ECOFLEX

With the support of the Energy Transition Fund

D4.1 Report on different smart charging management structure for electric vehicles

Version number and Date: Version 1; 28/02/2025
Author (s): Gilles Van Kriekinge (VUB), Saeed Naghdizadegan Jahromi (VUB)

Abstract for dissemination (PU)

In this document different charge management/energy management/charge scheduling algorithms will be investigated and developed. VUB will focus on the development of advanced charge scheduling algorithms focused on both implicit and explicit demand-side flexibility. Different configurations as multi-layered systems with multi-level control will be considered. The objective will be to make interact different charge scheduling algorithms and energy management systems in large system approaches. Particular attention will be given to investigation of stability issues linked to implicit demand-side flexibility with electric vehicles and capacity assurance inexplicit demand side flexibility.

Contents

List of abbreviations

BESS	Battery Energy Storage System
BRP	Balancing Responsible Party
BSP	Balancing Service Provider
DER	Distributed Energy Resources
DSO	Distribution System Operator
EMS	Energy Management System
EV	Electric Vehicle
FSP	Flexibility Service Provider
GEP	Green Energy Park
KPI	Key Performance Indicator
LES	Local Energy System
MV/LV	Medium-Voltage/Low-Voltage
MDP	Markov Decision Process
MPC	Model Predictive Control
PPO	Proximal Policy Optimization
RL	Reinforcement Learning
SOC	State-Of-Charge
TSO	Transmission System Operator
UFP	Universal Flexible Platform

List of Figures

1 Introduction

The transition to electric vehicles (EVs) in the transport sector presents significant challenges, particularly in managing the impact of EV charging on both local and global electric grids. Currently, the predominant method, known as uncoordinated charging, does not account for critical factors such as grid stability, electricity prices, solar energy production, building load demands, user charging needs, local grid constraints, and more. This approach can lead to increased peak power demand, voltage and frequency fluctuations, and a higher overall electricity demand, as documented in scientific research [1]. These issues highlight the need to coordinate the charging of electric vehicles, which is also often called “smart charging”.

Conventional methods to perform smart charging have primarily limit themselves to dynamic or static load balancing, which only limits charging power reactively based on responses to local grid measurements and grid limitations [2]. Instead, this task in the ECOFLEX project aims to go beyond the aforementioned method in order to significantly improve electric vehicle charging from economic, sustainable, and system-wide perspectives. From this, two distinct charging methods are presented, each fundamentally different and serving different purposes.

The first method presented in Section 2, is a model-based energy management system (EMS) that optimizes the charging power to minimize the cost of charging, as well as reducing the environmental impact of local energy systems. This method uses predictive information of the behaviour of the components (solar, building consumption, charging demand, …) to base the real-time decision on the future situation. The developed smart charging algorithm has two different modes; a) an implicit mode where the algorithm optimizes based on BELPEX day-ahead dynamic prices, and b) an explicit mode where the algorithm does implicit steering in addition to explicit steering from a market player. From this, Section 0 presents the real-world implementation of the explicit mode at Green Energy Park demonstration site.

The second method presented Section 4 is a model-free EMS that optimizes the charging power to maximize the profits when participating in the imbalance market. This method uses a reinforcement learning technique to learn the system's dynamics directly from market interactions. Through this learning process, the method formulates an optimal policy that reacts dynamically to the state of the EVs and the implicit steering of the market to minimize costs and ensure user satisfaction.

Finally, Section 5 presents an analysis of the aggregated behaviour of EMS at the neighbourhood or transformer level. Most EMS are designed to optimize their own objectives (i.e., selfish behaviour), which they usually do effectively. However, they do not account for the impact on neighbouring EMS. This section therefore investigates the aggregated effects of such selfish EMS behaviour.

2 Model-based Energy Management System

The model-based EMS has started in earlier research projects [3] and has continued to evolve within this project to address the remaining challenges encountered, to adapt it to work with external market player and to improve the overall performances. This model-based EMS uses a Model Predictive Control (MPC) approach to optimize the charging power through forecasts of the components inside the Local Energy System (LES) such as photovoltaic generation, building load and charging demand. A LES is defined as a system comprising generation units, non-controllable loads, controllable charging points, and a grid connection. This predictive optimization contributes to minimizing charging costs while simultaneously reducing the environmental impact of local energy systems. Additional information on this method is available in [4]. Modifications for market participation

The objective of the ECOFLEX project Task 4.1 is to add an "explicit mode" to the smart charging system. Explicit mode here refers to "explicit demand-side flexibility" and is defined as "*consumers that choose to participate in energy markets and receive payments in return for the load variation offered and accepted on the market*"[6]. The smart charging system prior to the project only worked in "implicit mode", meaning in response to (a) price signal(s) or tariff(s). The smart charging system presented in the previous sections should be adapted to be able to participate in energy markets in addition to the original implicit mode. The modified version of the smart charging system is schematically presented in Figure 1.

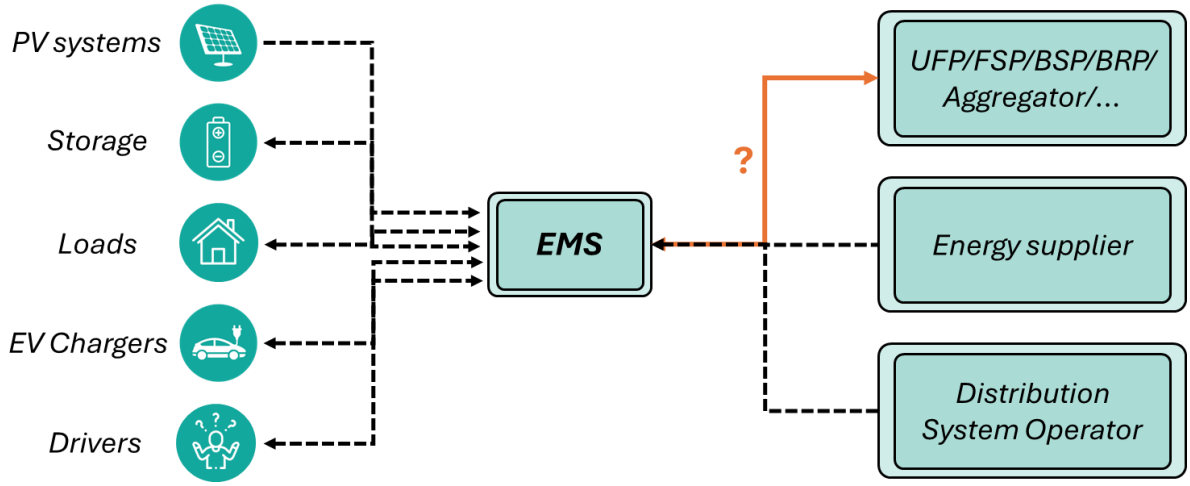

Figure 1**:** Scheme of interaction EMS and other market player

The smart charging system, shown as "EMS – Energy Management System", originally works in implicit mode by controlling its assets in order to optimize the interaction with the energy supplier (e.g. dynamic tariff) and the distribution system operator (e.g. capacity tariff). In the context of Task 4.1, the objective is to add an explicit mode by developing a new interaction with a market player (shown with orange arrow in Figure 1. In this task a market player can be a Flexibility Service Provider (FSP), a Balancing Responsible Party (BRP), an aggregator, or the Universal Flexible Platform (UFP) defined in this project. This section details the developments done to develop the interaction between the EMS and an external market player.

2.1.1 Requirements for an interaction with a market player

Two requirements are needed to establish an interaction between a market player and the EMS:

1) An incentive from the market player to encourage the EMS to adjust its control in a certain direction (consume more or consume less). This incentive is in the form of a price signal (e.g. euros/kWh).
2) A proof from the EMS to the market player that a control adjustment has been made. The proof shall be provided by submitting the expected grid power exchange (i.e. baseline) to the market player before the EMS receives the incentive from the market player.

For the second point, a sequence of events must be implemented to orchestrate the exchange of information between the EMS and the market player. In this project, in collaboration with Flexcity, the following chain of events—illustrated in Figure 2—has been established.

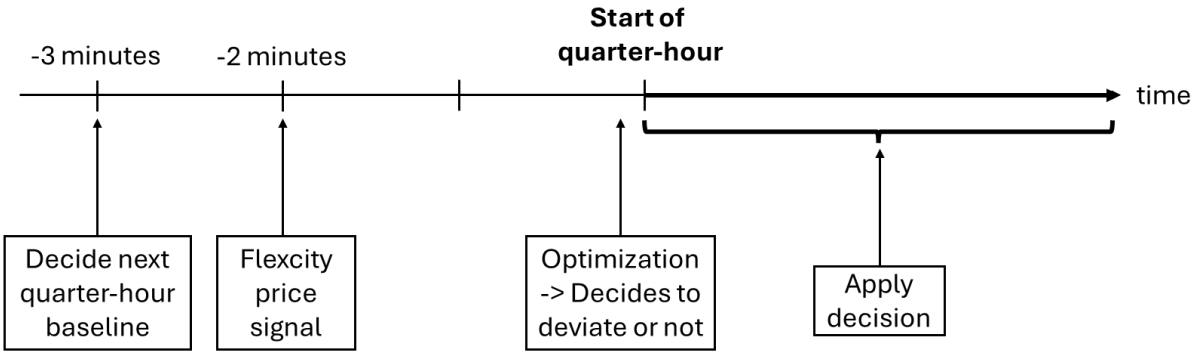

Figure 2: Chain of events between EMS and market player

The orchestration runs every quarter-hour and follows these steps:

1. **Three minutes before the quarter-hour:** an optimization is performed without the market player's price signal. This serves to calculate the baseline, unaffected by any external price incentives.
2. **Two minutes before the quarter-hour:** the market player sends a price signal indicating a desired positive or negative deviation.
3. **Right before the quarter-hour:** a new optimization is performed using the received price signal. The optimizer then determines whether deviating from the original baseline (calculated three minutes earlier) is beneficial or not.

2.1.2 Optimizer trade-off calculation

The final step in the sequence requires the optimizer to decide whether to deviate from the baseline. This decision involves assessing the immediate benefits of deviating now against the potential impact of having to compensate later. In other words, every kWh increase or decrease in the upcoming quarter-hour results in a corresponding decrease or increase in one or more of the subsequent quarter-hours (e.g., charging 1 kWh more now implies charging 1 kWh less later). To enable the optimizer to make this trade-off, a new component—called the *flex-based tariff*—is added to the objective function. This component, combined with the other parts of the function, allows the optimizer to quantify the gains and losses associated with any deviation from the baseline. The updated objective function is shown in Figure 3.

$$
\begin{aligned}
min\Bigg\{ & \\
& \sum_{t=1}^{T} C_t^{export} \times P_t^{grid,export} + C_t^{import} \times P_t^{grid,import} && \longrightarrow \text{Energy-based tariffs minimization} \\
& + \left(P_t^{import,max} - P^{import,monthly} \right) \times C_{peak} && \longrightarrow \text{Power-based tariffs minimization} \\
& - \left(P^{baseload} - (P_{t=1}^{grid,export} + P_{t=1}^{grid,import}) \right) \times C_{FSP} && \longrightarrow \text{Flex-based tariffs maximization} \\
& + \sum_{k=1}^{K_{AC}} \left(\frac{E_k^{target} - E_{t=T,k}}{E_k^{target}} \right) \times C_1 && \longrightarrow \text{Force to satisfy energy demand of drivers} \\
& + \sum_{k=1}^{K} \sum_{t=1}^{T} P_{t,k}^{ev} \times t \times w_1 && \longrightarrow \text{Charge as fast as possible} \\
\Bigg\} &
\end{aligned}
$$

Figure 3: New objective function that includes external market signals

3 Practical Implementations (link to WP6 / T6.1)

To enable testing and performance evaluation under realistic operating conditions, the algorithm as presented in Section 2 was implemented in a legacy EMS of VUB, deployed in a real-world environment The general overview of the smart charging system is presented in Figure 4.

