D4.3 Optimal selection of available flexibility (v1)

28 February 2025

opentunity

OPENTUNITYproject.eu

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. Horizon Europe Grant agreement N° 101096333.

Deliverable details

Title	WP	Version
Optimal selection of available flexibility (v1)	4	3

Contractual delivery date	Actual delivery date	Delivery type*	Dissemination**
M26 (February 2025)	M26 (February 2025)	R	PU

*Delivery type: R: Document, report; DEM: Demonstrator, pilot, prototype; DEC: Websites, patent fillings, videos, etc; OTHER; ETHICS: Ethics requirement; ORDP: Open Research Data Pilot.

Dissemination Level: **PU: Public; **CO**: Confidential, only for members of the consortium (including the Commission Services)

Author(s)	Organization
Janez Gregor Golja	UL
Klemen Peter Kosovinc	KOL

Version	Date	Person	Action	Status***
v0	7.1.2025	Janez Gregor Golja (UL)	First draft created	Draft
v1	7.2.2025	Janez Gregor Golja (UL), Klemen Peter Kosovinc (KOL)	Added contributions from KOL	Draft
v2	20.2.2025	Janez Gregor Golja (UL)	Document ready for review	Draft
v3	27.2.2025	Janez Gregor Golja (UL), Klemen Peter Kosovinc (KOL), Kore Adenuga (HIVE), Yannis Syrimpeis (QUE)	Peer review performed – final version	Final

***Status: Draft, Final, Approved, Submitted (to European Commission).

Authors (organization)
Janez Gregor Golja (UL), Tomi Medved (UL), Klemen Peter Kosovinc (KOL), Miha Valentinčič (AVA), Tomaž Buh (AMI), Edin Lakić (IRI UL), Matej Pečjak (IRI UL)

Keywords

Baseline forecasting, flexibility, electric vehicle, optimal selection of flexibility, energy markets, machine learning

Executive Summary

This deliverable presents the progress made in the first phase of the innovation "Optimal Selection of Available Flexibility", which consists of two key aspects: identifying market suitability of flexibility assets and developing an algorithm to assist aggregators in selecting which assets to activate. The initial phase established the foundation for both aspects by leveraging data from two fleets—EV charging stations and HEMS devices—to perform analyses and develop advanced baseline forecasting models. These models will be further utilized in the second phase to determine the optimal flexibility utilization options and serve as inputs for testing the optimal selection algorithm.

The document details the methodology used in this phase, including the integration of the two fleets into the KOL aggregation platform and an overview of the selected forecasting techniques. It also presents the process and results of baseline forecasting, which was conducted using different approaches: aggregated fleet-level and individual location forecasting for EVs, and a household-level approach for HEMS assets. Building upon these forecasts, the flexibility modelling process for different assets is explained, outlining how flexibility potential is quantified. Additionally, the deliverable explores various flexibility utilization options from a theoretical perspective, preparing the groundwork for the next phase where EV and HEMS fleet data and forecasting models will be used to assess their viability.

The insights and methodologies presented in this deliverable will serve as key inputs for the next stage, where both aspects of optimal selection will be demonstrated. This structured approach ensures that the findings in the next version of the deliverable are grounded in real-world data from both fleets, providing tangible results.

Copyright statement

The work described in this document has been conducted within the OPENTUNITY project. This document reflects only the OPENTUNITY Consortium view and the European Union is not responsible for any use that may be made of the information it contains.

This document and its content are the property of the OPENTUNITY Consortium. All rights relevant to this document are determined by the applicable laws. Access to this document does not grant any right or license on the document or its contents. This document or its contents are not to be used or treated in any manner inconsistent with the rights or interests of the OPENTUNITY Consortium or the Partners detriment and are not to be disclosed externally without prior written consent from the OPENTUNITY Partners.

Each OPENTUNITY Partner may use this document in conformity with the OPENTUNITY Consortium Grant Agreement provisions.

Index

List of figures

List of tables

1 Introduction

1.1 Purpose of the document

This deliverable provides an overview of the progress made in Task 4.4: Optimal Selection of Available Flexibility during its first phase. It outlines the methodology and results of baseline forecasting for both the electric vehicle (EV) charging station (CS) fleet and Home Energy Management System (HEMS) fleet. Additionally, it introduces the identification of flexibility utilization options, setting the groundwork for the next phase of the task.

1.2 Scope of the document

The document explains the approach taken in the first stage of Task 4.4, detailing the methodology applied and the results obtained from baseline forecasting using data from two integrated fleets within the KOL aggregation platform. It also defines flexibility for these assets and provides the theoretical background for various flexibility utilization options, including markets and optimization opportunities.

1.3 Structure of the document

The document is structured in the following sections:

- **Section 2** outlines the goals of the task and the approach taken, introducing the two dimensions of optimal selection. It also provides an overview of the aggregator's role and presents a high-level framework of the aggregator's decision-making process.
- **Section 3** details the methodology, including the workflow overview and the integration process of the two fleets: the EV CS fleet (AVA) and the Reduxi HEMS fleet (AMI). This section also covers the forecasting methods used, focusing on time series forecasting, with a literature review and a presentation of the techniques which were selected.
- **Section 4** presents the results of the baseline forecasting for both fleets. The EV fleet forecasting is divided into two approaches: fleet-level forecasting, which predicts aggregate charging demand, and individual location forecasting, which captures station-specific charging patterns. For the HEMS fleet, forecasting was performed at the household level, modelling individual assets separately, including smart meter consumption, photovoltaic (PV) generation, and heat pump (HP) consumption. Additionally, generic models for PV and HP were developed and tested.
- **Section 5** defines the flexibility estimation models for the modelled assets, detailing the technical constraints and user comfort requirements and formulating them into equations to quantify available flexibility.
- **Section 6** sets the stage for the second phase of the task, with the identification of flexibility utilization options, where the flexibility assets could be used, exploring their characteristics and barriers.

2 Task goals

The primary objective of task T4.4 is to develop a methodological and logical framework for the optimal selection of available flexibility across both small and large assets. This framework leverages advanced techniques—including big-data analytics, AI, and machine learning—to support energy market participation by aggregators. In addition to addressing technical factors, the framework must also consider user preferences, such as maintaining indoor comfort for HPs and ensuring a sufficient state of charge for EVs.

