economie
FOD Economie, K.M.O., Middenstand en Energie

FlexMyHeat

D2.1 – Overview of models for flexibility, local constraints, energy markets and flex controllers

Table of Contents

I. INTRODUCTION 5

II. BATTERY SMART CONTROL ALGORITHMS 6

II.1. Day-ahead Market 6

II.1.1. MDP Formulation 6

II.2. Imbalance Settlement Mechanism 6

II.2.1. MDP Formulation 7

II.3. Reinforcement Learning 7

III. RESULTS 9

III.1. Day-ahead Market 9

III.2. Imbalance Market Results 12

III.3. Social welfare 13

IV. CONCLUSION AND NEXT STEPS 15

V. REFERENCES AND INTERNET LINKS 16

Table of Figures

Figure 1 Average household monthly electricity bill in 2030 10
Figure 2 Contribution of different assets for scenario 3 in different seasons of 2030 12
Figure 3 Average daily peak power spread for scenario 3 12
Figure 4 Control logic of 3kW/ 5kWh battery in imbalance market 13

List of Tables

Table 1 Overview of the average annual electricity bill in 2030 for the day-ahead market..... 9
Table 2 Overview of the average annual electricity bill in 2030 for the imbalance market..... 12

I. INTRODUCTION

The FlexMyHeat project aims at understanding the role that heat pumps and decentralized storage solutions will play in 2030 and 2050 as a source of flexibility for the national electricity system.

The extra need for electricity by shifting from fossil fuel based heating systems to heat pumps in the upcoming years will increase peak loads in the Belgian electricity grids (on top of increased peak loads caused by other domains that get electrified such as mobility and industry) and will thus lead to challenges with respect to the energy security of supply, the net balance and (on a more local level) to congestion of the grid infrastructure.

However, by properly controlling these heat pumps in combination with local storage solutions, unlocking the available flexibility, this challenge can be turned into an opportunity for the grid, contributing to the national and regional balance of the Belgian electricity system.

The goal of FlexMyHeat is to quantitatively analyze the impact and value of the increased deployment of heat pumps and decentralized electrical/thermal storage in 2030 and 2050 on the Belgian electricity system, including proposed control/coordination strategies at (a combination of) various timescales, ranging from day-ahead markets to imbalance markets.

This quantitative assessment is performed for different scenarios:

- *Business-as-usual:* considering the heat pumps and possibly associated local storage as independent devices, only optimized for local objectives, i.e., maximizing PV self-consumption. Thus, no dynamic interaction from the grid side to exploit their flexibility.
- *Individual smart control:* optimized control of the flexibility opportunities offered by the heat pump or storage devices individually, so assuming that any other devices are only optimized for PV self-consumption maximization.
- *Integrated smart control:* combined optimization of both the heat pump and storage devices for local and market objectives

D1.1 focused on the business-as-usual scenarios while this deliverable focuses on the individual smart control assessment of battery systems. Finally, D3.1 will present the results of the individual smart control of heat pumps and thermal storage systems, and the results for the integrated smart control scenarios.

This deliverable is structured as follows:

- **Section II** describes the Reinforcement Learning (RL) based control algorithms for controlling a home battery for day-ahead and imbalance markets.
- **Section III** describes the results on energy cost reduction and peak load reduction for the two analyzed markets, and analyzes the consumption patterns of the battery storage, heat pump and thermal storage.
- Finally, in **Section IV**, we provide our conclusions and describe next steps.

II. BATTERY SMART CONTROL ALGORITHMS

As we discussed in the previous deliverable, a simple rule-based controller cannot fully unlock the potential for flexible assets. In this section, we introduce our data-driven controllers for batteries to participate in two electricity markets, i.e., day-ahead market and imbalance settlement mechanism.

II.1. Day-ahead Market

Controlling batteries unlocks multiple value streams—such as boosting household self-consumption, reducing energy bills, and participating in grid services—maximizing both economic and environmental benefits. However, developing an optimal controller for batteries in home energy management systems (HEMS) is a challenging task due to the involvement of sequential decision-making under uncertainties caused by PV generation and household consumption, including heat pump consumption. We use a reinforcement learning-based controller to deal with these uncertainties, where the agent learns a (near-)optimal policy by interacting with its environment [1].

