments 25, 26 and 65, and given the previous correlation between the apartments 26 and 65, this result is not unexpected. A third group emerges with the apartments 26, 27 and 65, where the correlation between the apartments 27 and 65 was noted in the earlier analysis.

Notably, smaller groups of two can also be identified. The apartments 35 and 65 are correlated, marking a change from the previous analysis. This observation suggests that the groups established over a ten-day span may not consistently remain accurate or identical. It implies a dynamic nature, subject to change over time, which could potentially lead to misinterpretations in predictive modeling. Despite this variability, utilizing the correlation from ten days prior might offer insights into potential groupings for the following day. Next, we aim to explore variations observed in another month. The study of the month of June is now showed in Figure 34.

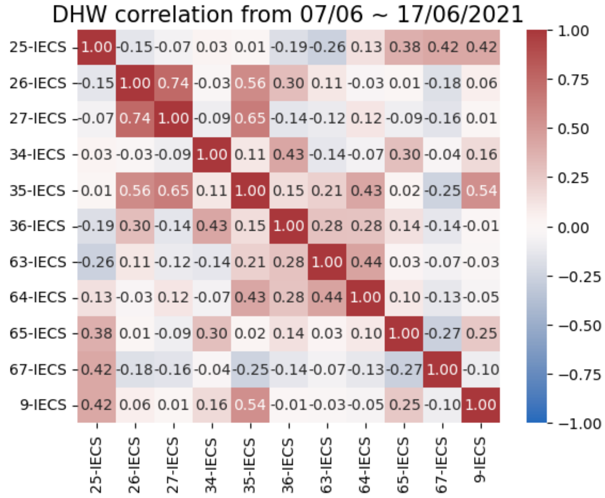

Figure 34: Correlation analysis

The aim of this examination is to discern disparities between the earlier observations and those derived from another month. We seek to see whether the groups previously identified remain the same or undergo significant changes.

A notable shift is observed in the apartment 25, which is not correlated with any other apartment this time. This observation contrasts to the previous analysis, where it shows high correlations, and underscores the dynamic nature of the relationship between the consumption. On the other hand, the apartment 26 is still correlated with the apartment 27. Additionally, the apartment 26 is correlated with the apartment 35, and intriguingly, the apartment 35 also shares a correlation with the apartment 27. Consequently, a first group can be formed, comprising these three apartments. However, this time, they do not exhibit the same relationship observed with the apartment 65 in the previous analysis. In this case, it appears that this is the only group that can be identified, except for a pair consisting of the apartments 35 and 9. The remaining apartments do not manifest correlations with one another, marking a significant change from the previous analysis.

We now aim to validate our hypothesis of forming group based on the correlation ten days before the prediction. Similar to our December analysis, we shifted our assessment one day after to examine if we could identify consistent groups. The Figure 35 represents the correlation between the apartments from June 8th to June 18th.

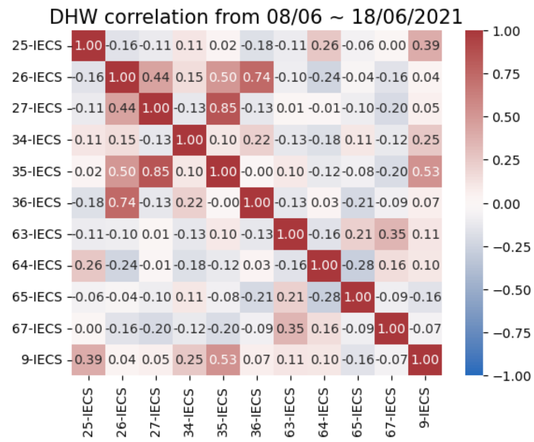

Figure 35: Correlation analysis

It appears that the apartment 26 still exhibits a correlation with the apartment 35. However, it is now linked with the apartment 36 instead of the apartment 27. Interestingly, the apartment 35 maintains its correlation with both apartments 26 and 27 but not with the apartment 36. Additionally, it continues to show correlation with the apartment 9.