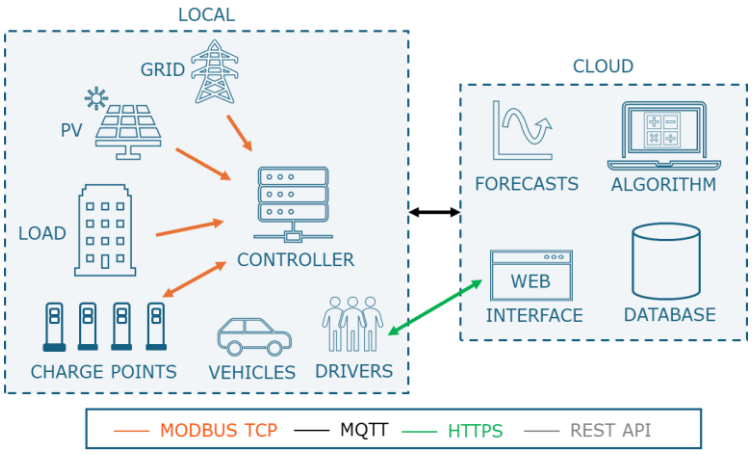

Figure 4: General scheme of the working principle of the smart charging system

The control of the charge points happens at two different levels, one at a cloud level followed by one at a local level. Each level is presented in the following subsections.

3.1 Local control level description

The local level consists of a controller that communicates with all the assets onsite and with the cloud. The controller has mainly three purposes:

1) The controller acts as a **gateway** between the assets onsite and the algorithm running in the cloud level. It transfers all state information from the assets towards the cloud (such as the current, the voltage, the power, etc), but also power setpoints from the cloud optimization towards the controllable assets such as the charge points.
2) The controller acts as a **dynamic load balancing** device which prevents overcurrent by balancing the charge points at all times. The balancing of the charge points is imperative because the algorithm running in the cloud is not able to react to second-time events such as a sudden drop in PV production, which could potentially overload the local energy system (LES) power constraints.
3) The controller acts as a **fallback position** in case of a communication issue with the cloud level. If a problem happens between the local and cloud levels, the controller takes over the control of the assets in default dynamic load balancing mode.

3.2 Cloud control level description

The cloud level contains different services such as the smart charging algorithm with its forecasts, the EV driver web interface and the database. These services are further detailed in the following paragraphs.

3.2.1 The algorithm

The algorithm is running in a dedicated server installed with an Intel® Xeon® Silver 4210R 2.40 GHz processor and 4 Gb of RAM. An example of 100 EVs staying 8 hours takes up to 1 second to optimize. The algorithm is implemented in Python and deployed in a Docker Container. The day-ahead forecast of the PV production is given by ELIA, the Transmission System Operator (TSO) in Belgium [5], and is updated every day at 5.40 pm. The algorithm is periodically triggered (every quarter-hour), and also event triggered (EV plugs in or plugs out).

3.2.2 The web interface

The algorithm uses information from the EV user on their charging needs in order to achieve better optimization results. Therefore, an interface is required between the driver and the algorithm in order to collect two important inputs:

- The expected departure time of the EV user,
- The energy need of the EV user.

Based on the previous research projects, and on the ECOFLEX project, two different interfaces can be used:

- From previous research projects, a temporary web interface has been developed, that can be accessed via a QR-code on the charge point. This temporary interface allowed to develop and commission the whole smart charging system. This interface is deployed at GEP demonstration site, and which can be accessed here: https://evscheduler.evergi.be/gep/
- From the ECOFLEX project, a new app interface has been developed by Pluginvest. Additional information is available in deliverable D4.2. This interface is deployed at Vanbreda demonstration site.

3.2.3 Communication Interfaces and Data Handling

All the data from the assets, accessed by the local controller, is stored in an InfluxDB database. The communication between the assets and the local controller is mainly Modbus TCP, while the communication between the local controller and the cloud environment is MQTT.
New functionalities, bug fixes and general improvements have been developed during the ECOFLEX project for smoother operation of the algorithm and to address specific real-world challenges. Enhancements have been done at both cloud and local levels. This new version of the algorithm has run for multiple months at GEP demonstration site.

3.3 Local control level changes

3.3.1 Handling loss of connection with cloud level

Significant work has been done to enhance the robustness of the smart charging algorithm running in the cloud environment. However, failures unrelated to the algorithm itself can still occur, such as server outages or loss of internet connection. To address this, a new feature has been implemented at the local level to enable operation without the cloud environment. During a loss of connection with the cloud, the local system will charge electric vehicles at maximum power in a dynamic load balancing mode to avoid exceeding the maximum grid power.

3.3.2 Enhancements of the local level algorithm

A specific functionality is required at the local level that computes the maximum charging power of the connected electric vehicles since such information is not exchanged between the charge point and the electric vehicle. From this, the maximum charging power is unknown at the charge point, hence unknown at the controller level and cloud level. Since the smart charging algorithm running in the cloud has to know the maximum power in order to optimize the charging process, a specific functionality has been implemented at the controller level which consists of charging the electric vehicle at maximum power at the very beginning of the charging session. Once the maximum power is known, the charging can resume its normal smart charging activities.

Some improvements were made during the ECOFLEX project in order to enhance this functionality :

- A first improvement was made in order to continue the dynamic load balancing of the other connected electric vehicles during the process of calculating the maximum power of one single electric vehicle.
- A second improvement was made to better calculate the maximum power of an electric vehicle. The reason is that such process is not trivial because it is a trade-off between
    - The time it takes to go to the maximum power which is different per brand, model, external temperatures, SOC of the battery, and other factors,
    - The time of such process should be that should be minimized

3.4 Cloud control level changes

3.4.1 Driver's interaction improvements

Interaction with the driver is a crucial aspect, as they are the owners of the assets being controlled. Prior research has shown that a large portion of the flexibility is lost due to inaccurate information provided to the system by the driver [4]. Therefore, a significant effort has been dedicated to improving the driver's experience to enhance the performance of the smart charging algorithm.

As first major novelty has been introduced which allows drivers to charge without using the web interface or app. This works by using a default charging profile: a default parking duration and a default energy need. This functionality is designed for frequent users of a charging facility who follow a typical usage pattern and hence do not need to scan a QR code or open an app every day. For example, at the GEP demonstration site, many drivers exhibit a typical employee charging behaviour: arriving in the morning and staying until the end of the working day. Given this consistent parking duration and an assumed energy need of 40 kWh, these drivers do not need to interact with the smart charging system. Several key points can be highlighted:

- The 40 kWh default value is based on historical charging session data and ensures that at least 75% (3rd quartile) of sessions are fully covered by this amount.
- The default parking time can be set as either a parking duration or a departure time. At the GEP site, using the departure time is more appropriate, as employees typically finish work around 5 p.m.

A second novelty has been implemented to allow a re-optimization of the charging profiles when a new charging preference is entered by a driver. This enables drivers to adapt their charging needs in response to unexpected events that may require a change in their plans.

A third major novelty has been introduced to enhance the driver's experience by designing a new friendlier web interface at the GEP demonstration site. A consultation with the drivers made it clear that the web interface was not well designed to be understandable without explicit explanations on how to use it. The difference between the old and new web interface is shown in Figure 5.

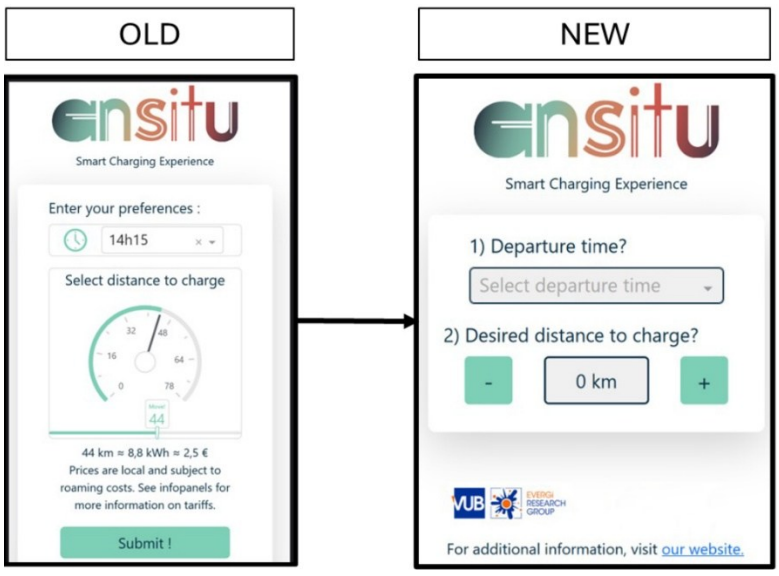

Figure 5: Design change of the web interface

The biggest changes in design are the following :

- The new design makes it clearer, compared to the old design, that there are two inputs needed by the driver, by showing two steps ("*1)*" and "*2)*"): First enter the departure time and then enter the distance to charge.
- The old design uses a slider, combined with a gauge, to change the energy demand. The use of two modules (slider and gauge) makes it not clear where to change the distance to charge. In addition, the slider makes it too easy for the driver to ask the maximum energy demand (simply slide to the right). This strongly reduces the flexibility of the driver towards smart charging. Consequently, the new design requires to click on "+" r "-" to add or reduce by steps of 20km.
- Reduce the amount of text to display to allow the driver to focus on the two inputs (steps).

Finally, a fourth change has been made to integrate the Pluginvest application with the smart charging system. Additional information is provided in deliverable "*D4.2 – Front end applications for charge scheduling*".

3.4.2 Load demand forecasts

The advanced smart charging algorithm needs a forecast of the uncontrolled load demand (e.g. office building consumption) to optimize the charging of electric vehicles. Such forecast is done using a specific neural network developed in [3] and which is trained once per week. However, a significant drawback of using such a neural network is the requirement for extensive historical data on the load demand. In [3], more than one year of historical data is used to train the neural network, which in practice, on a new site like Vanbreda demonstration site, is not available. In such a situation, a different method is required that works with limited historical data such as one week or one month length. There is a significant amount of forecasting methods available in literature, from simple to very complex ones. Authors in [17] presents a comparison between many different methods, and do show that a simple "naïve" forecast can have a similar performance to highly complex methods. The naïve forecast method predicts the last observed values assuming that the next behaviour is the same as the previous observed behaviour. The biggest advantage of such method is that it does not require extensive historical datasets, which is the initial problem mentioned previously with the training of neural networks. A naïve forecast method has been implemented in the smart charging system, with its working principle depending on the length of the available historical dataset.

- If the historical dataset is less than 24 hours long, the naïve forecast uses the last measured value to predict a constant value for the next 24 hours.
- If the historical dataset is less than one weeklong, the naïve forecast uses the previous day's measurements to forecast the next 24 hours.

- If the historical dataset is less than one month long, the naïve forecast uses the previous week's measurements to forecast the next 24 hours. For example, if a forecast is required for Tuesday, the measurements from the previous Tuesday are used.
- If the historical dataset is longer than one month, the naïve forecast is no longer used, and the neural network is employed instead.

3.4.3 Optimization problem adjustments

Two new constraints are introduced in the optimization problem which are a) the grid current limitation per phase, and b) the power limitation per EVSE in a double EVSE charge point.