We interpreted optimal selection in two ways:

1. **Market applicability and technical suitability of the flexibility asset:** In this interpretation, optimal selection involves identifying the markets where an asset can be most effectively utilized, based on its technical characteristics and forecast performance. To do so, we analyse data from both the Avantcar EV CS fleet and the Amibit Reduxi HEMS fleet. By developing baseline forecast models for each asset, we are able to assess how accurately we can predict their normal operation. These forecasts are critical because they not only indicate the potential deviations (i.e., flexibility) from the baseline, but they also inform us about the asset's technical suitability for specific market applications.
2. **Optimal selection algorithm:** The second way refers to the algorithmic process—often termed disaggregation—used by the aggregator to select which assets to activate when a bid is accepted. Here baseline and consequently flexibility forecasts play a key role for generating bids and disaggregation process.

Although flexibility forecasting is addressed in Task 4.1, we developed our own baseline models to gain a deeper understanding of our fleets' data and operational behaviours. These insights from our detailed data analysis and forecasting will serve as critical inputs for future work on both aspects of optimal selection.

To structure our work, we split the deliverables into two parts. The first deliverable lays the foundation by focusing on data analysis, baseline forecasting, and identifying potential flexibility utilization options. This groundwork is essential for the Slovenian pilot demonstration, where flexibility will be showcased through alternative means rather than via the NODES local flexibility market.

The second deliverable will build upon this foundation by evaluating the participation potential of the two flexibility sources within the identified market options. It will also include the formulation and testing of the optimal selection (disaggregation) algorithm—primarily aimed at local flexibility markets—in several OPENTUNITY pilot projects.

The following sections will elaborate on the aggregator's role and its operational process, which will help contextualize the work we did so far.

2.1 The role of aggregator

An aggregator acts as an intermediary between flexibility assets, their owners, and energy markets, utilizing these assets to provide flexibility services. By enabling end-users to actively participate in electricity markets and supporting grid operations, aggregators are essential to modern energy systems. They help overcome traditional market barriers that limit the participation of smaller-scale assets by pooling their flexibility.

Aggregators may be electricity retailers or independent entities [1], and they combine multiple customer loads or generated electricity for sale, purchase, or auction in energy markets. By managing a diverse portfolio of devices—ranging from small-scale generation units (like PVs and wind farms) to EVs and battery energy storage systems (BESS)—aggregators unlock two key economic values:

1. **System Value**: Aggregating flexibility increases the overall efficiency and stability of the power system.
2. **Private Value**: Aggregators provide asset owners with access to revenue streams that would otherwise be unavailable.

To effectively perform their role, aggregators must possess deep expertise in both the technical characteristics of distributed energy resources (DERs) and the dynamics of electricity markets. Their capabilities can be broadly divided into two core areas:

1. **Flexibility Expertise**: This involves understanding the operational constraints and user preferences of various distributed energy resources (DERs)—such as forecasting baseline consumption and flexibility potential—and translating these into market opportunities.
2. **Market Expertise**: Many DER owners lack detailed knowledge of electricity markets. Aggregators bridge this gap by acting as market representatives, leveraging their understanding of market structures and pricing dynamics to optimize participation and generate economic benefits.

Having outlined the role of aggregators in bridging DERs and energy markets, it is important to understand how these roles are operated in practice. In the following section, we provide high-level overview of the aggregator process, detailing the key data streams and operational steps that enable aggregators to combine and manage flexibility assets effectively.

2.2 Aggregator process overview

This section provides a high-level view of the aggregator's operational framework [2], describing the main components involved and their interactions. The goal is to illustrate how the "optimal selection of available flexibility" (also known as disaggregation) fits within the broader aggregation process.

A simple architecture of this process is shown in Figure 1

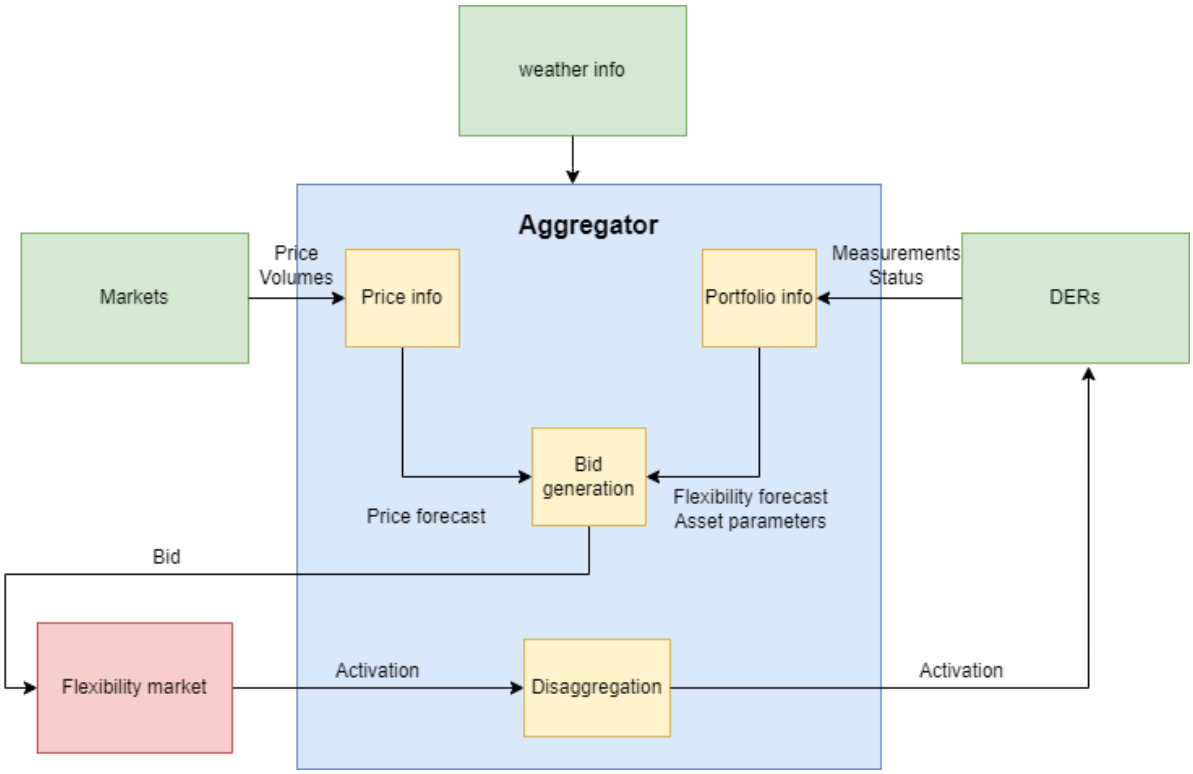

Figure 1: Architecture of the aggregator's process

The aggregator's main activity is to combine flexibility from multiple DERs and generate aggregated bids that maximize revenue in various energy or flexibility markets.