II.1.1. MDP Formulation

We formulate the home energy management problem as a Markov decision process (MDP) to minimize the daily energy cost of households by managing the (dis)charging of their home battery [2]. At each step, the state for each household is determined as follows

$$s_t = (t, SOC_t, \pi_t, P_t^{PV}, P_t^{load}, P_t^{HP})$$

where t is the hour of day, SOC_t shows the state of charge (SoC) of the battery at time t, π_t indicates the electricity price at time t, P_t^{PV}, P_t^{load}, and P_t^{HP} represent the PV generation of the household, the non-flexible load consumption, and the heat pump consumption, respectively.

The agent can take one of the 5 possible discretized actions as follows

$$a_t \in A, \qquad A = \{-P_{max}, -\frac{P_{max}}{2}, -\frac{P_{max}}{4}, -\frac{P_{max}}{10}, 0, \frac{P_{max}}{10}, \frac{P_{max}}{4}, \frac{P_{max}}{2}, P_{max}\}$$

where P_{max} is the maximum (dis-)charging power of the battery. The agent makes a decision every hour.

Since the RL agent aims to maximize the total reward, the reward function is defined as the negative of the energy cost, as shown below.

$$r_t = \begin{cases} -P_t^{agg}\, \pi_t^{buy}, & P_t^{agg} > 0 \\ -P_t^{agg}\, \pi_t^{inj}, & P_t^{agg} \leq 0 \end{cases}$$

$$P_t^{agg} = P_t^{load} + P_t^{HP} + a\_t - P_t^{PV}$$

We consider that households are exposed to the day-ahead market and the injection offtake prices are calculated based on the price formula for the dynamic price contract.

II.2. Imbalance Settlement Mechanism

To help maintain the grid balance, transmission system operators (TSOs) outsource part of the required corrective balancing actions to balance responsible parties (BRPs). At the end of each imbalance settlement period (ISP), unbalanced BRPs are exposed to an imbalance price, calculated based on the total system imbalance to penalize deviations against the system balance. In certain European countries, e.g., Belgium, BRPs can also be remunerated if their unbalanced portfolio contributes to restoring grid balance, i.e., if their deviation is opposite to

the direction of the system imbalance. However, performing implicit balancing is challenging, as it involves risks mainly stemming from the volatility of imbalance prices and prediction errors in RES production. The fact that the imbalance price is calculated at the end of the 15-minute period means there is uncertainty for BRPs regarding the price they can get by deviating from their nomination to help balancing the grid. Hence, the activation of flexibility by BRPs will take place if they estimate with enough confidence that the final imbalance price will be greater than their activation costs, by monitoring closely the near-real-time imbalance data published by the TSO, namely Elia in Belgium.

II.2.1. MDP Formulation

The implicit balancing strategy we propose aims to optimize the imbalance cost of the BRP by controlling battery actions. To this end, we model the problem as a MDP and solve it as a stochastic sequential decision-making problem [3].

The state at each time step is expressed as

$$s_t = (T_{qh}, qh, mo, SOC_t, \hat{\pi}_t^{imb})$$

where $T_{qh} \in [0,14]$ is the minute of the quarter hour, $qh \in [0,95]$ is the quarter hour of the day, $mo \in \{1,2,\dots,12\}$ represents the month of the year, and $\hat{\pi}_t^{imb}$ is the indicative imbalance price of the current quarter-hour. Indicative prices are minute-based prices published by Elia in real-time, based on the latest grid situation, providing more information to BRPs. Since the actual imbalance prices are calculated ex-post, our agent can only make decisions based on these indicative imbalance prices.