The original algorithm equations use power variables (in kW) instead of current/voltage variables to simplify the problem. Consequently, the grid power limitation is based on power rather than on the current per phase. This approach assumes that the local grid is balanced, with equal current per phase. However, in practice, this assumption does not hold for single-phase charging sessions. Therefore, additional constraints are added to limit the current per phase. Note that it is still assumed that uncontrolled production and consumption assets are balanced, and only the charge points can be unbalanced (single- vs three-phase charging sessions).

The original algorithm does not account for an important power limitation that applies to double EVSE (Electric Vehicle Supply Equipment) charge points. Each individual EVSE can provide up to 32A three-phase (22kW) charging power. However, the combined power of the two EVSEs should not exceed 32A three-phase (22kW). In other words, while one EV can charge at up to 22kW on a single EVSE, if two EVs are charging simultaneously, each will be limited to 11kW to ensure the total power does not exceed 22kW. Such type of chargers are more or more present in EV parking lots and are also present in the Vanbreda demonstration site (Alfen). Therefore, a new constraint is added in the optimization problem that limits the power accordingly.

3.5 Experimental Results

The experimental set-up where the smart charging algorithm is tested is located in the Green Energy Park in Flanders, Belgium. It contains an open-access parking lot with multiple charge points. The open accessibility of the parking lot allows to have random EV users that do not have prior knowledge on smart charging technology, contributing to a more realistic experiment. An aerial view of the site is shown in Figure 6.

Figure 6: A real picture of the living lab Green Energy Park

The electricity prices used in the experiment are composed of three main components:

- Energy-based price: BELPEX day-ahead price
- Power-based price: Capacity price
- Flex-based price from the external market player

More specifically, the LES consists of a grid connection limited to 44 kW. It contains a building consumer, a 6.4 kWh PV modules East-West oriented with a 4kW SMA inverter, and four 32A ABB TACW22 charge point, two single-phase 32A ABB TACW7-4 charge points, and two ALFEN Eve Double Pro-line charge points. This configuration is illustrated in Figure 7.

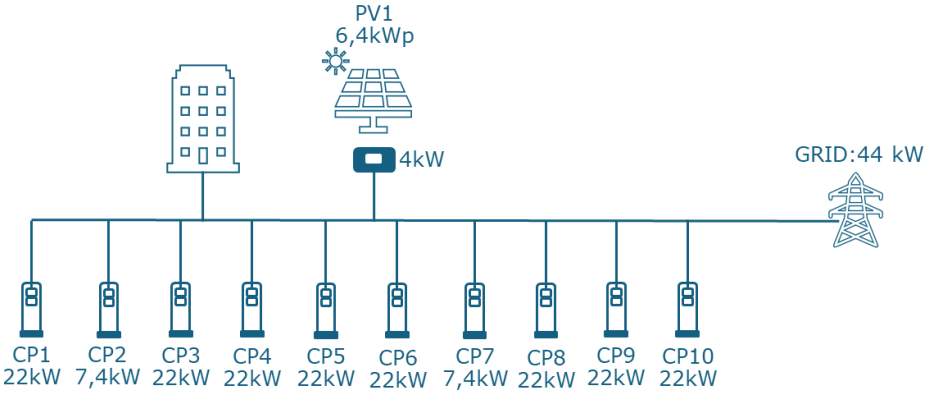

Figure 7: Electrical scheme of the experiment

3.5.1 Experiment Data Analysis

The experiment took place from the 4th of April 2025 to the 4th of July 2025, which corresponds to 3 months of data. The data collected in the system was pre-processed to deal with problems like network issues, server issues, newly implemented unstable codes, and more. Therefore, any problem identified at a certain time led to straightforward deletion of the day. From this, 11 days have been deleted, leading to a total number of acceptable experiment days of 80.

To gain a better understanding of how the assets of the LES behave in the experimental set-up, Figure 8 shows the daily energy balance of the LES by presenting the daily energy variability per asset.

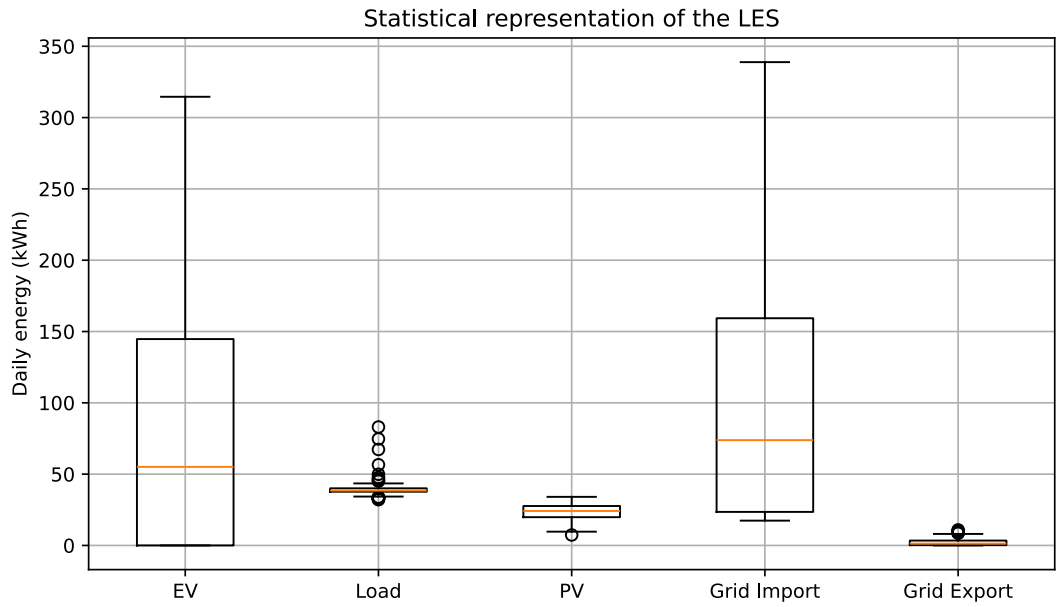

Figure 8: Daily energy variability analysis per asset in the LES

From Figure 8, it can be observed that the load demand and PV production show low variability in daily energy consumption compared to the charge points, with interquartile ranges of 12.9 kWh and 7.5 kWh for the load and PV, respectively, versus 144 kWh for the EVs. It is also clear that the energy demand of the EVs does exceed both the load demand and the PV production. This is reflected in the fact that the grid export is close to zero. These observations suggest that the PV installation is undersized, which limits its impact on the overall energy system.

3.5.2 Real-world results: example

To illustrate the approach, a concrete results of 1 day behaviour is presented in Figure 9. The day includes five charging sessions; 2 sessions at 11 kW, 1 session at 7 kW, and 2 sessions at 3.7 kW, which are non-flexible (i.e., required to charge at maximum power and cannot be shifted). The example is shown in Figure 9 which consists of two subplots:

- The upper plot shows the grid power exchange, comparing the baseline power (expected consumption) with the measured power. Deviations between the two are highlighted using a colour code:
    - Green indicates a beneficial (correct) deviation, meaning a deviation from the baseline that lead to an overall cost reduction.

- Red indicates a costly (incorrect) deviation, meaning a deviation which leads to an overall cost increase. This can happen when there is unforeseen behaviour in the system or when imbalance price forecasts deviate from the validated price.
- Grey indicates a request for deviation is ignored because the opportunity cost is too high.

- The lower plot displays the imbalance prices, including both the forecasted price provided by Flexcity and the validated imbalance price from ELIA. A grey-shaded zone marks the price intervals that are considered too uncertain or risky to act upon; within this range, no flexibility actions are taken.

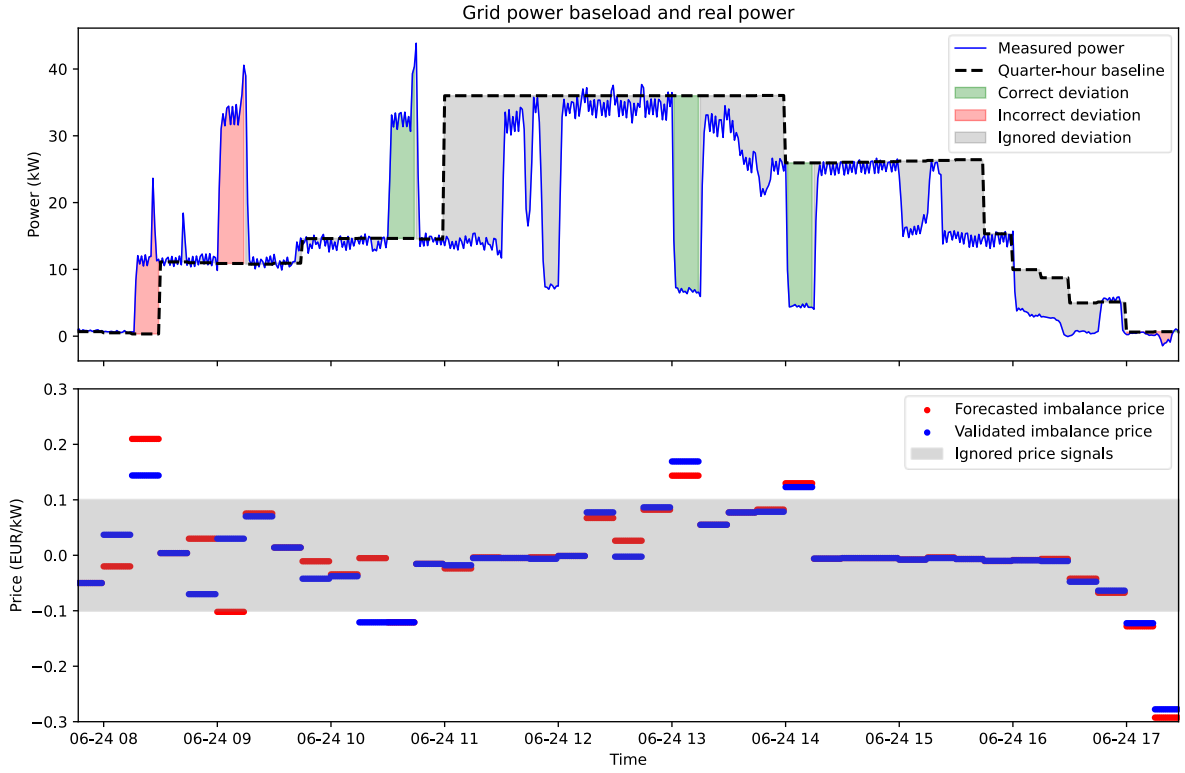

Figure 9: Performance example of one day scheduling of 5 charging sessions

In this example, five deviations from the baseline occurred. The first deviation was not beneficial, caused by an unexpected electric vehicle plugging in with a non-flexible charging session, requiring charging at maximum power and leaving no room for adjustment. The second deviation was also not cost-beneficial and resulted from a forecasting error: while the forecasted imbalance price was -101.3 €/MWh (suggesting a profitable upward deviation), the validated imbalance price turned out to be +30.9 €/MWh, leading to a financial loss. The remaining three deviations were correct, as indicated in green on the graph. These include one upward deviation of +17 kW over 15 minutes, and two downward deviations of -30 kW and -21 kW, each also lasting 15 minutes.

3.5.3 Real-world results: overall performance

Table 1 breaks down the financial outcomes of the charging strategy, detailing both gains and losses. The losses can arise from deviations between scheduled and actual energy consumption, which can be triggered by events such as inaccurate imbalance price forecasts or dynamic changes in the charging load, including the unexpected arrival or early departure of an EV.