The aggregator's activities begin with gathering essential information from three main streams:

1. **Market stream**: The aggregator collects relevant market data, such as spot prices, intraday prices, and historical regulation prices. These datasets are stored in the aggregator's database and analyzed for patterns like seasonality or price trends, forming the basis for estimating the value of flexibility in upcoming markets.
2. **DER stream**: The aggregator regularly obtains real-time and historical measurements from DERs, including their status, availability, and power consumption. It also records static characteristics (e.g., capacity, power limits, or maximum activation intervals) that affect how each resource can provide flexibility. Based on these inputs, the aggregator constructs a portfolio view of available flexibility, estimating both quantity and associated costs.
3. **Weather information stream**: External weather data (e.g., temperature, solar irradiation) is used for analysis and forecasting.

Using these information streams, the aggregator proceeds with the following steps:

1. **Bid definition**:
   - The aggregator generates aggregated bids as price-volume points, balancing flexibility opportunities against revenue potential. The strategy includes considerations of several markets and pricing dynamics to maximize revenue.
   - Bids do not merely reflect activation costs but incorporate strategic pricing to optimize the aggregator's profitability.
2. **Disaggregation**:
   - Before finalizing a bid, the aggregator often performs a preliminary disaggregation to verify feasibility. This step involves translating each potential bid volume into specific DER setpoints, ensuring the portfolio can actually deliver that level of flexibility. Preparing these instructions in advance speeds up the activation process later on.
3. **Submission**:
   - Once the aggregator determines a suitable bidding strategy, the final bids are submitted to the flexibility market. The submitted bids and their subsequent outcomes are stored for statistical analysis and future improvements.

When the market operator activates a bid, two additional processes come into play:

4. **Bid reference and disaggregation**:
   - The aggregator retrieves the activated bid details and may re-check the portfolio's status, DER measurements, or updated forecasts to ensure the requested flexibility can still be delivered. If necessary, it refines the asset-level disaggregation based on real-time conditions, such as a DER's unavailability.
5. **Setpoint communication**:
   - The aggregator sends specific instructions to each DER, translating the disaggregated activation into actionable commands. Throughout the activation period, the aggregator continuously monitors performance, looking for deviations that might compromise the committed flexibility. If deviations occur, corrective measures are taken, such as adjusting the disaggregation to replace unresponsive assets.
   - The goal of this process is to deliver the assigned flexibility in the most reliable and efficient manner, ensuring compliance with market obligations.

The work completed so far represents the initial phase of the aggregation process. We have successfully established an information flow from flexibility assets to the aggregator and developed baseline forecasting models that are critical for both identifying the optimal market application of each asset type and estimating its flexibility potential. These forecasts serve as essential inputs for the subsequent disaggregation process. This deliverable details our methodologies and results, setting the stage for the next developments within the scope of T4.4.

3 Methodology

This chapter details the methodology employed during the initial phase of the Task 4.4 timeline. It provides an overview of the overall workflow, describes the process used to integrate the EV CSs and HEMS devices to the KOL aggregator platform, and reviews the forecasting techniques applied to generate baseline forecasts for both fleets.

3.1 Workflow overview

The work was carried out in five main phases, which are shown in Figure 2. Each phase builds on the outcomes of the previous one, starting from device integration and data collection, through baseline and flexibility forecasting, and concluding with the identification of potential flexibility utilization options. This phased approach ensures that every step—ranging from raw data acquisition to the final assessment of flexibility utilization options identification—has a clear objective and outcome, laying a solid foundation for subsequent tasks in T4.4.

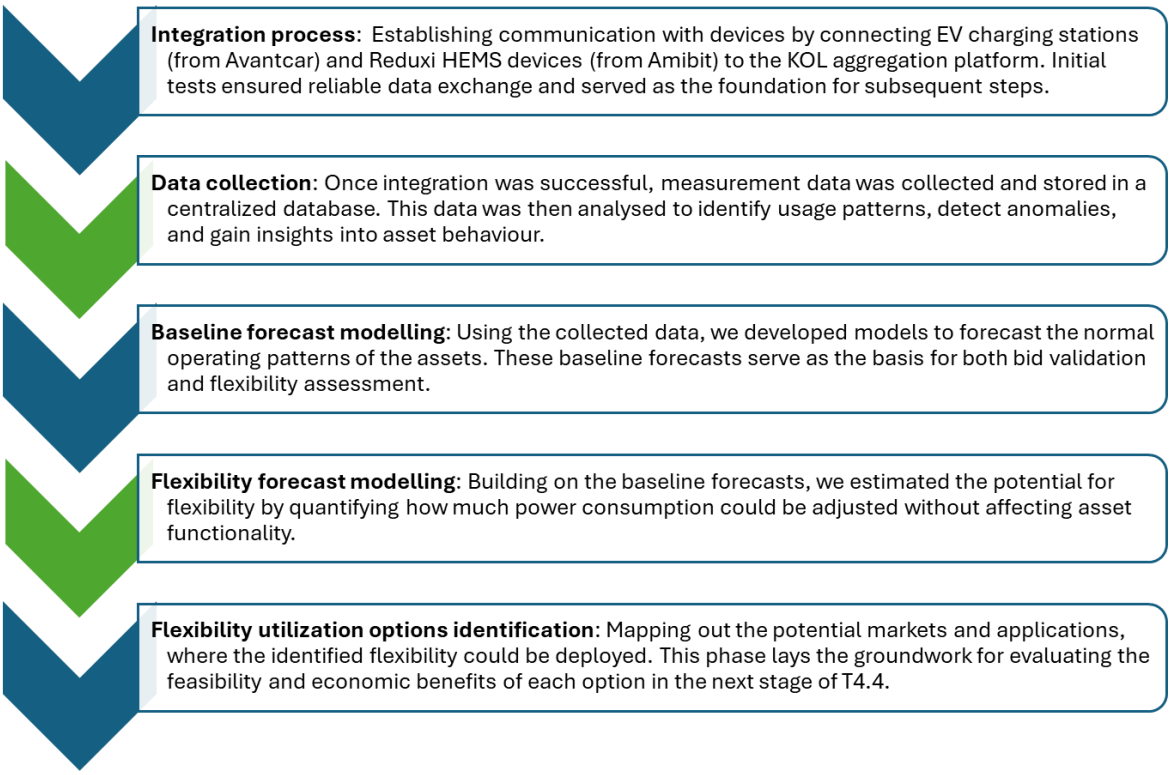

Figure 2: Task 4.4 workflow overview during stage 1

3.2 Integration process

The integration process of EV CSs and HEMS devices was divided into three phases as shown in Figure 3.