This observation emphasizes the challenge of forming the same group as in our previous analysis. To make accurate prediction, it becomes clear that we have to examine the past results and correlations, enabling the formation of groups. Forecasting the consumption one month in advance with an hourly resampling over ten days seems challenging. Nevertheless, predict the future consumption of a group becomes more feasible by analyzing the consumption patterns of prior days.

v. Correlation analysis of the apartment together using the three-hours average of a ten-day period

Our current objective is to identify patterns by examining the average consumption of apartments over a span of ten days, with a three-hour sampling interval. Our goal is to observe daily variations divided into distinct periods. To start our study, we look at the Figure 36, which provides information about the DHW consumption.

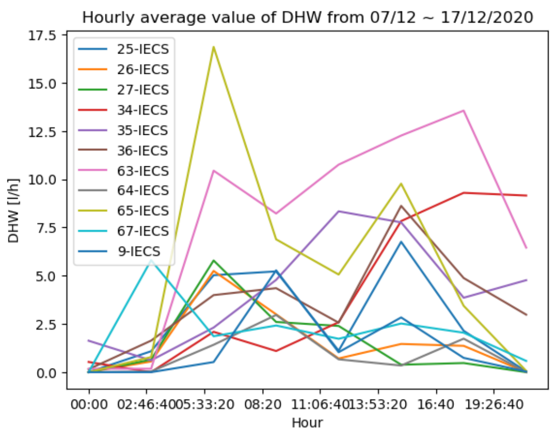

Figure 36 : Range of DHW consumption for each apartment with a resample of three-hour

As observed previously, the highest overall consumption still occurs in the late afternoon. Notably, the apartments 35, 36 and 65 appear to share the same pattern in the late afternoon, while the apartments 26 and 27 exhibit an identical pattern in the morning. To discuss these initial visualizations, we plan to conduct a correlation analysis. As said in the protocol, our attention is directed towards a specific ten-day period. The initial analysis focuses on the timeframe from December 7th to December 17th and is illustrated in Figure 37.

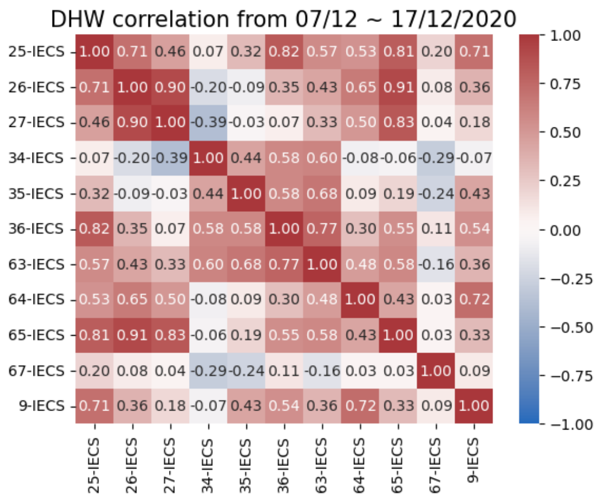

Figure 37 : Correlation analysis

In Figure 37, the correlation analysis for a three-hour resampling reveals anticipated patterns among apartments. The apartment 25 exhibits a high correlation with the apartments 36 and 65, with correlation coefficients of 0.82 and 0.81 respectively, surpassing by far our threshold of 0.5. Furthermore, the apartment 25 shows a notable correlation with the apartment 26 and 9, with correlation coefficients of 0.71. Additionally, it is also correlated with the apartments 63 and 64, although the coefficients are slightly lower, falling below 0.6.

Moving to apartment 26, it is strongly correlated with the apartments 25, 27, 63 and 65, reaching high coefficients around 0.9, far above the threshold.

The apartment 36 exhibits correlations with the apartments 25, 34, 35, 63, 65 and 9, even though the majority of the correlation coefficients are lower compared to those observed previously. Specifically, they are around 0.55, except for the correlations with the apartments 25 and 63.

The apartment 63 appears to share patterns with the apartments 25, 34, 35, 36 and 65. While the apartment 64 is correlated with the apartments 25, 26, 27 and 9. Then, the apartment 65 exhibits correlation with the apartments 25, 26, 27, 36 and 63. Lastly, the apartment 9 shows some correlation with the apartments 25, 36 and 64.

Based on this initial analysis, the most significant group comprises the apartments 25, 36, 63 and 65. Further examination suggests the formation of additional groups, such as the apartments 26, 27 and 65. Notably, the apartments 34 and 35 stand out as the only apartments correlated with two others, while the apartment 67 lacks any correlation with the other apartments.