The action space includes 3 possible actions as follows

$$a_t \in A, \qquad A = \{-P_{max}, 0, P_{max}\}$$

Since the RL agent aims to maximize arbitrage revenue, the reward function is defined as the negative of the imbalance cost, as below

$$r_t = -a_t \pi_{qh}^{imb} \Delta t$$

System dynamics in RL are modeled using a state transition probability function. The battery dynamics form part of the system dynamics in our problem, which we explicitly formulate using a battery linear model with a constant round-trip efficiency for (dis)charging. However, the transition function is primarily unknown, due to dependency on uncertainties in imbalance prices, regulation state, weather, and wind forecast error. By interacting with the environment, the RL agent can implicitly learn the transition function and these uncertainties.

II.3. Reinforcement Learning

We will solve the formulated MDP problems using RL methods. Recently, RL, as a model-free method, has attracted researchers' attention due to its remarkable performance in solving complex sequential decision-making problems such as playing games, robotic control, and autonomous driving. RL aims to learn a policy that maximizes the expected long-term reward. We focused on policy gradient RL methods where the policy is directly learned by an actor network. Among existing policy gradient methods, we choose soft actor-critic (SAC) for its superior sample efficiency and stability [4].

In this policy gradient-based method, the policy is learned by an actor network π_φ and the Q-function is approximated by a critic network Q_θ. The goal of the actor is to maximize the expected reward as well as maximize the entropy of the actor to encourage the agent to further explore the environment. The loss function of the actor network (J_π) is defined as

$$J_\pi(\varphi) = \mathbb{E}_{s\sim\rho^\pi, a\sim\pi_\varphi}\left[\alpha \ln \pi_\varphi(a|s) - Q_\theta(s,a)\right]$$

The critic network estimates the soft Q-value values. The loss function of the critic network (L_Q) is calculated as follows.

$$L_Q(\theta) = \mathbb{E}_{(s_t,a_t)\sim D}[(Q_\theta(s_t,a_t) - y_t)^2]$$

$$y_t = r(s_t,a_t) + \gamma \mathbb{E}_{a_{t+1}\sim\pi_\varphi}\left[Q_{\theta'}(s_{t+1},a_{t+1}) - \alpha \ln \pi_\varphi(a_{t+1}|s_{t+1})\right]$$

y_t is an estimated soft-Q value that is calculated by a modified Bellman equation (soft Bellman equation). In this step, the target value for the Q function is computed according to the maximum entropy objective. For calculating the target value for training the Q_θ network, using the same network can be prone to divergence in Q_θ updates. To solve this problem, another Q network is defined and parametrized by θ'. After each episode when Q_θ is updated, the parameters of $Q_{\theta'}$ are updated according to the following equation with $\tau \ll 1$ to slowly track the learned network.

$$\theta' = \tau\theta + (1-\tau)\theta'$$

III. RESULTS

This section aims to study the impact of applying the proposed battery controller in 2030 while heat pumps and thermal storage are controlled using rule-based business-as-usual control logic, thus only optimizing the self-consumption of local PV power. We used a similar case study to the one described in detail in deliverable D.1.1, i.e., 120 households across five different locations in Belgium with a PV installation.

The smart control scenarios are defined the same as the business-as-usual scenarios to be able to compare with as follows:

- Base scenario: base load with PV without any flexible asset
- Scenario 1: base scenario with battery
- Scenario 2: base scenario with battery and heat pump
- Scenario 3: base scenario with battery, heat pump, and thermal storage

III.1. Day-ahead Market

The results of controlling household batteries in the day-ahead market are summarized in Table 1. The smart control of batteries reduced the annual electricity bill on average by 3.26%, 4.88%, and 2.13% compared to the rule-based control logic in Scenarios 1, 2, and 3, respectively. Figure 2 explains the reason for this cost reduction. In the rule-based controller, the battery begins charging as soon as there is excess PV generation, leading to it being fully charged before the afternoon. As a result, peak battery consumption occurs in the morning. This means the household will miss the opportunity to charge the battery during the cheaper hours around early afternoon. In other words, the excess PV will be sold at lower prices. However, the RL agent charges the batteries around early afternoon, when the price is cheaper. In addition, the batteries will be charged more during cheap hours to avoid purchasing electricity from the grid in the evening when the price is more expensive. Average monthly electricity bills, shown in Figure 1, demonstrate that the benefit of using the RL agent is not limited to a certain month/season for an average household with the annual electricity consumption of 2653 kWh. The RL agent provides a constant cost reduction over the rule-based controller. It is worth mentioning that the model was trained on 2023 data, which contained fewer instances of negative prices. However, with negative prices now occurring more frequently during the spring and summer, it is important to train season-specific or even month-specific models to better address this issue.