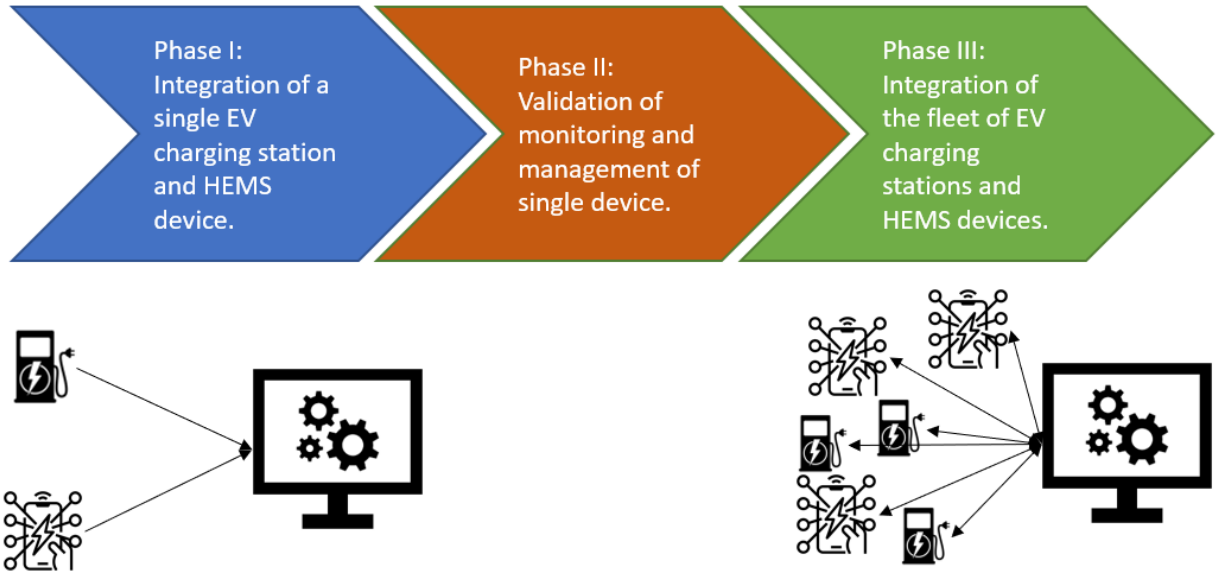

Figure 3: Integration of process EV charging points and HEMS devices.

- **Phase I** focused on connecting a single EV CS and a single HEMS device to the aggregator platform. The HEMS device was integrated via an MQTT server set up on the aggregator side. Once configured, the HEMS device began sending JSON-formatted data to the platform, where it was parsed, validated, and stored in a central database for further analysis. For the EV CS, an Application Programming Interface (API) was employed. Avantcar provided the necessary authorization credentials, allowing the aggregator platform to periodically request JSON-formatted measurement data. This data was similarly parsed, validated, and stored in the database.
- **Phase II** concentrated on assessing connectivity and reliability. Once both devices were successfully integrated, measurement data flowed continuously to the aggregator platform. During this phase, remote management capabilities for the HEMS device were tested. By sending setpoints to the PV plant via the aggregator platform, the PV's output could be curtailed as needed, confirming that control commands were effectively transmitted and executed.
- **Phase III** involved identifying the types of EV CSs in Avantcar's fleet and the appliances connected to Amibit's HEMS devices. Through discussions in the Slovenian pilot, CSs and units with the greatest flexibility potential were selected. A total of 70 CSs—covering car-sharing and operating categories—were integrated into the aggregator platform. Meanwhile, 51 HEMS devices were connected, encompassing a variety of connected device types. This portfolio laid the groundwork to start working on baseline forecasting.

For EV charging points, the devices were categorized as follows:

a. **Car-sharing charging points**: These provide limited flexibility, particularly at night, when charging schedules can be adjusted to ensure batteries are fully charged by a specific time while still optimizing the overall charging process.

b. **Operating charging points**: These are used for vehicles collected by Avantcar at the end of the day, including company cars. They offer flexibility mainly during nighttime and to some extent during the day.
c. **Fast charging points**: Designed for rapid charging to minimize downtime, these high-power units are optimized for speed rather than flexibility, so flexibility is not a primary consideration for them.

Based on these classifications, a representative portfolio of car-sharing and operating charging points was selected, resulting in the successful integration of a total of 70 charging points into the KOL aggregator platform.

For HEMS devices, the following appliance categories were identified as relevant for flexibility: PV power plants, BESS, HVAC devices, and EV charging points. Amibit enabled the integration of 51 HEMS devices into the KOL aggregator platform, resulting in a diverse portfolio that includes 51 smart meters, 43 PV power plants, 10 BESS, 13 heating, ventilation, and air conditioning (HVAC) systems, and 19 EV CSs.

The following subsections detail the data structures for both Avantcar's EV CSs and Amibit's HEMS devices. For each portfolio, metadata and measurement data tables have been prepared and will be explained. Additionally, the functional requirements for data handling within the aggregator platform—ensuring scalability, reliability, and flexibility—are discussed.

3.2.1 AVA EV charging stations

The Avantcar's CSs metadata parameters are presented in Table 1.

Table 1: Avantcar's CSs metadata

Parameter	Description	Example Value
Measuring_point_name	Internal name of the charging point	AcantCar EvCharger 6
Asset_name	Information about the asset location	A2GO SI LJ AC FMF D
Asset_status	Indicates if the charging point is active (for long-term analysis)	Active
Asset_network_status	Network status of the charging point	Available
Asset_remote_id	Internal ID of the charging point	21734
Measuring_point_remote_id	Internal measuring point ID	6
Measuring_point_max_power_w	Maximum power capacity of the CS (in watts)	22,000
Measuring_point_current_type	Type of current provided (AC or DC)	AC

Measuring_point_max_voltage_v	Maximum voltage at the CS connection (in volts)	400
Measuring_point_hardware_status	Current operational status of the CS	Charging

3.2.2 AMI Reduxi HEMS units

The Amibit's HEMS devices metadata parameters are presented in Table 2.