In the upcoming analysis, we aim to observe how these groups evolves by shifting our data analysis by one day. The data for analysis spans from December 8th to December 18th.

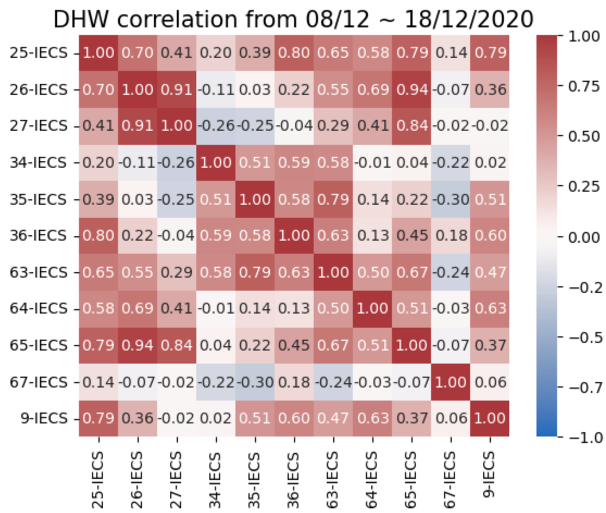

Figure 38 : Correlation analysis

In Figure 38, the correlation analysis among each apartment in the study case is done. The apartment 25 maintains its high correlation with the apartments 36 and 65, with coefficients still around 0.8, showing minimal change from the previous analysis. Interestingly, the apartment 9 now exhibits an even higher correlation with the apartment 25 compared to the previous analysis. Additionally, the apartment 9 continues to show strong correlations, around 0.7, with apartments 26 and 63. This indicates that the correlation between apartments 25 and 26 remains relatively stable, while the correlation with the apartment 63 increases. Notably, the correlation for the apartment 64 remains nearly the same.

Additionally, our objective is to examine whether the previously identified group remain consistent. The initial group encompasses the apartments 25, 36, 64 and 65 changes as the apartment 36 is no longer correlated with the apartment 65. This highlights the dynamic nature of domestic hot water consumption and its continuous fluctuations. However, the second group, consisting of the apartments 26, 27 and 65, persists in this analysis. It's notable that the correlation coefficients in this case were higher than those observed from December 7th to December 17th.

Our focus now shifts to observing the correlation between each apartment based on an analysis for another month. To achieve this, we shift our analysis month to June. The initial analysis, presented in Figure 39, covers the period from June 7th to June 17th.

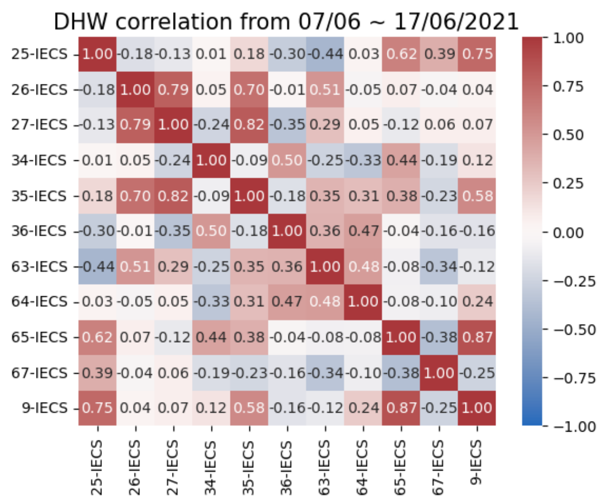

Figure 39 : Correlation analysis

In comparison to the month of December, there is a lower level of correlation among the apartments in June, as evident from the visual inspection of the figure. The color tone is now more centered around 0 on the scale, whereas in the previous analysis, it leaned more towards 0.75. Despite this shift, the apartment 25 remains correlated with the apartments 65 and 9, although not with others.

Examining the correlation between apartments 9 and 65 reveals a very high correlation, allowing the formation of a first group encompassing the apartments 25, 65 and 9.