Table 1 Overview of the average annual electricity bill in 2030 for the day-ahead market

Control Logic	*Base (€)*	*S 1 (€)*	*S 2 (€)*	*S 3 (€)*
RBC	557.34	370.56	835.45	837.69
RL	557.34	383.04	878.29	856

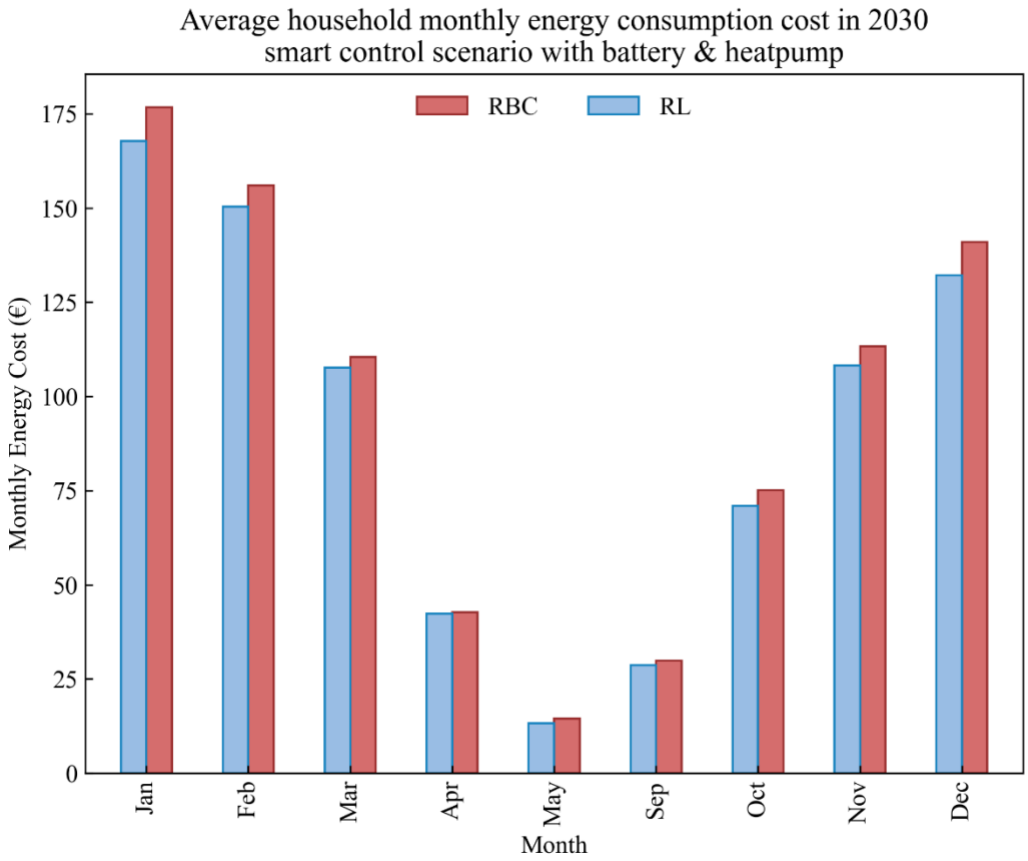
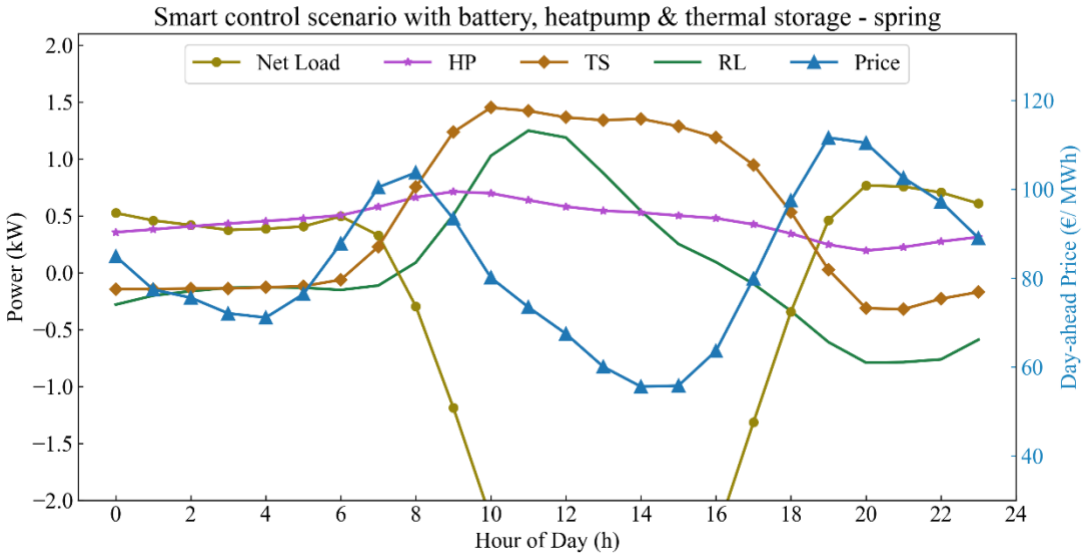
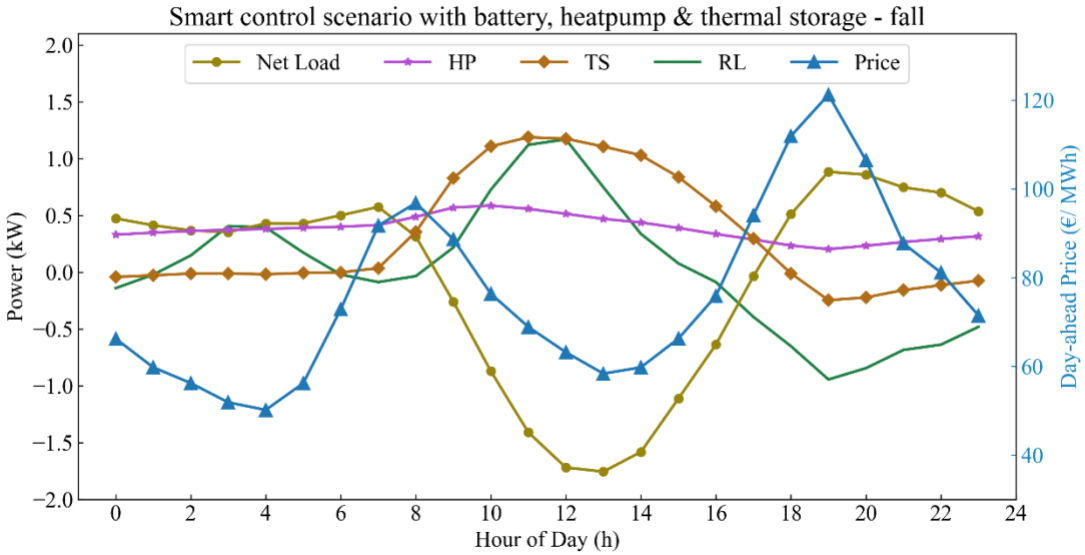


Figure 1 Average household monthly electricity bill in 2030

To better understand the contribution of each asset to the cost reduction, the average energy consumption of each asset during winter is shown in Figure 2. The battery consumption profile exhibits two main peaks: one in the early morning when electricity prices are low, and another around noon when PV generation is high. The thermal storage is charged during hours with an excess of PV to avoid selling to the grid and to maximize self-consumption. In the evening, when space heating is needed, the thermal storage is used first. This explains the dip in the heat pump profile around 17:00. Once the heat pumps begin operating for space heating and the thermal storage is completely empty, the batteries start discharging to supply the heat pumps with electricity.