Table 1: Gains, losses and total balance of deviations

Gains/losses on deviations	Using forecast imbalance price	Using real validated imbalance price
Gains	€108.03	€102.39
Losses	€-52.39	€-50.35
Total balance	€55.64	€52.03

The results show that using a perfect forecast or the actual imbalance price leads to very similar outcomes, with total net gains of approximately €50 over a three-month period. This small difference suggests that the price forecast used is either already highly accurate or that the applied strategy is relatively conservative, participation occurs only when the forecasted imbalance price falls outside a moderate range (between -100 and +100 €/MWh). Additionally, gains are roughly double the losses, indicating that even with some uncertainty, the strategy yields a positive net result. This gain reduces the original electricity bill (energy-based and power-based tariffs) of 1108 euros by 4.7%.

Still, the results should be interpreted with caution. The observed gains and losses are closely tied to the specific use case, the chosen thresholds for price signals, and the flexibility available in the system. Moreover, to fully isolate the impact of reacting to price signals, a comparative simulation without external price inputs – but with smart charging logic – should be performed. This would allow for a more rigorous assessment of the added value of market-driven energy management.

3.5.4 Future work

Two main improvements are identified for future iterations of this approach. First, the acceptable price signal range should be further optimized. It is possible to adjust this range – currently set between -100 €/MWh and +100 €/MWh – to improve the trade-off between better or worse financial outcomes versus higher or lower risk. Second, the EMS should be configured to strictly follow the baseline even when price signals are ignored. In real market conditions, it is likely that market actors expect each EMS to track its forecasted baseline, regardless of whether or not it reacts to imbalance prices.

4 Model-free Energy Management System

This section of the deliverable was presented at the 38th International Electric Vehicle Symposium and Exhibition (EVS38) [7].

4.1 Introduction

Unlike traditional model-based approaches that depend heavily on accurate system modelling and prediction of future market prices and user behaviours, often challenging given the nonlinear, nonconvex system dynamics and volatile market conditions, reinforcement learning (RL) can learn optimal control policies directly from real-time data and interactions with the environment. RL is particularly well-suited for managing the integration of EVs and battery energy storage systems (BESS) in ancillary service markets due to the high uncertainty and dynamic nature of these environments. Model-free characteristics allow RL to adapt to sudden changes in imbalance market dynamics, and unpredictable EV behaviours such as variable

arrival/departure times and initial state of charge, without relying on explicit forecasting or scenario generation. Consequently, RL offers a flexible, data-driven approach that can continuously improve decision-making in complex, stochastic settings, overcoming the limitations of conventional optimization and model predictive control methods.

4.1.1 Model-free reinforcement learning

Model-free RL methods do not require prior knowledge or an explicit model of the environment. Instead, by interacting directly with the environment, the agent learns to adapt itself to the system dynamics and accommodate for inherent uncertainties through experience to maximize cumulative reward [8]. This method simplicity is especially useful when the environment is complex, stochastic, or partly observable since it does not require an understanding of how actions affect the environment. Although it comes at the cost of reduced sample efficiency, meaning the agent may need more interactions to learn an effective policy. Model-free RL is depicted in **Error! Reference source not found.**. In contrast, model-based RL simulates future states and rewards using a model of the environment's dynamics and reward function, allowing the agent to plan and make decisions without solely relying on direct interaction experience.

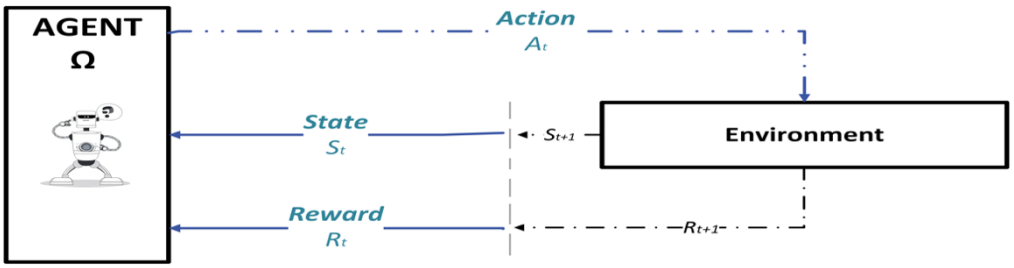

Figure 10: Model-Free reinforcement learning diagram

Some popular model-free RL methods are Q-learning, Deep Q-Networks, policy gradient methods, and Actor-Critic methods. Proximal Policy Optimization (PPO), as a member of policy gradient category, is chosen among these due to its balance of stability, sample efficiency, and ease of implementation, making it suitable for a wide range of applications [9]. The policy gradient method's purpose is to directly optimize the policy, which is the strategy the agent uses to decide what action to take in each state by adjusting the parameters to maximize the anticipated cumulative reward. The set of interactions of an agent with its environment (actions, states, rewards) from a starting state to an ending state is called an episode. PPO is developed to address the instability and high variance inherent in earlier policy gradient methods, which frequently resulted in unstable learning processes and suboptimal performance. By incorporating a clipping mechanism within the objective function of the new policy, PPO ensures that policy updates are significant enough to yield meaningful improvements while remaining sufficiently constrained to prevent instability. In PPO, deep neural networks are utilized as function approximators to represent both the policy and the value function, allowing the algorithm to efficiently navigate complex environments and large state spaces.

4.1.2 Markov Decision Process Formulation for EVs Smart Charging

The EV smart charging can be formulated as Markov Decision Process (MDP) which provides a mathematical structure for stochastic sequential decision-making problems. The MDP is formally defined as a 5-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma)$. Here $\mathcal{S}$ represents the state space, $\mathcal{A}$ denotes the discrete action space, $\mathcal{R}: \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ indicates the immediate reward function, $\mathcal{P}: \mathcal{S} \times \mathcal{S} \times \mathcal{A} \rightarrow [0,1]$ signifies the unknown state transition probability distribution. The discount factor $\gamma \in (0,1]$ indicates the importance placed on future rewards relative to immediate ones [10]. At each time step t, the agent observes the current environment state $s_t \in \mathcal{S}$ and selects an action $a_t \in \mathcal{A}$ in response. As a result of this action, the agent receives a reward value $\mathcal{R}(s_t, a_t)$ and transitions to a new state $s_{t+1} \in \mathcal{S}$ according to the state transition probability distribution $\mathcal{P}(s_{t+1} \mid s_t, a_t)$. The MDP formulation for EV smart charging within the imbalance market can be structured as follows:

State:

In this study, these selected episode for the RL is a randomly selected specific day and charger. Once training is completed, the developed policy is implemented across all chargers, enabling efficient management of the EV charging process based on the learned strategies. At each timestep, the observation space, referring to the set of possible inputs or states the agent can perceive from its environment, is defined as follows:

$$S_t = \left(T_t, T^{dep}, SOC_t, \hat{\pi}_t^{imb}\right) \quad (1)$$

Where T_t is the timestep on that day, T^{dep} represents the specific timestep during the day when the EV, is scheduled to depart. SOC_t is the SOC of the EV at the timestep *t*, and the $\hat{\pi}_t^{imb}$ is the forecasted imbalance price of the system. A forecast of the imbalance price is used because TSO calculates it at the end of each quarter-hour based on the activated balancing resources. It is assumed that EV users provide their SOC^{target} and departure time (T^{dep}) when they connect their car to the charger, and that this information is known in advance.

Action

It requires defining appropriate actions that the agent can select during its interaction with the environment. For this study, a discrete action space is considered, containing three possible actions, as detailed below:

$$a_t \in A, \quad A = \{0, \frac{1}{2} \times P_{max}^{ch}, P_{max}^{ch}\} \quad (2)$$

Where P_{max}^{ch} is the maximum rate of the EV charger in kW. Thus, a_t represents the action chosen by the agent, which can be idle, charge at half rate, or charge at the maximum rate. While the two actions (0 and P_{max}^{ch}) are straightforward—indicating whether the EV is charging with maximum rate or not—introducing a third action ($\frac{1}{2} \times P_{max}^{ch}$,) allows for more nuanced control over the charging process. Selecting the half-rate charging option allow the agent to charge during periods of slightly elevated electricity prices without completely forgoing the charging process.

Reward function

The reward function is depicted in equation (3). It consists of a cost-based term and a term to capture the degree to which the charge is successfully completed. The complexity of the

reward function (Equations 3-10) is a direct consequence of the multi-objective nature of the EV charging problem in the context of imbalance market participation, where the primary objective is to maximize charging revenues (minimize charging cost), while trying to fulfil the charging needs of the driver, meaning the EV reaches its target SOC by the departure time set by the driver. Simple reward functions, such as those based solely on imbalance price cannot adequately capture this essential trade-off. Therefore, a well-designed reward system with a clear reward-penalty structure is implemented to promote desirable behaviours, such as achieving the target SOC and minimizing expenses, while also discouraging adverse outcomes like excessive costs or insufficient energy delivered to EVs. This is particularly important, given the complexities of EV smart charging in the imbalance market, which include stochastic imbalance prices, varying EV arrival and departure times, and differing initial SOCs.

The reward at each timestep (3), comprises two components: the stepwise cost and energy delivery-related term, reflecting the agent's need to align with the EV driver's charging preferences. The weighted factors α and β are employed to adjust the influence of each respective term.

$$r_t = \alpha \times Cost_t + \beta \times Energy\_delivery\_factor_t \quad (3)$$

$$Cost_t = -\frac{a_t}{P_{max}^{ch}} \times \frac{\hat{\pi}_t^{imb}}{\hat{\pi}_{max}^{imb}} \times b_t \qquad \in [-1, 1] \quad (4)$$

The cost formula in equation (4) is the chosen action a_t by the agent at time t, normalized by the maximum charging power P_{max}^{ch}, multiplied with the imbalance price $\hat{\pi}_t^{imb}$, normalized by the maximum imbalance price ($\hat{\pi}_{max}^{\text{imb}}$) observed in the dataset, and b_t, which is a binary variable that takes the value of 1 when the EV is connected to the charger and 0 when it is not connected. is Normalization is done to ensure $Cost_t$ ranges from [-1, 1]. The imbalance price can be either negative or positive: a negative price means money is received from the TSO, while a positive price indicates a payment to the TSO is required. To respect this convention, a negative sign is added to the cost formulation in equation (4).