The apartment 26 remains correlated with the apartment 27, and in this analysis, it is also linked to the apartments 35 and 63. Notably, the correlation coefficients are nearly equal to the previous ones. The apartment 27 seems to be correlated with the apartment 35 but not with the apartment 63, forming another group comprising the apartments 26, 27 and 63.

The apartments 34 and 36 continue to be correlated with each other exclusively maintaining the same pattern as in the previous analysis. However, the apartment 36, which was correlated with several apartments in December, now only correlates with the apartment 34.

The final observation is that the apartments 64 and 67 still lack correlations with other apartments in the study case. Interestingly, this remains consistent for the apartment 67, while for the apartment 64, it marks a change from the previous analysis.

The dynamic nature of the analysis, which represents the primary challenge in predicting consumption, is clear in this examination. Furthermore, it becomes apparent that the approach of grouping apartments based on correlation does not yield consistent results. The changing correlation patterns among apartments indicate that they are not consistently correlated with each other. Our hypothesis is then rejected due to the results' fluctuations.

vi. Correlation analysis of the apartments and their average with themselves

In this section, we aim to investigate the hypothesis suggesting that apartments may exhibits temporal correlation within their own consumption patterns, indicating a tendency for apartments to follow similar hot water usage patterns over time. Additionally, we explore the hypothesis that the average consumption of apartments tends to correlate more rapidly with itself than that of individual apartments. This proposition is grounded in the idea that predicting group behavior is often more straightforward than predicting individual behavior.

To explore these hypotheses, we conducted three analyses that can be found on the code 'Correlation analysis of apartment to themselves'. Initially, we assessed the correlation of individual apartments and their average across different months, spanning from December 2020 to June 2021. We employed various resample times, progressively increasing them when the correlation coefficients were insufficient (less than 0.5). Subsequently, we compared pairs of consecutive months, such as December to January, January to February and so forth, investigating the influence of different time steps on the correlation coefficients until optimal efficiency was achieved if possible. Finally, we compared two sets of three months data: December to February versus March to June. This allowed us to examine whether there were notable differences in correlation patterns between these distinct time frames.

Initially, we begin by comparing the monthly consumption patterns, starting with a resample interval of 10 minutes. We then increase the time-step to assess consumption on an hourly, daily, and finally, five-day bases. The following table presents the outcomes obtained through hourly resampling.

Table 1: Hourly correlation coefficient of the apartments and their average

Ap.	Dec.	Jan.	Feb.	Mar.	Apr.	May
25	-0.057	0.26	0.35	0.079	0.34	0.
26	-0.069	0.09	0.2	0.1	0.08	0.
27	-0.11	0.36	0.32	0.35	0.36	0.
34	-0.017	0.15	0.19	0.13	0.11	0.
35	-0.12	0.32	0.28	0.15	0.04	0.0
36	-0.093	0.37	0.31	0.26	0.2	0.
63	-0.11	0.13	0.33	0.24	0.11	0.
64	-0.12	0.36	0.21	0.16	0.17	0.
65	-0.18	0.21	0.15	0.17	0.17	0.
67	-0.02	0.36	-0.021	0.23	0.28	0.
9	-0.11	0.34	0.25	0.27	0.34	0.
mean	-0.288	0.48	0.52	0.52	0.38	0.

The examination of hourly consumption patterns for each apartment in December reveals a lack of correlation, as indicated by consistently negative values. Notably, even the average consumption for December does not display any interrelation. In January, however, there is a reversal in the tendency, with some coefficients hovering around 0.3 (through insufficient for a significant correlation) they are higher than those observed for December. Despite this, the average correlation coefficient for January is the highest within the set and approaches a level indicative of correlation. The positive correlation trend remains the same in February, March, April, May and June, with a few exceptions. Notably, there is a correlation in the average values of February and March, with a coefficient of 0.52. This suggests the possibility that having a water tank with a one-hour storage capacity could be viable for both months.

To explore whether this correlation tendency extends to more months and apartments with increased resample time, we proceed to examine the values daily. The subsequent table presents the consumption values for apartments across the months.