Smart control scenario with battery, heatpump & thermal storage - spring
Net Load
HP
TS
RL
Price
Power (kW)
Day-ahead Price (€/ MWh)
Hour of Day (h)


Smart control scenario with battery, heatpump & thermal storage - fall
Net Load
HP
TS
RL
Price
Power (kW)
Day-ahead Price (€/ MWh)
Hour of Day (h)

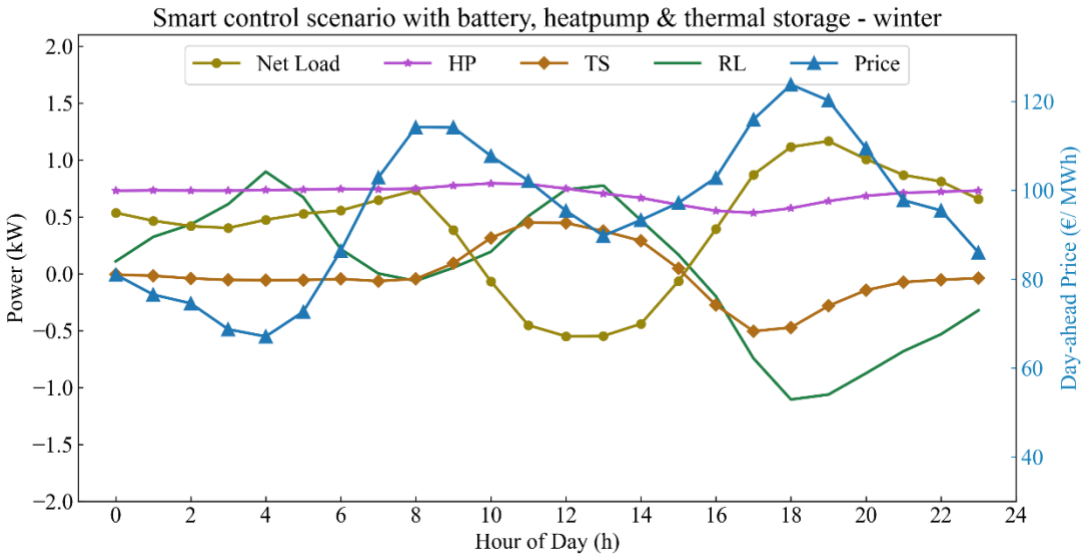

Smart control scenario with battery, heatpump & thermal storage - winter
Net Load
HP
TS
RL
Price
Power (kW)
Day-ahead Price (€/ MWh)
Hour of Day (h)

Figure 2 Contribution of different assets for scenario 3 in different seasons of 2030

Figure 3 shows the effect of the smart controller on the daily peak power reduction. Based on the results, by adopting the smart control logic, the trained RL agent can reduce the daily peak power on average by 11.5% and 5.6% compared to the base and rule-based controller cases.