Table 2: Amibit's Reduxi HEMS metadata

Parameter	Description	Example Value
Name	Includes the HEMS ID and channel ID of the connected appliance.	MGV4-BHY9-XHRX-YF2C---2 (HEMS ID: MGV4-BHY9-XHRX-YF2C; Channel: 2)
Category	Specifies the type of connected appliance. This should be configured by the user; if not set, it defaults to NULL.	EV Charger
Other data	Contains additional details such as the internal ID and internal name of the connected appliance. If not configured by the user, default values are applied.	Internal ID: 2; Internal Name: SolarEdge - Garage

3.2.3 Functional requirements

When integrating Avantcar's EV CSs and Ambit's HEMS devices into the KOL aggregator platform, the following requirements were, or will be, considered to ensure the system can accommodate future growth and evolving business needs:

- **Scalability**: The identification table and time-series data tables must be designed to support expansion, allowing for the addition of new parameters as required.
- **Automation**: Data retrieval processes should operate autonomously, executing at defined intervals. When new CSs or HEMS devices are introduced, their data must be integrated automatically into the identification table and time-series data tables.
- **Performance**: The platform must be capable of reliably managing data retrieval and processing for a minimum of 1.000 CSs and HEMS devices, with the ability to scale to accommodate larger numbers.
- **Reliability**: Data retrieval must occur consistently at a defined interval. In the event of a failure, a retry mechanism should ensure that data is successfully captured within the same timeframe.
- **Flexibility**: The system should allow for future reorganization of CSs into measurement points, accommodating potential requirements for registration with entities such ancillary systems market operators such as ELES.

- **Modularity**: The platform's architecture should prioritize modularity and scalability, enabling addition of new features or replacement of components without being tightly coupled to specific technologies.

3.3 Forecasting methods

Forecasting is the process of predicting future outcomes based on historical data, providing critical insights for planning, decision-making, and optimization across various domains [3]. In the energy sector, forecasting is especially important for managing distributed energy resources, enabling demand-side flexibility, and maintaining grid stability. By anticipating energy consumption, flexibility potential, and renewable generation, forecasting helps balance supply and demand, optimize scheduling, and ensure efficient operations.

Forecasting in the energy field often involves time-series data, where observations are recorded sequentially over time. Time-series forecasting differs from other predictive tasks due to its reliance on temporal dependencies, with future values influenced by historical patterns. Time-series data in energy systems presents unique challenges, such as seasonality, nonstationary, and sensitivity to external factors like weather or user behaviour. Addressing these complexities requires robust forecasting techniques capable of capturing both short-term variations and long-term trends.

Forecasting techniques can be broadly categorized into three main groups: statistical methods, machine learning, and deep learning.

1. **Statistical Methods** are often used for time-series forecasting due to their simplicity and interpretability. Methods such as Autoregressive Integrated Moving Average (ARIMA), Seasonal Decomposition of Time Series (STL), and Exponential Smoothing focus on identifying patterns in historical data. These models assume that future values are linearly related to past values, making them well-suited for stationary and seasonal datasets, but lack the ability to capture nonlinear relationships.
2. **Machine learning (ML)** techniques expand on statistical approaches by capturing nonlinear relationships and interactions in the data. Methods such as Support Vector Machines (SVM), Random Forests, Gradient Boosting Machines (GBMs), and XGBoost are capable of handling diverse datasets with complex patterns. ML models often require feature engineering to extract meaningful predictors, such as calendar effects or weather conditions.
3. **Deep Learning (DL)** models, particularly neural networks, have revolutionized time-series forecasting by learning directly from raw data without extensive feature engineering. Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Temporal Convolutional Networks (TCNs) are widely used for energy applications. These models excel at capturing temporal dependencies and sequential patterns in data.

3.3.1 Literature review

Forecasting techniques have become increasingly important in the context of energy systems, particularly with the integration of various DERs into the grid continues to grow. Accurate forecasting

of energy consumption and flexibility plays a key role in optimizing operations, improving scheduling, and enabling demand-side management. By predicting future energy usage patterns and flexibility availability, these techniques help ensure grid stability, enhance resource allocation, and support the efficient integration of distributed energy resources.

To understand the state-of-the-art in forecasting, we reviewed literature focused on methodologies applied to EVs, HEMS, and broader residential devices. Our aim was to identify diverse approaches—from traditional statistical models to advanced machine learning and deep learning techniques—and assess their applicability in predicting energy consumption and flexibility. The key findings from our review are summarized in Table 3.

Table 3: Literature review findings regarding forecasting techniques

Approach	Reference	Methodology & Application
Statistical Methods	[4]	Clustering and statistical modelling to analyse real-world EV charging data and identify patterns that inform demand response strategies
	[5]	Load-shifting potential in diverse EV fleets by applying averaging and regression-based techniques to capture consumption trends.
Machine learning	[6]	Support Vector Regression (SVR) to forecast residential energy consumption, demonstrating capability in modelling nonlinear relationships.
	[7]	Ensemble methods (e.g., Random Forests) to predict charging demand for heterogeneous EV fleets, highlighting adaptability across different operational conditions.
	[8]	XGBoost for handling non-linear and volatile load profiles, especially when combined with decomposition techniques for preprocessing.
Deep learning	[9]	Deep neural networks (DNNs) to forecast EV charging demand by capturing complex temporal patterns and long-term dependencies.
	[10]	Feed-forward neural networks for predicting aggregated residential loads and flexibility, emphasizing their strength in modelling sequential data.
	[11]	LSTM for forecasting fast-charging power demand in EV fleets, showing its effectiveness in capturing rapid temporal variations.

3.3.2 Selected forecast methods

Building on insights from the literature review and the characteristics of our data, two primary forecasting methods were selected for the baseline forecasting: XGBoost and LSTM. XGBoost excels at modelling non-linear relationships and is robust to noisy, sparse, or missing data due to its regularization features and scalability. In contrast, LSTM networks are well suited for capturing temporal dependencies and long-term trends in sequential data—a critical advantage for time-series forecasting in energy systems.

Both methods were applied separately to predict baseline consumption, which consequently helps identify the flexibility potential. This comparative evaluation not only aims to identify the most suitable technique for each forecasting task but also to assess how accurately the consumption patterns can be captured.

In parallel, several tree-based ensemble algorithms were also investigated for forecasting consumption in Avantcar's EV charging fleet. Specifically, the following methods were explored:

- Random Forest Regressor (RFR) [12]
- Light GBM (LGBM) [13]
- Histogram-based Gradient Boosting Regression Tree (HGBRT) [14]

The basis of these tree-based approaches is the decision tree—a predictive model that recursively splits data based on feature values to arrive at a final prediction. For example, in a typical decision tree, shown in Figure 4, the algorithm progressively divides the data by evaluating questions about different features, ultimately following a path from the root to a leaf node that represents a specific combination of decision rules.