Table 2: Correlation coefficient of the apartment and the average with a daily resample

Ap.	Dec.	Jan.	Feb.	Mar.	Apr.	May
25	0.07	0.29	0.043	0.019	0.03	0.555
26	-0.092	0.005	0.31	0.084	0.23	-0.244
27	-0.103	-0.17	-0.19	-0.14	-0.22	-0.17
34	0.037	0.1	0.32	-0.084	-0.16	0.2
35	-0.442	-0.25	0.091	0.18	0.19	-0.17
36	0.461	-0.0419	-0.04	0.12	0.03	-0.04
63	-0.12	-0.077	0.18	-0.067	-0.44	0.035
64	0.248	0.038	-0.18	-0.24	-0.21	0.39
65	-0.044	0.16	0.5	0.041	0.12	0.13
67	0.317	0.0039	-0.22	-0.012	0.12	0.76
9	0.52	-0.09	0.37	0.27	-0.22	0.7
mean	0.101	-0.2	0.58	0.23	-0.28	-0.006

Upon employing a daily-resample, notable changes become apparent. Specifically, in December, where apartment 9 exhibits a correlation with itself, and apartment 36 also shows a proximity in correlation. Some values increase, including the average correlation coefficient for December, indicating a shift towards a stronger correlation. However, in January, there is a noticeable decline in correlation coefficients, with several turning negative, indicating a decreasing tendency for apartments to correlate with themselves. This pattern persists across the other months, where overall correlation coefficients are lower.

Notably, the sole consistency lies in the average correlation for February, which maintains a correlation with a coefficient of 0.58. To further explore this trend, we aim to investigate if increasing the time step to five days will alter the observed patterns. This is made with the expectation that an increased time step may lead to a smoothing effect on the values. The subsequent table presents the results for the months with a five-day resample basis.

Table 3 : Correlation coefficient of the apartment and the average with a five-day basis resample

Ap.	Dec.	Jan.	Feb.	Mar.	Apr.	May
25	0.986	-0.11	0.88	0.95	0.46	0.75
26	-0.441	-0.62	0.29	0.68	-0.36	-0.36
27	-0.505	-0.52	-0.94	-0.6	0.49	-0.97
34	-0.67	-0.998	-0.91	0.64	-0.58	-0.98
35	-0.578	-0.35	0.42	0.29	0.99	0.99
36	-0.412	0.33	0.3	0.88	-0.5	-0.73
63	0.61	-0.36	0.96	0.36	-0.31	0.37
64	-0.856	0.91	-0.65	-0.74	0.27	0.88
65	0.388	0.62	0.79	0.49	-0.93	0.64
67	0.426	0.85	-0.53	0.43	-0.99	0.88
9	0.92	-0.07	-0.99	0.58	-0.95	0.68
mean	0.504	-0.72	0.89	0.9	-0.89	-0.45

As expected, a significant shift in the data occurs. Across each month, certain apartments now exhibit correlations with themselves. In December, the apartments 25, 63 and 9 display self-correlation, and the average value demonstrates a notable correlation as well. January follows suit, with three apartments (64, 65, 67) showing self-correlation, a change from the lack of such correlations previously observed. The trend continues in February, maintaining high average correlation values and self-correlation in the apartments 25, 63 and 65.

Moving to March, five apartments (25, 26, 34, 36, 9) exhibit self-correlation, and the average correlation is also high. May shows a pattern similar to January, with numerous negative coefficients, but some apartments (25, 35, 64, 65, 67, 9) display self-correlation. April stands out with predominantly negative values, except for the apartment 35.

Notably, the apartments 25, 64, 65, and 9 consistently show self-correlation throughout the months. This suggests the potential for predicting consumption for these apartments on a five-day basis and the possibility to identify a water tank with a five-day storage capacity to harness energy from solar panels. On its own, the average is not as promising as we anticipated, as it reveals an equal occurrence of self-correlation and the lack thereof.

While these findings are promising, we aim to assess whether consumption correlation persists over time by comparing not one month to itself but two months together. The following table presents our results.