The $Energy\_delivery\_factor$ component, as outlined in Equation (5), comprises two essential terms: charging progress (6), which measures the change in the EV SOC, and the penalty component (10), which measures the deviation of the SOC (SOC_t) to the target SOC (SOC^{target}), both at each time step t and at the end of the charging session ($t = T^{\text{dep}}$). The charging progress term component quantifies the increase in the SOC of the EV by calculating the change in SOC over time, scaled by an urgency factor. The $Time\_Sensitive\_Urgency\_Factor$ in (7) is a dynamic metric that quantifies the urgency based on the remaining time until departure. It acts as a crucial gauge of urgency, facilitating decision-making in time-sensitive situations. Without considering the remaining time, the RL agent might prioritize minimizing cost too heavily in the early stages, leaving insufficient time to reach the target SOC as the departure time approaches. In the equation (7), ρ represents a baseline level of urgency, while μ indicates the sensitivity to time, defining how urgency increases as the departure time approaches. Equation (7) is normalized in (8) by dividing by $\rho + \mu$, which represents maximum possible value of (7). These values were calculated based on trial and error in this study. Also ΔSOC_t divided by ΔSOC_{max} equation (9) to normalize the SOC change. Where C^{ev} is the maximum capacity of the EV battery and η^{ch} is the efficiency of the charger in equation (9).

$$Energy\_delivery\_factor_t = \text{Charging\_Progress}_t^{norm} + Penalty_t^{norm} \quad (5)$$

$$Charging\_Progress_t = \Delta SOC_t \times Time\_Senssitive\_Urgency\_Factor \quad (6)$$

$$Time\_Senssitive\_Urgency\_Factor = \rho + \frac{\mu}{max(1,\ T^{dep} - t)} \quad (7)$$

$$Charging\_Progress_t^{norm} = \frac{\Delta SOC_t}{\Delta SOC_{max}} \times \frac{\rho + \frac{\mu}{max(1, I^{dep} - t)}}{\rho + \mu} \quad \in [0,1] \quad (8)$$

$$\Delta SOC_{max} = \frac{P_{max}^{ch} \times \eta^{ch} \times \Delta t}{C^{ev}} \quad (9)$$

$$Penalty_t = \begin{cases} -\dfrac{(SOC^{target} - SOC_t) \times b_t}{\beta \times SOC^{target}} & \text{if } t < T^{\text{dep}} \\ -\dfrac{(SOC^{\text{target}} - SOC_{T^{dep}})}{SOC^{target}} & \text{if } t = T^{\text{dep}} \text{ and } SOC_{T^{dep}} < SOC^{\text{target}} \\ 1 & \text{if } t = T^{\text{dep}} \text{ and } SOC_{T^{dep}} = SOC^{\text{target}} \end{cases} \quad (10)$$

The penalty component designed as equation (10), when the time t has not yet reached the departure timestep (T^{dep}), the penalty is calculated as the normalized deviation from the requested SOC^{target}. This time dependent signal allows the RL model to receive feedback every action step, in contrast to the other component which is a single penalty at the departure time which would not offer sufficient opportunity for the model to adjust its strategy. To avoid this component having disproportionate influence due to being calculated at every connected timestep before departure, the first condition is divided by β, the scaling factor of the full $Energy\_delivery\_factor$. Upon reaching timestep T^{dep}, the penalty factor assumes a negative value if the departure SOC is less than the level requested by the EV driver, and a positive value of 1 if the SOC satisfies the specified threshold. The penalty factor falls within the range [-1, 1].

State Transition

Within the MDP framework, system dynamics are characterized by a state transition probability function, denoted as P. In the context of the EV smart charging problem, this probability function is unknown due to the uncertainties related to each EV presence and imbalance price. Consequently, the agent seeks to estimate the state probability distribution through its interactions with the environment. Nonetheless, the state transition for the SOC_t is governed by a_t and can be explicitly represented as follows:

$$SOC_{t+1} = SOC_t + \frac{a_t \times \eta^{ch} \times \Delta t}{C^{ev}} \quad (11)$$

4.1.3 Benchmark charging methods

Uncoordinated charging method

Then uncoordinated charging method is the current charging method where EVs charge exclusively on the driver's preferences and schedule, disregarding any broader effects on the power grid or charging infrastructure. Typically, EVs begin charging at a maximum rate as soon as they connect to the charger and continue to consume energy until fully charged [11].

Model Predictive Control charging method

The model predictive control charging method effectively addresses the challenges posed by the uncoordinated charging of EVs. Unlike uncoordinated charging, MPC utilizes a proactive approach by considering energy demands and optimizing charging schedules to minimize costs while reaching the desired SOC for users. At each timestep, MPC repeatedly solves an optimization problem that considers current conditions, including grid status, energy prices, and EV charging requirements future energy demand and consumption. We use the model developed in [13] applied to our use case. This smart charging control not only meets the energy demands of EV drivers but also achieves global minimization of charging costs.

4.2 Case Study

The proposed framework is depicted in **Error! Reference source not found.**.

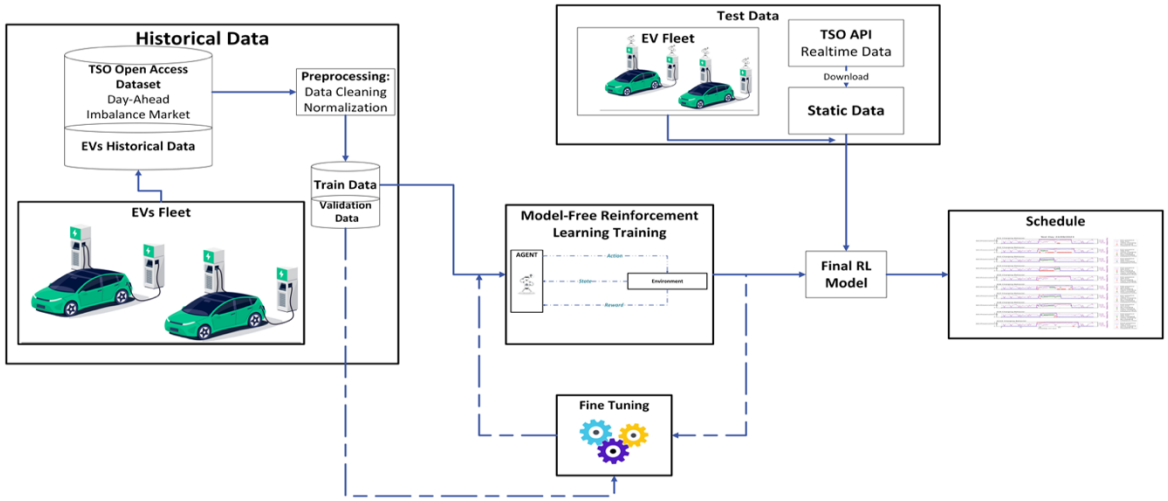

Figure 11: Proposed framework for EV smart charging in the imbalance market

The simulation consists of an office equipped with 10 EV chargers, operating under exposure to imbalance market prices. The historical data required for the imbalance and day-ahead markets was collected from the TSO, while the EV data consists of a historical dataset of charging sessions at a commercial office building in Belgium. After preprocessing, this data is divided into training and validation datasets to train and fine-tune the RL model. Finally, a test dataset is used to evaluate the final model's capabilities.

4.2.1 Dataset electricity market

In this study, the electricity data comprises day-ahead and imbalance prices sourced from the Elia open data website, which can be accessed at [5] (the historical data section in Figure 5). The year 2023 was chosen to represent current market conditions due to its recent and relevant data. Elia publishes two types of imbalance prices: the 15-minute-based price and the 1-minute-based price. The 15-minute-based price serves as the reference for imbalance settlements for BRPs. It represents the real imbalance price calculated at the end of each quarter-hour period and is subject to a validation process by TSO to ensure accuracy. In contrast, the 1-minute-based prices are derived from non-validated data. These prices reflect the instantaneous system imbalance and the cumulative activated regulation volumes on a minute-by-minute basis. They are intended to provide BRPs with additional insights into the grid's real-time status and the adjustments being made [12].

BRPs manage imbalances across their entire portfolios, and EV parking lots—whether through aggregators or direct contracts with BRPs—can actively participate in the imbalance market by offering flexibility services. There is no minimum size requirement, as it is dependent on the contractual agreement with the BRP. As depicted in Figure 12 the frequency of negative imbalance prices is particularly high during peak office parking hours (8:00–18:00), with the highest occurrence rate of 35% around noon. Most imbalance prices, in 2023, are concentrated in the pricing bin ranging from -69.95 €/MWh to 344.83 €/MWh, a range that spans approximately 414.8 €/MWh. With an appropriate strategy targeting this pricing window, there is significant potential to decrease the cost of charging. The maximum observed imbalance price during the simulation period was 1450 €/MWh, which was used to

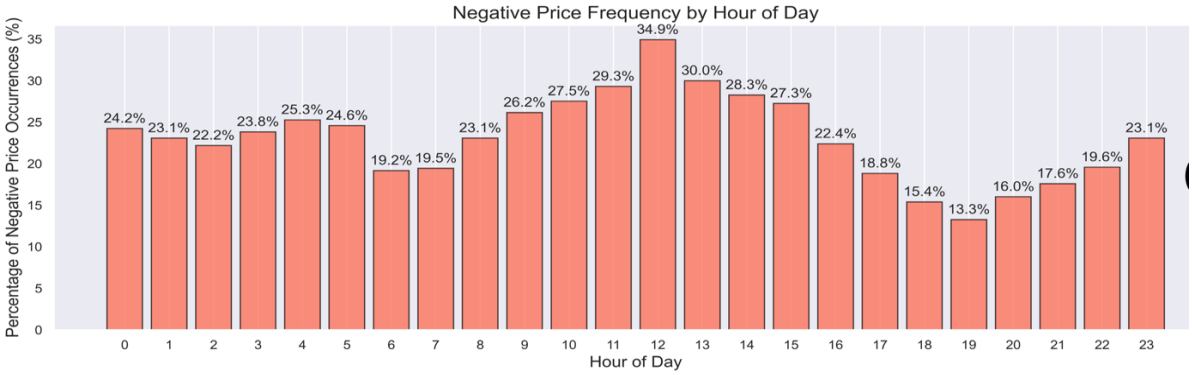

Figure 12: Belgium's market of imbalance price for 2023, Negative price frequency percentage in each hour

normalize the imbalance price data.

In electricity markets, a baseline refers to the scheduled level of energy consumption or generation, serving as a reference point to compare against actual levels and identify imbalances. In the use case, we define a zero-baseline consumption to be cleared day-ahead. By defining a zero-baseline scenario, every power of the EV parking facilities is considered a deviation from the baseline and will be exposed to the imbalance prices, and do not engage the day-ahead market. The setup allows for a focused assessment of the implications of relying exclusively on the imbalance market, giving insight into both its potential cost advantages and risks. Since the EV parking lot participates with a zero baseline and combining individual EVs does not yield additional advantages in this context, a unified smart

charging strategy has been developed and applied to all chargers. The approach is scalable as it standardizes decision-making across the entire charging infrastructure, eliminating the need for individualized control logic. This uniformity enables integration and expansion, allowing the system to easily adapt to additional chargers and growing EV fleets.

4.2.2 Dataset charging sessions

The case study focuses on an office parking lot equipped with 10 charging points. The charging sessions, defined by their arrival times, parking durations, and energy requirements, are derived from a historical dataset of charging sessions at a commercial office building in Belgium, as detailed in [13]. The simulation used data from a 90-day period between July 1st and September 30th, 2023. It was trained on the first 70 days, validated on the following 10 days, and tested on the final 10 days. The chargers are considered to have maximum power of 11 kW (P_{max}^{ch}). The charger's efficiency (η^{ch}) is considered at 90%. For simplicity, it is assumed that all EVs have a similar 70 kWh battery capacity (C^{ev}). It is also assumed that each charger can accommodate one charging session per day. The arrival and departure times, as well as the SOC, can assume any values in range of (0,1]. The SOC^{target} is set to 1, indicating that EV drivers aim to fully charge their vehicles. In this paper, after several trials and adjustments, the parameters α and β were set to 1 and 32, respectively. In equation (8) the ρ and μ are 0.25 and 2, respectively. These values adjust the magnitude of rewards and penalties at each timestep, thereby considering that the impact of each component is balanced and proportionate within the optimization framework.

4.2.3 Reinforcement Learning hyperparameter

Reinforcement learning hyperparameters play an important role in training agents because they control how quickly and effectively an agent learns optimal policies. A fine-tuning of hyperparameters can enhance performance and convergence rates, whereas a poor choice may result in slow learning or suboptimal solutions. In this study, the Optuna library [14] is utilized to identify the optimal hyperparameters for this problem. Optuna is a robust framework for hyperparameter optimization that updates the search for the best parameters using advanced optimization techniques. The result of the hyperparameters is reported in Table 2.