Decision trees are intuitive and capable of capturing feature interactions, but they can be prone to overfitting and instability. To overcome these limitations, ensemble methods such as RFR, LGBM, and HGBRT combine multiple trees to improve predictive accuracy and stability.

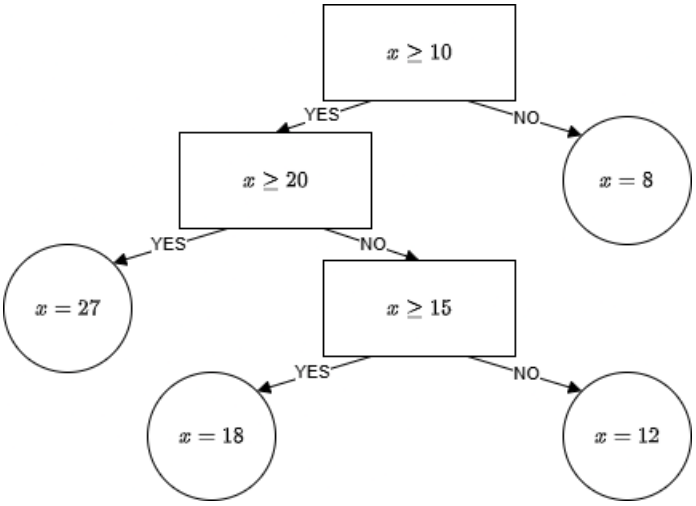

Figure 4: A flowchart visualization of the decision tree method.

The chosen methods are briefly described below:

3.3.2.1 Long Short-Term Memory (LSTM)

LSTM networks are a specialized type of recurrent neural network (RNN) designed to model temporal dependencies in sequential data. Developed by Hochreiter and Schmidhuber, [15] LSTMs overcome the limitations of traditional RNNs—namely, the difficulty in learning long-term dependencies due to vanishing or exploding gradients—making them well-suited for time series forecasting where past events influence future outcomes.

The core architecture of an LSTM revolves around a memory cell that stores information over extended periods. This memory is regulated by three key gates:

- **Input Gate**: Determines which new information is added to the memory cell.
- **Forget Gate**: Decides which information should be discarded from the cell, ensuring that only relevant data is retained.
- **Output Gate**: Controls the extraction of information from the cell for use as output or for passing to subsequent cells.

Together, these gates allow the network to selectively update, retain, or remove information, enabling it to capture both short-term fluctuations and long-term trends. Additionally, LSTMs maintain a hidden state that acts as a form of short-term memory, updated at each time step based on the input, previous hidden state, and cell state.

Key Features of LSTMs

- **Sequential Data Handling**: LSTMs excel at modelling data with inherent sequential structures, such as time-series data.
- **Selective Memory Retention**: By leveraging gates, LSTMs can decide what information to keep or forget, enabling the network to focus on important features.
- **Robust to Long-Term Dependencies**: LSTMs can capture dependencies over extended time horizons, which is critical for forecasting tasks where events from the distant past affect future outcomes.

3.3.2.2 Extreme Gradient Boosting (XGBoost)

XGBoost is a high-performance implementation of the gradient boosting algorithm designed for speed, efficiency, and scalability. Developed by Chen and Guestrin [16], it is renowned for its ability to handle structured data and address a wide range of regression, classification, and ranking tasks. In time-series forecasting, XGBoost is particularly valued for modelling complex, non-linear relationships that traditional methods may struggle with.

XGBoost works by iteratively training an ensemble of decision trees, where each new tree is built to correct the errors (residuals) of the previous trees. This process leverages gradient descent to minimize the loss function, with the incorporation of second-order derivatives enhancing the optimization process and speeding up convergence. A significant strength of XGBoost is its built-in regularization, which penalizes overly complex models (for example, by controlling the number of leaves per tree) to prevent overfitting and improve generalization.

Additional key features of XGBoost include:

- **Scalability**: XGBoost supports parallel and distributed computing, making it suitable for large datasets.
- **Efficient Memory Usage**: Its sparsity-aware algorithms and optimized data structures allow it to handle missing values and sparse data effectively.
- **Feature Importance**: The model provides metrics that help identify which features are most influential in driving predictions, enhancing interpretability.

3.3.2.3 Random Forest Regressor (RFR)

RFR is an ensemble learning method that constructs multiple decision trees and combines their predictions by averaging, which typically leads to more accurate and stable forecasts than a single tree could achieve. In RFR, bootstrap sampling is used to generate different subsets of the original data, ensuring that each tree is trained on a slightly different version of the dataset. This randomness helps to reduce overfitting. Additionally, at each decision split, only a random subset of features is considered, a process known as feature bagging, which further diversifies the trees. Each tree is built by recursively splitting the data based on the best available feature and split point until a stopping criterion—such as a maximum depth or a minimum number of samples in a leaf—is met. Finally, the predictions from all the individual trees are averaged, a process known as bagging, which helps to reduce prediction variance and enhance overall model stability.

3.3.2.4 LightGMB (LGBM)

Unlike Random Forest, which builds trees in parallel and averages their outputs, LGBM employs gradient boosting to construct trees sequentially. Each new tree in LGBM is built to correct the errors (residuals) of the combined ensemble of preceding trees. The core of LGBM is its leaf-wise growth strategy: instead of growing all branches evenly, it expands the branch (leaf) that promises the largest reduction in prediction error. The process begins with a simple initial prediction (often the average of the target values) and then iteratively refines predictions by:

- Calculating residuals for each data point.
- Building a small tree (usually with 8–32 leaves) to predict these residuals.
- Updating the overall prediction by adding the new tree's contribution, scaled by a learning rate.
- Repeating the above steps until the desired number of trees is reached.

LGBM incorporates several optimizations to enhance its efficiency:

- **Gradient-based One-Side Sampling (GOSS):** It retains all data points with large gradients (where the model's error is high) and randomly samples those with small gradients, focusing on the most informative data.
- **Exclusive Feature Bundling (EFB)**: It groups mutually exclusive features together to reduce memory usage and computational cost.
- **Histogram-based Learning**: Continuous features are binned into discrete intervals, speeding up the split finding process.