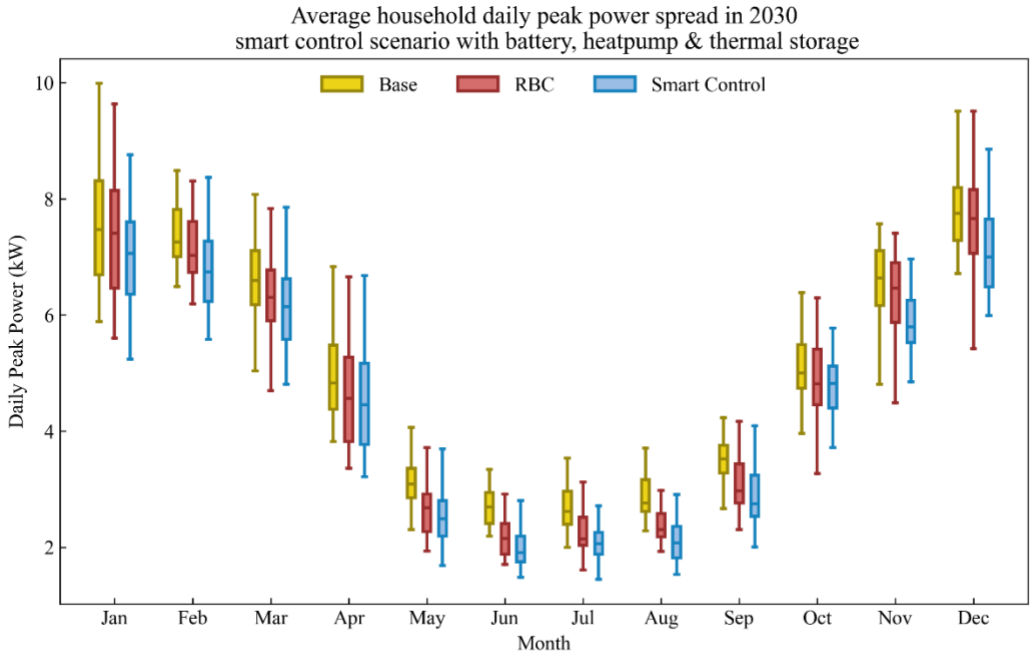


Figure 3 Average daily peak power spread for scenario 3

III.2. Imbalance Market Results

Table 2 also shows the impact of controlling the batteries for the imbalance market, on the average annual electricity bill of households. The results are presented for different household revenue-sharing rates, ranging from 20% to 40%. The remaining profit will be shared among other stakeholders involved, such as balance responsible parties (BRPs) — for participating in the market and managing portfolios —, energy suppliers — for providing customer contracts —, etc. According to the results, the households can decrease their annual electricity cost by 367.8€ with a 20% revenue-sharing rate. The battery consumes an average of 5.7 cycles per day. Figure 4 shows the control logic of the RL agent in the imbalance market. If the imbalance price is lower than -200€, the agent charge batteries. At very high prices (above 600€), the agent always discharges the batteries. When the price is between -200€ and 600€, the agent decision depends on the battery state of charge (SoC). When the SoC is low, the agent becomes more conservative with discharging. This ensures that the battery is not completely depleted, reserving some energy for future arbitrage opportunities.

Table 2 Overview of the average annual electricity bill in 2030 for the imbalance market

revenue-sharing rate	*Control Logic*	*Base (€)*	*S 1 (€)*	*S 2 (€)*	*S 3 (€)*
0%	No control	557.34	557.34	1073.22	1018.02

20%	RL	557.34	189.59	705.47	650.27
30%	RL	557.34	5.72	521.60	466.40
40%	RL	557.34	-178.15	337.73	282.53

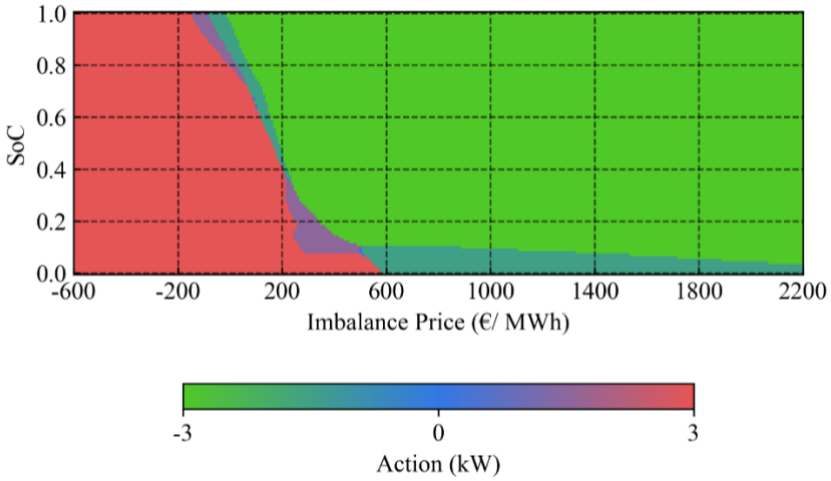


Figure 4 Control logic of 3kW/ 5kWh battery in imbalance market

III.3. Social welfare

The mechanism of imbalance market relies on the following principle:

- The TSO (Elia) receives bids for aFRR (automatic Frequency Restoration Reserve) for each quarter. Those bids are made by electricity producers or big industrial players who sell their flexibility upward or downward at a certain price.
- In order to unleash an even larger flexibility potential, the TSO created the imbalance market, which allows the Balance Responsible Parties (BRP) to deviate from their equilibrium position if that helps maintaining the grid in balance. The BRPs are incentivised thanks to a price signal, i.e. the imbalance market price, which is published in near-real-time on the TSO open data platform.
- When a BRP decides to activate some flexible assets in order to deviate from his allocation for the quarter, which effectively helped the grid, he will be rewarded at the imbalance market price, which follows a formula based on the aFRR marginal price during that quarter, but that is **strictly lower** than the aFRR price. The reason behind this is that the action of the BRP has avoided the need to recourse to aFRR provider. As a result, when a BRP reacted to an imbalance market price by activating upward or downward volumes, he is basically helping to maintain the power grid at a lower cost than it would have been with an aFRR provider.

  We call this **"social welfare"** as it reduces grid balancing costs. The TSO (Elia) is paying the BRP but then passes on this cost to all unbalanced BRPs during the quarter - not those who are unbalanced in a way that helped the grid. These costs are generally passed on to final consumers, hence social welfare implies lower balancing costs for the final consumers, which are citizens and companies.

- The BRP will then share his imbalance revenue with the other stakeholders that were involved:
    - The asset owner for the use of his asset
    - The Flexibility Service Provider (FSP) for the aggregation of assets and contracting with asset owners

As a result, the mechanism of imbalance market generates two types of revenues:

- Imbalance revenues - which are shared among the BRP, the FSP and the asset owner
- Social welfare, which is very complex to quantify, and consists of:
    - Lower balancing costs passed on to final consumers
    - Capacity Reserve Mechanism (CRM) costs savings, i.e. the mechanism through which the TSO can conclude capacity contracts to ensure the security of power supply in Belgium. Bidders offer their availability to produce when needed.

In its Adequacy & Flexibility study of June 2025, ELIA has estimated that by 2036, end-user flexibility can deliver €350 to €500 million in annual savings, with about €260 million from balancing savings, and between €90 and €260 million from CRM-related savings. These social welfare estimates encompass not only batteries, but also change in consumers behaviours, but nonetheless they show a significant savings potential by leveraging end-user flexibility.

IV. Conclusion and Next Steps

The results conclude that smart battery control provides greater electricity cost reduction compared to the rule-based controller. The results of participating in the day-ahead and imbalance markets indicate that controlling the battery in the imbalance market is more profitable for households, despite its higher uncertainty and greater complexity.

The next step is to study the impact of smart control for heat pumps and thermal storage. Additionally, developing a multi-asset controller to manage flexible assets simultaneously will help optimize electricity cost reduction and peak power management.

V. REFERENCES AND INTERNET LINKS

[1] G. Gokhale, B. Claessens, and C. Develder, “Sample efficient reinforcement learning for building control: Leveraging physics informed latent representations,” in *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*, 2024, pp. 496–497.

[2] S. S. K. Madahi, T. Van Puyvelde, G. Gokhale, B. Claessens, and C. Develder, “Multi-source Transfer Learning in Reinforcement Learning-based Home Battery Controller,” in *Proceedings of the 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 2024, pp. 341–345.

[3] S. Soroush Karimi Madahi, G. Gokhale, M.-S. Verwee, B. Claessens, and C. Develder, “Control Policy Correction Framework for Reinforcement Learning-based Energy Arbitrage Strategies,” in *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*, 2024, pp. 123–133.

[4] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in *International conference on machine learning*, PMLR, 2018, pp. 1861–1870.