Table 2: PPO Hyperparameters Optimized by Optuna

Hyperparameter	Explanation	Range	Optimal Value
learning_rate	The learning rate	[$1e^{-5}$ - $1e^{-2}$]	$17e^{-5}$
n_steps	The number of steps to run for each environment per update	[192 - 672]	576
batch_size	Minibatch size	[192 - 672]	480
clip_range	Clipping parameter	[0.1 – 0.5]	0.34
gamma	Discount factor	[0.9 – 0.999]	0.985
ent_coef	controls the weight of the entropy term in the loss function, promoting	[0.1 - 0.8]	0.34

	exploration by encouraging a more diverse action distribution		
vf	Value Network hidden layer size	[32 - 256]	[32, 64]
pi	Policy Network hidden layer size	[32 - 256]	[128, 128]

The simulation is executed in a Python environment, utilizing Stable-baselines3 [15] for reinforcement learning. This framework provides a collection of reliable reinforcement learning algorithms in Python, offering a user-friendly and efficient interface for training and assessing RL agents. Given the complexity of the problem, the PPO model was trained for 20000 episodes, with each episode representing a full day. For each episode, the RL chooses a random day and a random EV on that day, then trains and gathers the required information to use as experience.

4.2.4 Key performance indicator

This subsection outlines the criteria used to evaluate the performance of charging strategies in both the day-ahead and imbalance markets. To evaluate the technical and economic performance of the configurations, as well as the primary reward equation presented in equation (3), the following key performance indicators (KPIs) are utilized:

- *Charging cost:* the total cost related to pay or receive from the market (€/kWh).
- *Energy delivery:* Defined as the percentage of the total energy supplied to EVs during charging relative to the total energy required to achieve a fully charged state, expressed as a percentage (%).

4.3 Results

4.3.1 Day-ahead market

To facilitate a comprehensive comparison, the first scenario is in the day-ahead market using two methods: uncoordinated and MPC charging. Table 3 presents the findings for the period from 21 to 30 September 2023. A total of 83 charging sessions took place during this period. The total cost for uncoordinated charging amounts to €216.60 for charging 2282.23 kWh. The MPC method achieved a price reduction to €170.4 while successfully charging all EVs, resulting in 100% energy delivery.

Table 3: Comparative results of charging scenarios

Charging Simulation	Market	kWh Charged	Total Cost (€)	Cost per kWh (€/kWh)	Percentage of Energy delivery
Uncoordinated charging	Day-Ahead	2282.26	216.76	0.094	100
MPC method	Day-Ahead	2282.26	170.43	0.074	100
Uncoordinated Charging in Imbalance price	Imbalance	2282.26	205.38	0.09	100

Knowing 15-min Imbalance price	Imbalance	2206.93	-45.84	-0.027	96.7
1-min imbalance price	Imbalance	2079.14	-29.47	-0.01	91.1

4.3.2 Uncoordinated charging in the imbalance market

This scenario is designed to assess the impact of exposing charging operations to the imbalance market price in the absence of smart charging algorithms. To ensure a fair comparison across all the scenarios, a simulation was also conducted over the same 10-day test period and reported in Table 3. As shown in Table 3, this scenario resulted in a slight reduction of 5.2% in total cost compared to uncoordinated charging in the day-ahead market. Nonetheless, it is observed that it still could not outperform the MPC method in the day-ahead market.

4.3.3 Imbalance market with knowledge of 15-minute prices

As mentioned in Section 4.1, the actual imbalance price is revealed at the end of each 15-minute interval. RL method was tested under the assumption that only the next 15-minute imbalance prices were perfectly known in advance. This idealized scenario provided a simplified environment to demonstrate the RL model's theoretical capabilities. The RL agent must develop an optimal policy without knowledge of future prices while observing the state represented by $S_t = (T_t, T^{dep}, SOC_t, \hat{\pi}_t^{imb})$.

In this scenario, each day consists of 96 timesteps (24 × 4). The simulation run on a MacBook Pro (with M3 MAX CPU and 48 GB RAM) took 190 minutes. The total cost is €-45.84, indicating that the BRP receives money from the TSO. The total energy delivery is 2206.93 kWh, resulting in an energy delivery rate of 96.7%. Among the 83 charging sessions, two EVs departed with SOC levels below 90%, registering values of 85% and 88%, respectively.

In Figure 13, the policy developed by the RL agent using PPO is displayed.

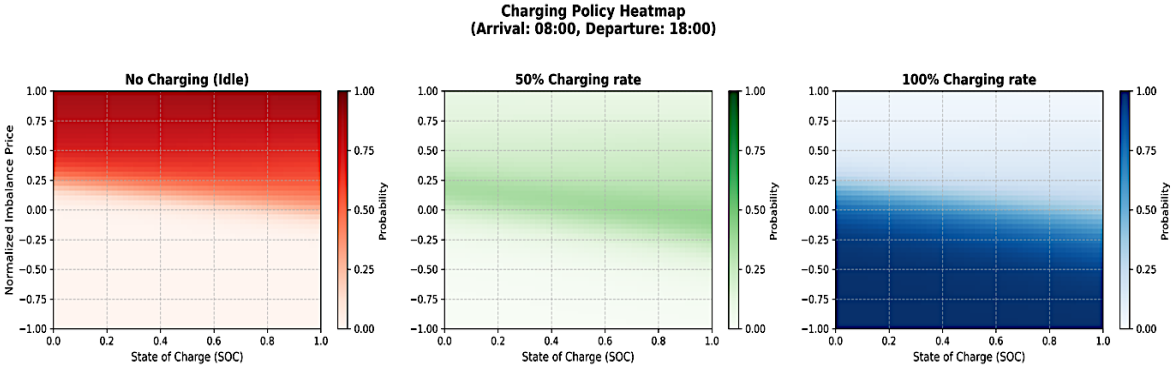

Figure 13: Policy developed with PPO by knowing 15-minute imbalance price

The heatmaps illustrate a dynamic charging policy in function of the imbalance price and SOC. To reduce the four-dimensional state space of Equation (1) to a 2-D plot based on SOC and imbalance price, the other two variables are fixed. For this plot, the arrival time is set at 8:00 AM and the departure at around 6:00 PM, reflecting typical office hours in Belgium. The probability distribution generated by the PPO algorithm (shown in Figure 7) reflects how

likely the agent is to choose each possible action in a given state. This framework allows the agent to balance exploration and exploitation by giving higher probabilities to actions that are more beneficial, while still occasionally selecting less likely actions. The probability distribution refines over time as the agent accumulates experience, providing a more optimal policy through iterative learning processes. When the training phase is complete, the learned policy is used deterministically. At each time step, based on the SOC and the normalized imbalance price while also considering the known departure time, the agent consults the policy and selects the action with the highest probability.

As it is seen from Figure 13, when the SOC is low (0% to ~30%) the model shows a preference for charging at the maximum rate (100%) even when faced with slightly positive prices (up to ~0.25). This aggressive charging behaviour rapidly increases SOC to prevent energy deficits, possibly disregarding cost considerations. As the SOC increases, the model starts to adjust its charging strategy. It tends to reduce the charging rate in response to elevated prices. The width of the no-charge charging also expands slightly during this stage, reflecting a more cautious approach. When the SOC exceeds 70%, the model shows a notable tendency to defer charging, even amid low positive or negative prices. This suggests a strategic wait for further price reductions before engaging in charging at the 100% charging rate. The model shifts from an aggressive, cost-neutral approach at low SOC to a balanced strategy at intermediate SOC, and finally to a conservative, price-conscious position at high SOC.

Figure 14 illustrates the charging progress for a representative day, where the RL agent applies the policy outlined in Figure 13 to make state-dependent decisions for a sample state. The left vertical axis of this plot shows SOC values in yellow and the EV presence with a green line. The right vertical axis displays the normalized imbalance price, which is depicted in purple.

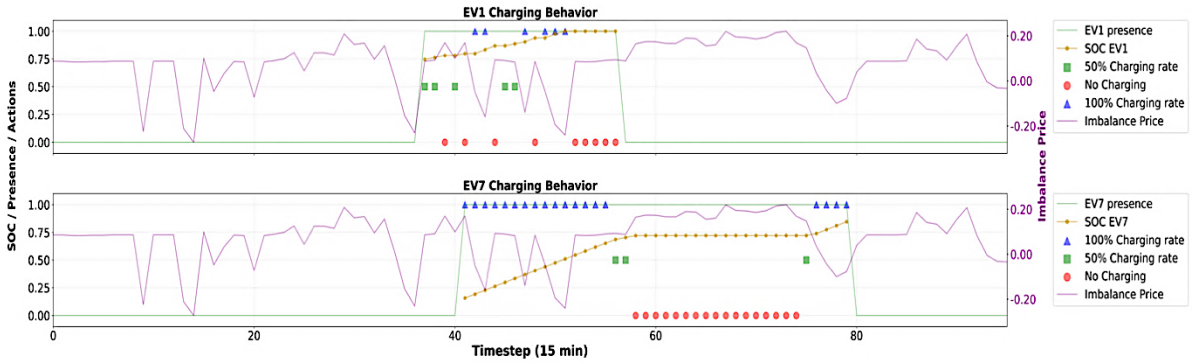

Figure 14: Charging progress of sample day (23/September/2023)

For instance, EV1, which connects to the charger with a moderately high SOC (75%), employs a cautious charging strategy by initially selecting a 50% charging rate. This approach enables EV1 to mitigate exposure to periods of marginally elevated imbalance prices, strategically deferring charging until more advantageous pricing conditions arise. Conversely, for EV7, which connects with a low SOC (21%), the agent adopts a full charging rate despite a moderately high positive imbalance price to promptly elevate the SOC to an intermediate level; when more favourable pricing conditions arise later, additional charging is implemented, resulting in a final SOC of 95%, which is deemed acceptable. The RL agent does not guarantee a globally optimal result; rather, it iteratively approximates the best possible pricing outcome by exploiting available observations and experiences.

4.3.4 Imbalance market with knowledge of 1-minute prices

Given that having prior knowledge of the exact next 15-minute imbalance price is unlikely, this section assumes that the 1-minute data discussed in Section 4.1 can be used as a prediction to participate in the imbalance market. These 1-minute prices offer an approximation of market conditions, allowing the model to account for real-world conditions. In this section, the RL agent makes decisions at one-minute intervals, consistent with the one-minute granularity of the forecasted imbalance prices. However, the real total cost is calculated using the actual imbalance price, which is validated at the end of 15-minute intervals with the TSO. Due to the increased number of timesteps, this scenario training took 280 minutes.