3.3.2.5 Histogram-based Gradient Boosting Regression Tree (HGBRT)

HGBRT shares similarities with LGBM in using boosting and histogram-based learning but differs in its approach and design priorities. HGBRT uses a traditional level-wise growth strategy: all nodes at a given tree depth are grown before proceeding to the next level. The process includes:

- **Feature Binning**: Continuous features are pre-processed into integer bins at the start, which simplifies the split search.
- **Tree Building**: For each level, histograms of gradient statistics are computed for each feature in each node, and the best split is selected based on these statistics.
- **Missing Value Handling**: HGBRT inherently supports missing values by learning the optimal way to handle them during training.

The key differences between LGBM and HGBRT lie in:

- **Tree Growth Strategy:** LGBM grows trees leaf-wise (expanding the most promising leaf first), while HGBRT grows trees level-wise.
- **Feature Handling:** LGBM adapts feature binning dynamically; HGBRT uses a fixed binning approach.
- **Memory Management:** LGBM is optimized for minimal memory usage, whereas HGBRT's approach may use more memory but offers predictability.
- **Ease of Use:** LGBM provides more hyperparameter tuning options for greater control, whereas HGBRT is designed to work effectively with default settings.

Both LGBM and HGBRT typically yield better prediction accuracy than Random Forest Regressor, especially in time-series forecasting tasks with complex, multi-scale patterns. Their ability to handle missing values, outliers, and feature interactions makes them highly suitable for modelling the daily, weekly, and seasonal variations observed in energy consumption data.

Tree-based methods like RFR, LGBM and HGBRT are particularly effective for time series forecasting. In our case, this data includes daily patterns (peak usage during business hours), weekly patterns (different user behaviour on weekends), monthly patterns (seasonal variations) and yearly patterns (long-term trends in adoption). These methods are great at capturing multi-scale patterns because each split in a tree can focus on a different temporal scale. A single tree might have one branch that captures daily variations and another that captures seasonal trends. When you combine multiple trees into a forest, you get an even more robust ability to model these complex temporal relationships.

These methods also easily handle practical challenges common in time series data. They can work with missing values (which might occur during sensor failures or system downtime), deal with outliers (such as unusual usage patterns during special events), and automatically discover relevant interactions between different features (like how weather conditions might affect charging patterns differently during weekdays versus weekends). Their ability to model non-linear relationships and adapt to evolving patterns over time makes them effective for forecasting tasks where traditional linear methods might struggle to capture the full complexity of the data.

4 Baseline forecast modelling

4.1 Baseline calculation practices

Baselines play a central role in the flexibility provision process by serving as a reference against which actual performance is compared. For example, a Transmission system operator (TSO) or Distribution system operator (DSO) relies on a baseline to verify that a flexibility service was delivered as contracted. In traditional markets, large assets often have fixed schedules that naturally serve as baselines. However, for DERs that are typically aggregated, an alternative methodology for baseline estimation is required.

A baseline is defined as an estimation of what the energy consumption or generation would have been if no flexibility activation had occurred. This estimation is critical because DERs are usually managed in aggregate by retailers or balancing service providers (BSPs), and independent aggregators depend on accurate baselines to:

- Appropriately allocate activations within their portfolios, and
- Ensure compliance with market requirements.

Various baseline calculation methodologies have been proposed and can generally be categorized into two groups:

1. **Historical Data Approaches**: These methods estimate the baseline by analysing past energy consumption or generation data while incorporating additional variables such as weather or calendar effects. Common techniques include:
    - **Forecasting Models**: Advanced tools like regression models and machine learning techniques use historical data and external factors to predict the baseline curve.
    - **Averaging Methods**: Baselines are computed by averaging recent days' consumption data (e.g., the past five days).
    - **High X of Y**: This approach selects the X highest consumption days from the last Y eligible days (e.g., weekdays for a weekday activation) and averages them to form the baseline.
2. **Static Baselines**: These simpler methods, such as the "meter before/meter after" approach, compare the energy demand immediately before and after an activation period. While they are easy to implement, static baselines do not adjust dynamically over the activation period, making them less suitable for scenarios with significant variability.

The selection of a baseline methodology should consider several factors:

- **Accuracy**: The method must reliably estimate what consumption would have been without flexibility activation, thereby minimizing verification errors.
- **Simplicity**: The methodology should be straightforward for Flexibility service providers (FSPs) to understand and implement.
- **Efficacy**: The baseline should be aligned with the market product objectives, ensuring that the incentives promote the desired behaviours.
- **Integrity**: The calculation must be robust against manipulation (e.g., preventing artificial inflation of the baseline prior to activation).

For our case we focused on the baseline methods utilizing historical data. In particular, the forecasting techniques described in Section 3.3.2 were used to capture the variability and temporal patterns inherent in the datasets collected from both EV CSs and HEMS devices.

The results of the baseline forecasts for the EV charging fleet and the HEMS fleet are presented in the following sections, along with analyses of their performance and implications for flexibility provision.

4.2 Avantcar EV charging fleet

This section presents the process and results of forecasting the baseline charging demand for the EV fleet on two levels. One approach, led by KOL, focused on forecasting the aggregated operation of the entire EV CS fleet, while the work conducted by UL concentrated on evaluating selected forecasting methods using data from individual CS locations (with each location comprising between 2 and 10 CSs).

For this study, only historical charging power data was used as the input, meaning that supplementary information—such as the duration of charging sessions or the state of charge of the vehicles—is not available. Consequently, the forecasting efforts focused on estimating the charging demand at both the fleet and individual location levels. The accuracy achieved through these forecasts is crucial, as it can be used in further analyses regarding flexibility potential.

Forecasting charging demand is especially challenging due to the inherent variability in user behaviour. For example, car-sharing services experience major fluctuations in charging patterns as users' arrival and departure times, and their usage behaviour, differ considerably from day to day. From a statistical standpoint, individual locations with a low number of CSs present further difficulties because the smaller sample size results in higher variance and lower statistical reliability.

Despite these challenges, the developed forecasting models provide valuable insights into both the opportunities and limitations of predicting charging demand for such fleets, and they will serve as critical inputs for the next phase of T4.4.

4.2.1 Fleet level

Figure 5 presents the aggregated power profile from all CSs. Two extended periods of missing data—occurring between May and July—were observed, which are attributed to maintenance activities on Avantcar's infrastructure that disrupted data collection. Additionally, a significant seasonal variation is evident; lower power consumption is observed from December to February, likely reflecting reduced EV availability and demand during this period. Consequently, the frequency of charging events is lower, leading to a reduction in total power consumption. For subsequent analysis, the period from July to December is selected, as charging activity was more consistent and stable, an essential condition for effectively training forecasting algorithms.