Table 4 : Daily basis correlation between two months

	Dec-Jan	Jan-Feb	Feb-Mar	Mar-Ap	Ap-May
25	-0.031	0.15	0.18	0.28	0.18
26	-0.0001	0.49	-0.12	0.45	0.36
27	0.143	0.36	0.4	0.41	-0.18
34	-0.181	-0.1	-0.17	0.2	0.23
35	0.117	0.185	0.24	0.36	0.29
36	0.129	0.14	0.12	0.19	0.02
63	-0.065	0.46	0.49	-0.007	0.22
64	-0.127	0.34	0.26	-0.11	0.04
65	-0.189	-0.34	0.47	-0.02	0.26
67	0.27	0.304	-0.05	0.42	0.16
9	-0.0418	0.44	0.51	0.3	0.19
mean	0.2	0.38	0.66	0.52	0.67

Now, we aim to examine the potential self-correlation of apartments over a two-month period, starting with a daily basis resample. For December and January, as well as January and February, there doesn't appear to be self-correlation for each apartment or their average. The trend continues between February and March, with only the apartment 9 showing self-correlation and a correlation observed for the average. The other coefficients are promising but still under the threshold of 0.5. When analyzed individually, March and April display correlation with the average, as does April and May.

Notable, it seems that over a two-month span, the average exhibits more promise than when considering only one month. However, we are interested in exploring how these correlations evolve when the time step is increased, aiming to determine if correlations strengthen or decrease. To investigate this, we increase the time step to five-day basis and the results can be examined in the following table.

Table 5 : Five-day basis correlation for the apartment and their average comparing two months

	Dec-Jan	Jan-Feb	Feb-Mar	Mar-Ap	Ap-May
25	-0.03	0.28	-0.33	0.67	0.58
26	-0.22	0.23	-0.67	0.4	0.39
27	0.44	0.73	0.45	0.33	0.18
34	-0.6	0.62	-0.65	0.33	0.28
35	0.13	-0.5	0.09	0.22	0.87
36	0.63	-0.13	0.31	0.18	-0.63
63	-0.33	0.76	0.78	-0.48	0.04
64	-0.23	0.14	0.05	0.32	0.66
65	-0.84	-0.76	0.8	-0.38	0.26
67	0.59	0.45	0.06	-0.15	0.47
9	0.18	-0.28	0.35	-0.433	0.18
mean	-0.1	-0.14	0.81	-0.21	0.47

As previously noted, the increase in resample time leads to more correlation of apartments with themselves. While December and January continue to maintain their trend of not being self-correlated, some apartments, such as 36 and 67, now exhibit self-correlation through the months. Similar observations are made for January to February, where the apartments 27, 34 and 63 show self-correlation. However, the tendency is less promising for February to March, with diminishing coefficients, although the average still demonstrates self-correlation for those months. March to April and April to May follow a similar pattern, with one apartment showing self-correlation (25 for March and April, and 35 for April and May), while others do not.

Notably, the average is not as promising as seen before, with only one instance of self-correlation between February and March, compared to the three instances observed previously. It appears that finding a tank for several apartments, using a daily resample might be more interesting. This implies the need for a storage tank that can accumulate solar panel energy for multiple apartments over one day.

To further validate this tendency, we aim to observe if it persists with a ten-day resample and whether a one-day resample remains more advantageous. The results are presented below.

Table 6 : Ten-day basis correlation of the apartment and their average for two months.

10-days	Dec-Jan	Jan-Feb	Feb-Mar	Mar-Ap	Ap-May
25	0.999	0.34	0.39	-0.54	0.97
26	-0.632	0.85	0.35	0.51	0.78
27	0.387	0.66	0.51	0.6	0.71
34	-0.553	0.999	0.94	0.72	-0.24
35	0.082	-0.74	0.89	0.29	0.8
36	0.64	-0.99	0.99	0.67	-0.58
63	-0.822	0.87	0.89	-0.71	0.95
64	-0.565	-0.95	-0.17	0.61	0.96
65	-0.937	-0.74	0.8	-0.49	0.1
67	0.629	-0.28	-0.94	-0.13	0.62
9	-0.395	0.59	-0.22	-0.91	-0.95
mean	-0.28	-0.28	0.99	-0.81	0.68

In comparison to the previous analysis, December and January still exhibit some negative coefficients, but now three apartments (25, 36 and 67) show self-correlation between both months. It's noteworthy that the apartments 36 and 67 were also the ones displaying self-correlation with a five-day resample. Concerning January to February, the tendency toward self-correlation intensifies, with now five apartments (26, 27, 34, 63, 9) self-correlating. This trend becomes even more pronounced between February and March, with six apartments (27, 34, 35, 36, 63, 65) self-correlating, and the average still showing self-correlation. March and April also reveal some self-correlation among apartments, including 26, 27, 36 and 64. Finally, April to May show self-correlation with the apartments 25, 26, 27, 35, 63, 64 and 67, with their average also exhibiting self-correlation.