Here, the final policy is similar to the 15-minute policy. It is depicted in Figure 15**Error!**

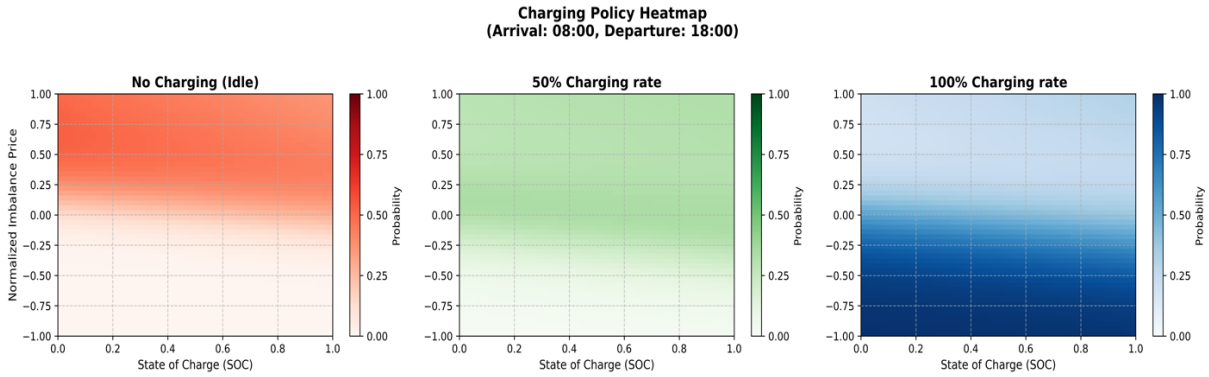

Figure 15: Policy developed with PPO for a 1-minute imbalance price

Reference source not found..

At low SOC levels (0%–30%), the agent predominantly selects a 100% charging rate, displaying insensitivity to moderately positive normalized imbalance prices (up to ~0.25) to quickly accumulate energy. In the intermediate SOC range (30%–70%), the likelihood of using the full charging rate declines under higher prices, with a shift toward either a 50% rate or no charging, reflecting growing cost awareness. At high SOC levels (above 70%), the agent mainly opts for no charging, even when prices are neutral or slightly negative, indicating a strategy to take advantage of potentially lower future prices.

Figure 16 illustrates that for most of the time, the minute-based predictions (shown in black) closely track the actual imbalance prices (shown in purple) measured at the end of each 15-minute interval.

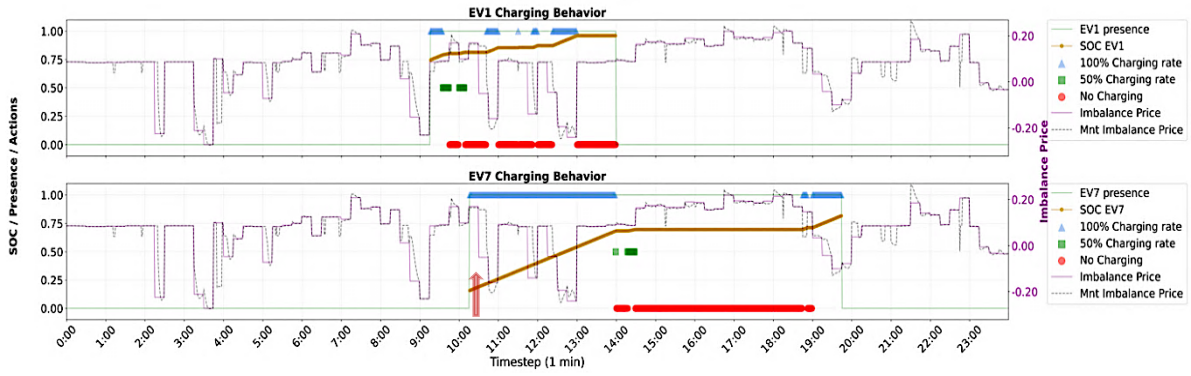

Figure 16: Charging progress of sample day (23/September/2023) in minute granularity

However, discrepancies do occur. For instance, between 10:00 and 11:00 AM, the predicted imbalance price was approximately 0.2 (normalized), while the actual imbalance price was negative. In this scenario, the agent refrained from charging to avoid high costs, thereby forfeiting an appropriate charging opportunity. As reported in Table 2, compared to the previous scenario, this approach resulted in an approximate 35% increase in cost. Nonetheless, it continues to exhibit clear advantages relative to uncoordinated charging or exclusive participation in the day-ahead market. Among the 83 EV charging sessions, nine vehicles departed with a state of charge (SOC) below 90%, ranging from 77% to 89%. To achieve a full charge, there's a trade-off between penalty and cost optimization—a higher penalty can boost energy delivery but also increase overall costs. For example, when β was raised from 32 to 50 in a simulation, the total cost increased by 28.3%, while energy delivery increased by 2%.

4.4 Discussion

The results demonstrate the advantages of smart charging strategies in the day-ahead and imbalance markets. The MPC method in the day-ahead market reduced costs by about 22% compared to uncoordinated charging, highlighting its effectiveness in optimizing expenses without compromising energy delivery. In the imbalance market, uncoordinated charging yielded a total cost of €205.38, a decrease of 5.2% compared to the day-ahead market, but it was still outperformed by the MPC method. Furthermore, the scheduling charging with the RL using the 15-minute imbalance price data resulted in enhanced financial performance with a cost of €-45.84 and corresponding to an energy delivery rate of 96.7%. The RL based on using 1-minute predictions of the imbalance price resulted in €-29.47 and the delivery rate of 91.1. Compared to the 15-minute imbalance price scenario, this is an increase in costs of approximately 35%, and a reduction of energy delivery by 5.8%. These findings indicate that while costs increased in the scenario utilizing 1-minute predictions, this approach, which reflects more closely the real market conditions, still yields good results and thus provides a practical, usable framework. Despite the increased costs, this method still resulted in lower expenses compared to the uncoordinated charging and day-ahead market approaches.

5 Large-scale Impact of Energy Management Systems

5.1 Introduction

The objective of this section is to assess the impact, in a large-scale system context, of aggregating different charge scheduling algorithms and energy management systems. Specifically, the home energy management system (HEMS) developed and evaluated in Deliverable 5.2, the smart charging system described in Section 2, and the reinforcement learning algorithm described in Section 4 will be simulated together to evaluate their combined impact at a medium-voltage/low-voltage (MV/LV) transformer. The end goal is to assess not only the benefits of EMS at the MV/LV level, but also the potential drawbacks of selfish EMSs which could induce unstable behaviors at the aggregated level.

5.2 Case Study

To address the objective mentioned above, a case study is built in which the three types of EMS are simulated together across multiple sites. According to [16], in Belgium, an MV/LV transformer typically serves between 20 and 114 consumers, with a power limit ranging from 160 kVA to 400 kVA. Based on this, the use case to be simulated is shown in Figure 17.

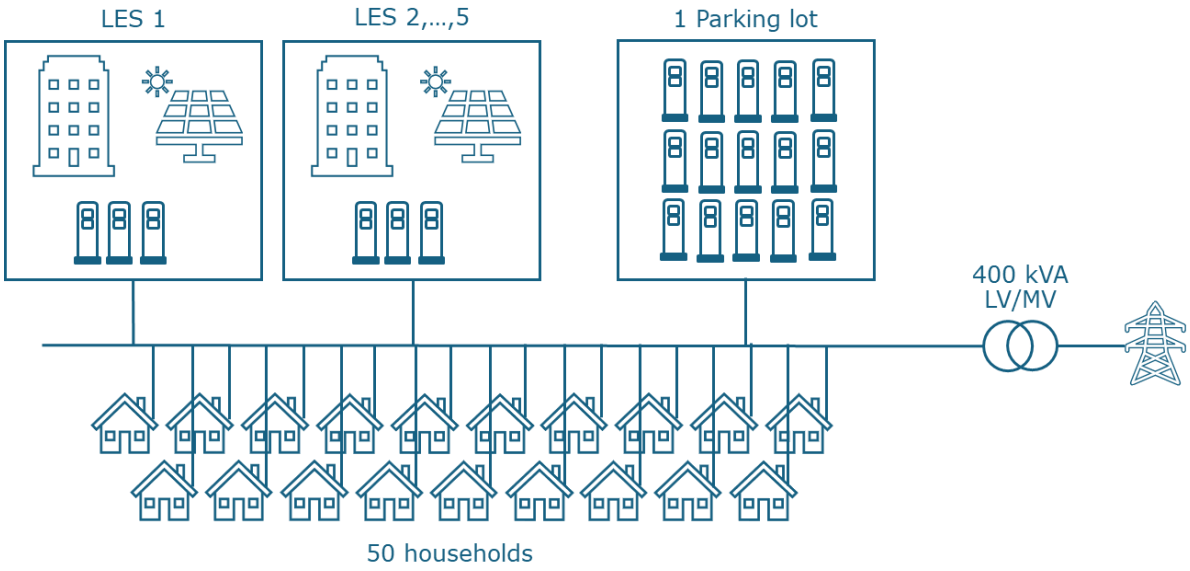

Figure 17: Simplified scheme of the use case

Households with HEMS

The simulation includes 50 households equipped with a home energy management system (HEMS). This configuration has already been modelled and analysed in Deliverable 5.2, and the corresponding power profiles from that work will be reused here. This ensures consistency with previous analyses and allows the focus of the present work to remain on the aggregated system behaviour.

Local energy systems with model-based EMS

The case study also incorporates five LES operating with the model-based EMS presented in Section 2. In the context of this report, each LES emulates an office building equipped with photovoltaic (PV) panels and an associated parking lot for EV charging. The five LES differ in size, with the installed PV capacity, number of charging points, and typical load profiles detailed in Table 4.

Table 4: The different LES characteristics

LES ID	# Charge points	Size of PV [kWp]	Yearly building consumption [MWh]
1	10	50	84.4
2	15	75	126.6
3	20	100	168.7
4	25	125	210.9
5	30	150	253.3

Large parking lot with model-free EMS

The case study also includes one large parking facility managed by the model-free EMS presented in Section 4. The site represents a high-density EV charging hub with an average of 100 employee-type commuters. This type of commuter can be characterized as morning arrival and afternoon departure, which are parked every day of the working days, but charge only 1 time of 3. The parking lot contains 50 charge points, which is enough to satisfy the charging demand of these 100 employees.

Electricity Tariffs

All EMS in the simulation operate under the same electricity pricing scheme, combining dynamic energy tariffs from 2023 with capacity-based tariffs. This common framework ensures comparability across use cases and allows for a consistent evaluation of EMS performance.

Simulation

The case study is simulated on the full year 2023 with 15-minute timestep resolution.

5.3 Results

The results of the simulated case study are presented in Figure 18, where the full year is summarized into one-day by displaying the mean, 75th percentile, and maximum values per quarter-hour. Furthermore, the figure contains two columns: a) the first shows the results without EMS, and b) the second shows the results with EMSs. The aggregated power results of the houses, offices, parking lot, and the total system are shown on lines 1, 2, 3, and 4, respectively.