Notably, certain apartments consistently show high self-correlation across both months, namely 26, 27, 34, 35 and 63. It may be worthwhile to consider predictions for these apartments using a ten-day basis resample to estimate hot water consumption and determine the required tank for each apartment. Alternatively, an average-based prediction on a daily basis might be more suitable, as the average shows less self-correlation with a ten-day basis and is linked to a bigger tank.

Moving forward, we aim to explore the correlation between three-months, starting with a daily basis, as in the previous analysis.

Table 7 : Daily basis correlation coefficients for the apartments and their average comparing three months.

Ap.	01 Dec ~ 28 Feb to 01 Mar ~ 28 May
25	-0.156
26	0.031
27	-0.043
34	-0.083
35	-0.0357
36	0.092
63	-0.153
64	0.144
65	0.146
67	0.398
9	0.2
mean	0.237

It is noteworthy that none of the apartments, including the average value, demonstrate self-correlation when considering three months together. This indicates a lack of correlation for the apartments with themselves when comparing both sets of three months.

To explore whether the observed trend tends to shift or remain consistent, we proceed to increase the time step to ten days. The results are shown below.

Table 8 : Ten-day basis correlation coefficients for the apartments and their average comparing three months.

Ap.	01 Dec ~ 28 Feb to 01 Mar ~ 28 May
25	0.0218
26	0.652
27	-0.086
34	0.16
35	0.183
36	-0.169
63	-0.584
64	0.264
65	0.085
67	0.787
9	0.0667
mean	0.0123

The increase in time step introduces a smoothing effect on the values, as evidenced by the correlation of the apartments 26 and 67 with themselves across both periods. This indicates that their consumption from December to February aligns with that from March to May. However, it's crucial to note that this correlation is observed only in two out of eleven apartments. Additionally, the average still exhibits a very low level of correlation.

As in previous analyses, we continue to increase the time step for a more thorough examination. The results are presented below.

Table 9 : Monthly basis correlation coefficients for the apartments and their average comparing three-months.

Ap.	01 Dec ~ 28 Feb to 01 Mar ~ 28 May
25	-0.34
26	0.727
27	0.22
34	0.557
35	0.874
36	0.036
63	-0.874
64	-0.966
65	-0.648
67	0.988
9	-0.131
mean	-0.525

In this observation, we note that only four apartments (26, 34, 35, 67) exhibit self-correlation during the period. This implies that the consumption patterns of these four apartments from December to February align with those from March to May. However, even the average value displays a very low level of correlation and does not indicate self-correlation.

Comparing this experience with the previous analyses, it becomes apparent that finding correspondences on a three-month basis is more challenging. It seems easier to identify consumption correspondences between two consecutive months than across a span of three months. This difficulty may be attributed to seasonal changes, as the consumption patterns in December may not align with those in May due to variations in weather and habits.

IV- Conclusion

In conclusion, upon visualizing the data, it appeared that there could be some correlation, with patterns seeming to repeat. Initially, the average consumption of all apartments suggested certain units sharing similar patterns. However, when we did the correlation analysis to assess if apartments were correlated together, the coefficients proved to be highly fluctuating, making it a challenge to identify consistent correlations. In one case, a specific group was identified, but the addition of a sole day to the case altered the composition of the group, highlighting the dynamic nature of the process.

Subsequently, we explored our second hypothesis by visualizing apartment consumption over five days to detect recurring patterns. While some apartments exhibited seemingly similar peak patterns, a direct comparison using two dataframes revealed inconsistencies. Perhaps comparing dataframes in this manner was not the optimal approach. Upon further reflection, exploring the visualization figures to identify consistent patterns on a few days might be more effective. Using the Dynamic Time Warping method could have been more efficient too. For instance, when visualizing the data for December, it appeared that some values were replicating themselves. However, the method of comparing two dataframes for December did not reveal significant correlations among the apartments. It's possible that dividing the month into two large dataframe might not have been the most suitable approach, and a more nuanced comparison of smaller time set could be more insightful.

Despite the initial unexpected results, increasing the resampling time revealed that certain apartments indeed exhibit patterns consistent with themselves. For example, if individual tanks are assigned to each apartment, using a resample of five or ten days may be effective, as it appeared to bring together multiple apartments when comparing consumption within a month of between two months. However, when extending the comparison to three months, such self-correlations were not as prominent, likely due to significant variations in consumption. It's worth noting that our analysis did not consider external factors such as changes in weather or other related variables. Incorporating these factors might provide a more comprehensive understanding of consumption patterns.

The average value consistently showed variations; for instance, when comparing one month, correlations were higher for the five-day basis, whereas comparing two months, the daily correlations proved to be more efficient. Despite some promising results when comparing one or two months, it remains challenging to identify a consistent trend that could serve as a foundation for predicting consumption.

Finally, in our analysis, we employed a smoothing approach by using average values for preprocessing the data. Exploring alternative preprocessing methods such as scaling or detrending could be beneficial to assess whether our results undergo any significant changes.

V- References

1. Ibrahim Ali Kachalla, Christian Ghiaus. Reinterpreting electric water boiler energy prediction: state-of-the-art review of influence factors, techniques, and future directions. Energy, 22/11/2023.
2. Kaiser Ahmed, Petri Pysly, Jarek Kurnitski. Monthly domestic hot water profiles for energy calculation in Finnish apartment buildings. Energy and Buildings, Vol 97, 2015, Pages 77-85. ISSN ISSN 0378-7788, https://doi.org/10.1016/j.enbuild.2015.03.051 [Consulted on: 30/10/2023]

3. Lukas G. Swan, V. Ismet Ugursal, Ian Beausoleil-Morrison. Occupant related household energy consumption in Canada: Estimation using a bottom-up neural-network technique. Energy and Buildings, Volume 43, Issues 2–3, 2011, Pages 326-337. URL: https://doi.org/10.1016/j.enbuild.2010.09.021 [Consulted: 31/10/2023]

4. Cao, Sheng & Hou, Shengya & Lu, Jie. Predictive control based on occupant behavior prediction for domestic hot water system using data mining algorithm. Energy Science & Engineering. 2019. 7. 10.1002/ese3.341. URL: https://www.researchgate.net/publication/332371972_Predictive_control_based_on_occupant_behavior_prediction_for_domestic_hot_water_system_using_data_mining_algorithm [Consulted on: 31/10/2023]

5. Mohammad Ebrahim Banihabib, Pezhman Mousavi-Mirkalaei. Extended linear and non-linear auto-regressive models for forecasting the urban water consumption of a fast-growing city in an arid region. Sustainable Cities and Society, Volume 48, 2019. URL : https://doi.org/10.1016/j.scs.2019.101585 [Consulted: 03/11/2023]

6. Wang, Zhaocai & Wu, Xian & Wang, Huifang & Wu, Tunhua. Prediction and analysis of domestic water consumption based on optimized grey and Markov model. Water Supply. 2021. 21. 10.2166/ws.2021.146. URL: https://www.researchgate.net/publication/351674528_Prediction_and_analysis_of_domestic_water_consumption_based_on_optimized_grey_and_Markov_model/references [Consulted on: 31/10/2023]

7. Lazzari Florencia, Mor Gerar, Cipriano Jordi, Gabaldon Eloin Grillone Benedetton Chemisana Daniel, Solsona Francesc. User behavior models to forecast electricity consumption of residential customers based on smart metering data. Published on 21/02/2022 by Elsevier Ltd. (Online) Accessible on: https://www.sciencedirect.com/science/article/pii/S2352484722005078 [Consulted on 19/10/2023]

8. Jae Yong Lee, Taesu Yim. Energy and flow demand analysis of domestic hot water in an apartment complex using a smart meter. Energy, Volume 229 2021, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2021.120678 [Consulted on: 30/10/2023]

9. Wojciech Rzeznik, Ilona Rzeznik, Pawel Hara. Comparison of Real and Forecasted Domestic Consumption and Demand for Heat Power Buildings in Poland. Energies 2022, 15, 6871. URL: https://doi.org/10.3390/en15196871 [Consulted on: 31/10/2023]