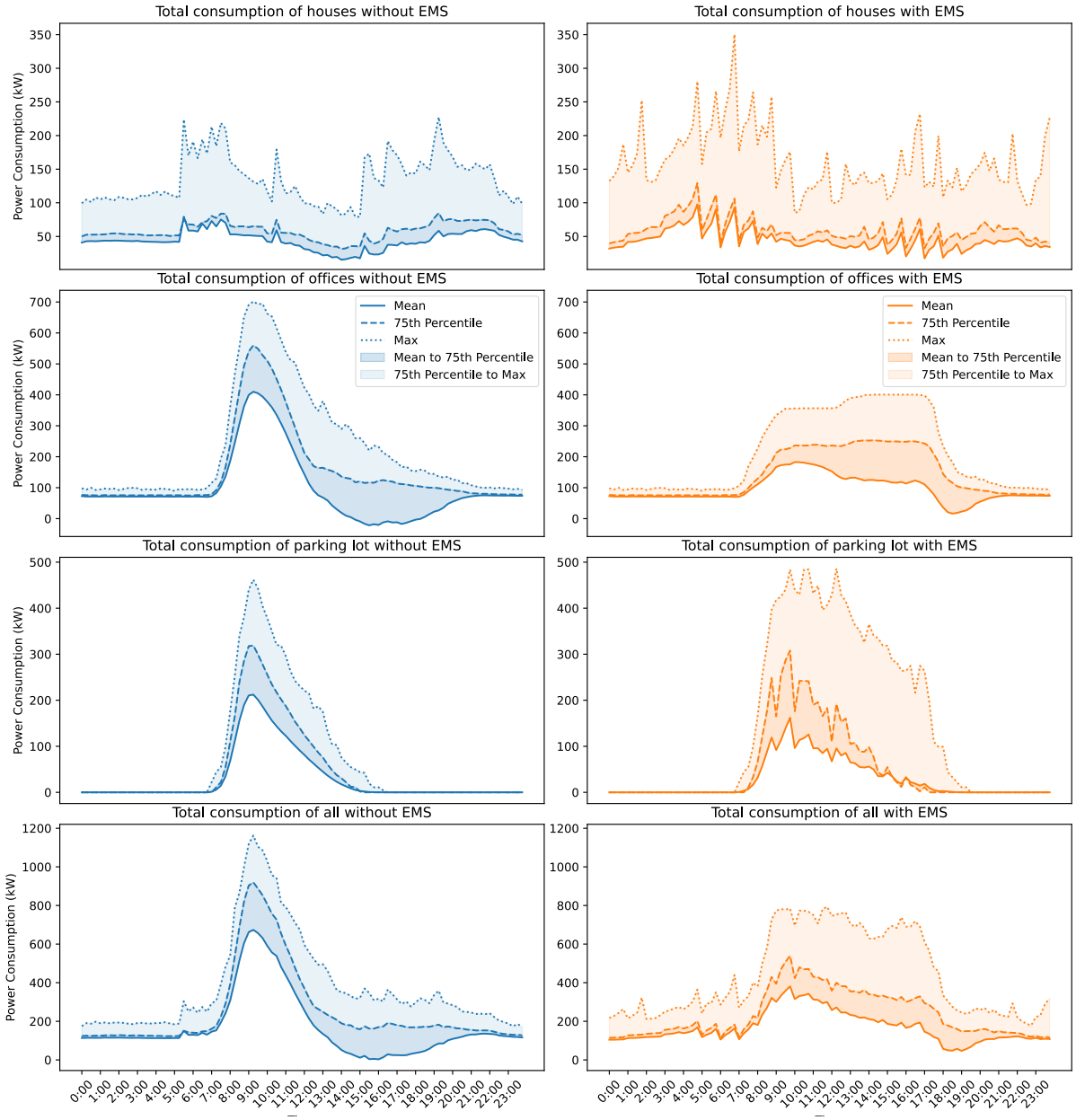

Figure 18: Results of total consumption per case and with and without EMS

Total consumption of houses

It can be observed on the first line, which represents the total consumption of the houses, that adding EMSs results in higher peak powers. For example, the maximum peak increased from 227.5 kW without EMS to 350.2 kW with EMS, corresponding to a 53.9% increase. It is worth noting that this 350.2 kW peak occurred only once during the full-year simulation. Another noteworthy observation can be made in the morning and late afternoon periods, where additional peak powers are visible in sawtooth pattern. These additional peaks can be explained mainly by the operation of the heat pump rather than the electric vehicles. The heat pump typically runs in the morning and evening when households require heating. In addition, it tends to switch on at the end of each hour to take advantage of lower prices and to compensate for thermal losses. Additional explanations and figures are available in Deliverable 5.2.

Total consumption of offices

It can be observed on the second line, which represents the total consumption of the LES, that adding EMSs results in **lower** peak powers. For instance, the highest mean peak power decreased from 410.2 kW without EMS to 182.7 kW with EMS, corresponding to a 55.5% reduction. It should be noted, however, that the simulation with EMS assumes a perfect forecast of production and consumption assets, which is not the case in reality. Nevertheless, the results demonstrate that the EMS strongly reduces peak power consumption, as it was designed to do.

Total consumption of parking lot

It can be observed on the third line, which represents the total consumption of the parking lot, that adding an EMS results in peak powers of similar intensity but occurring at different times. With the EMS, these peaks are distributed across the working hours. The reason the peak powers are not reduced is that the EMS was not designed for peak shaving, but rather to take advantage of imbalance prices. In other words, while such peak powers may appear problematic at the local level (i.e., for the DSO), they are in fact advantageous at the national level (i.e., for the TSO).

Total consumption of the neighbourhood

It can be observed on the fourth line, which represents the total consumption of the neighbourhood, that adding EMSs results in lower peak powers. For example, the highest mean peak power decreased from 673.6 kW without EMS to 380.7 kW with EMS, corresponding to a 43.5% reduction. However, the largest peak with EMS reaches 793.2 kW, of which 61.3% originates from the parking lot alone. This indicates that a single parking lot with the EMS in imbalance mode can have a significant impact on the neighbourhood, which can be explained by the fact that this EMS does not account for peak power.

5.4 Discussion

The results clearly show that EMSs have a significant impact on the power consumption profiles of households, offices, parking lots, and actually the total neighbourhood. However, this impact can be either positive or negative depending on the use case, on the specific objective of the EMS, and on the control strategy employed. A fundamental limitation of current EMSs is that they are typically designed for a specific objective, without considering the broader context or the combined effect of multiple EMSs operating simultaneously. Moreover, this study focuses on a neighbourhood and demonstrates that certain EMSs can induce substantial peak powers. Extrapolating these results to a larger scale, such as a national grid, suggests that uncoordinated EMS deployment could potentially lead to serious, unexpected, and unmanaged peak power demand. This highlights the need for further investigations, which may indicate the necessity of coordinating EMS operation across multiple sites.

A key limitation of the present simulations is that they do not account for potential “feedback loops” at the system-level. For example, if a model-based EMS shifts demand to a later time to take advantage of lower prices, it may create imbalances at the grid level, which could be partially mitigated by other, model-free EMSs. Such interactions were not modelled or

simulated in this study. Finally, the simulation results are highly case-study dependent. To obtain a more comprehensive assessment of EMS impact at larger scales, further research is required, including a wider variety of scenarios, control strategies, and grid conditions.

6 Conclusions

This deliverable presents three different EMS structures for smart charging of EVs. Specifically, the study investigates the performance of these EMSs under explicit/implicit market signals and analyses the overall impact of their combined operation on the power system. Simulation results show that EMS implementation can significantly reduce peak demand at the neighbourhood level, achieving a 43.5% reduction in the highest mean peak power. However, the results also indicate that the largest instantaneous peak during EMS operation is primarily driven by the parking lot, which contributes 61.3% of the total peak. This occurs because the parking lot's EMS is not configured with peak reduction as a control objective. Overall, the findings demonstrate that although each EMS performs effectively within its own market framework and objectives, their simultaneous operation in a shared environment produces varying impacts across different power system stakeholders. While these EMSs can effectively mitigate or delay the system's peak demand, they may also lead to the emergence of secondary, smaller peaks at other times. These observations highlight the importance of developing coordinated and system-aware EMS strategies that promote collective efficiency rather than isolated optimization. Moreover, it is essential for system operators to consider these cumulative effects when designing market mechanisms and control strategies to ensure overall grid stability.

This study can be further extended to examine the impact of large-scale EMS deployment across scenarios beyond the neighbourhood level, with an in-depth analysis of the resulting implications from both DSO and TSO perspectives. By doing so, system operators can evaluate how their price signals influence EMS responses and assess the consequent effects on the overall power system. It is also worth considering the feedback dynamics whereby large-scale changes in EMS behaviour, driven by explicit or implicit signals, could influence future price formation and subsequently alter EMS responses once again.

References

[1] A. Mangipinto, F. Lombardi, F. D. Sanvito, M. Pavičević, S. Quoilin, and E. Colombo, 'Impact of mass-scale deployment of electric vehicles and benefits of smart charging across all European countries', *Appl. Energy*, vol. 312, p. 118676, Apr. 2022, doi: 10.1016/j.apenergy.2022.118676.

[2] M. Amjad, A. Ahmad, M. H. Rehmani, and T. Umer, 'A review of EVs charging: From the perspective of energy optimization, optimization approaches, and charging techniques', *Transp. Res. Part Transp. Environ.*, vol. 62, pp. 386–417, 2018, doi: 10.1016/j.trd.2018.03.006.

[3] G. Van Kriekinge, C. De Cauwer, N. Sapountzoglou, T. Coosemans, and M. Messagie, 'Peak shaving and cost minimization using model predictive control for uni- and bi-directional charging of electric vehicles', *Energy Rep.*, vol. 7, pp. 8760–8771, 2021, doi: 10.1016/j.egyr.2021.11.207.

[4] Van Kriekinge, Gilles, 'Real-World Implementation of Smart Charging : Challenges and Lessons Learned', presented at the EVS36, 2024.

[5] Elia, 'Open Data'. [Online]. Available: https://opendata.elia.be/pages/home/

[6] F. D'Ettorre *et al.*, 'Exploiting demand-side flexibility: State-of-the-art, open issues and social perspective', *Renew. Sustain. Energy Rev.*, vol. 165, p. 112605, Sept. 2022, doi: 10.1016/j.rser.2022.112605.

[7] S. N. Jahromi, G. Van Kriekinge, C. De Cauwer, and T. Coosemans, 'Scheduling Electric Vehicle Charging for Participation in the Belgian Imbalance Market Using Model-Free Reinforcement Learning', presented at the The 38th International Electric Vehicle Symposium & Exposition, EVS38, 2025, pp. 1–12.

[8] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, 'Proximal policy optimization algorithms', *ArXiv Prepr. ArXiv170706347*, 2017.

[9] Z. Zhu, Z. Hu, K. W. Chan, S. Bu, B. Zhou, and S. Xia, 'Reinforcement learning in deregulated energy market: A comprehensive review', *Appl. Energy*, vol. 329, p. 120212, 2023.

[10] R. S. Sutton, 'Reinforcement learning: An introduction', *Bradf. Book*, 2018.

[11] G. Van Kriekinge, C. De Cauwer, N. Sapountzoglou, T. Coosemans, and M. Messagie, 'Peak shaving and cost minimization using model predictive control for uni-and bi-directional charging of electric vehicles', *Energy Rep.*, vol. 7, pp. 8760–8771, 2021.

[12] 'End User Documentation "1-minute publications"', 2019. Accessed: Mar. 31, 2025. [Online]. Available: https://www.elia.be/-/media/project/elia/elia-site/grid-data/balancing/20190827_end-user-documentation-elia1-minute-publications.pdf

[13] G. Van Kriekinge, C. De Cauwer, N. Sapountzoglou, T. Coosemans, and M. Messagie, 'Electric vehicle charging sessions generator based on clustered driver behaviors', *World Electr. Veh. J.*, vol. 14, no. 2, p. 37, 2023.

[14] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, 'Optuna: A next-generation hyperparameter optimization framework', presented at the Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2019, pp. 2623–2631.

[15] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, 'Stable-baselines3: Reliable reinforcement learning implementations', *J. Mach. Learn. Res.*, vol. 22, no. 268, pp. 1–8, 2021.

[16] R. Guo, S. Meunier, C. Protopapadaki, and D. Saelens, 'A review of European low-voltage distribution networks', *Renew. Sustain. Energy Rev.*, vol. 173, p. 113056, Mar. 2023, doi: 10.1016/j.rser.2022.113056.

[17] N. Beck, J. Dovern, and S. Vogl, 'Mind the naive forecast! a rigorous evaluation of forecasting models for time series with low predictability', *Appl. Intell.*, vol. 55, no. 6, p. 395, Apr. 2025, doi: 10.1007/s10489-025-06268-w.