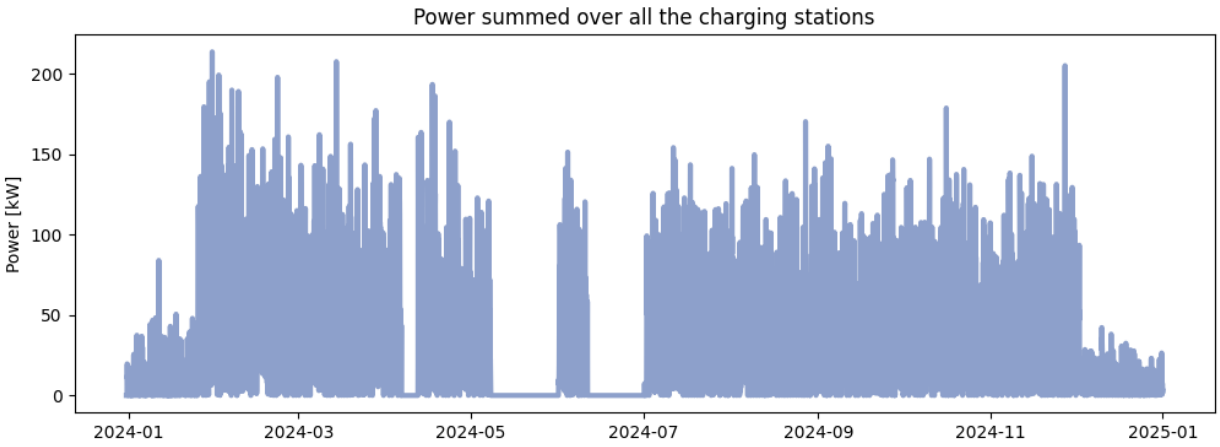

Figure 5: Total power summed over all CSs.

Figure 6 illustrates the number of active CSs throughout the observation period. The maximum recorded availability of EV CSs is 39, representing 54% of the total integrated portfolio. However, the number of active stations fluctuates significantly, ranging between 5 and 35 during the more "active" period (March to November) and between 0 and 10 during the "inactive" period (December to February). Consequently, during the model training phase (July to December), the available CS portfolio ranged between 14% and 42% of the total capacity, highlighting the variability in infrastructure utilization.

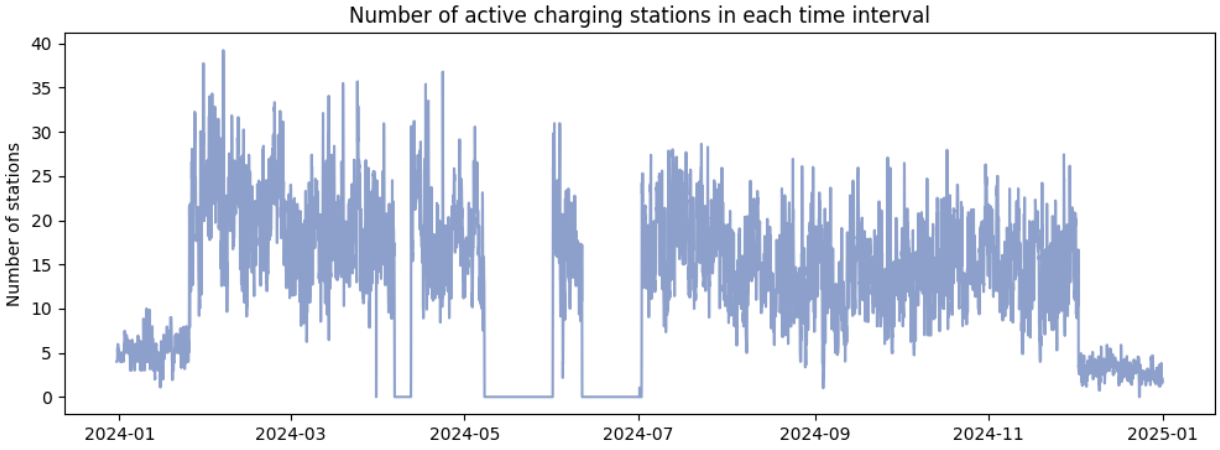

Figure 6: Average number of active EV CSs.

Figure 7 closely aligns with the previous figure, depicting the available power that can be drawn from all active CSs at any given moment. The available power exhibits substantial fluctuations, ranging from 100 kW to 700 kW during the more "active" period (March to November) and between 0 and 200 kW during the "inactive" period (December to February).

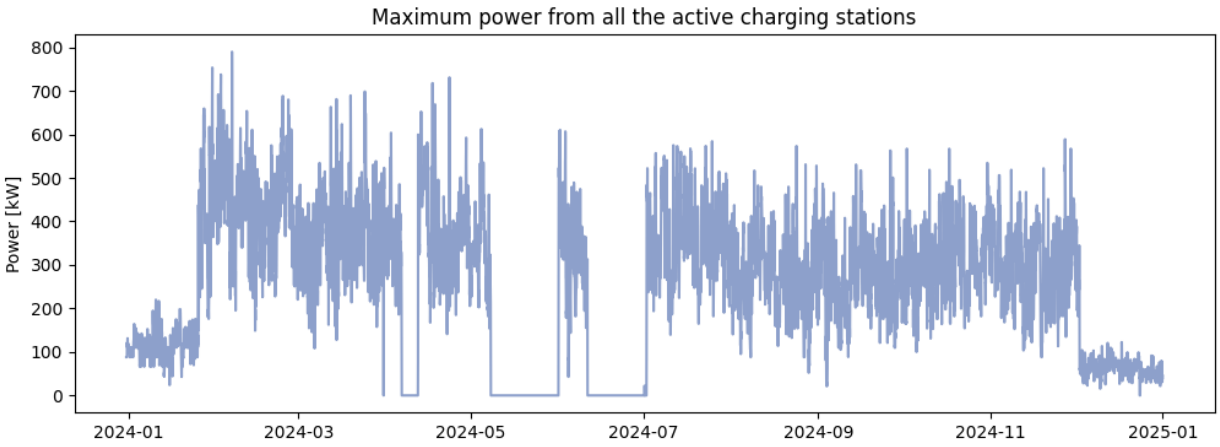

Figure 7: Total available power from all active CSs.

Figure 8 illustrates the daily energy demand required to charge all stations within the Avantcar portfolio. The total daily energy consumption fluctuates, reaching up to 2,000 kWh during the more "active" period (March to November) and up to 300 kWh during the "inactive" period (December to February). Assuming an average daily consumption of 1,000 kWh and considering that a single charging session requires approximately 30 kWh (sufficient for 150 km, based on an efficiency of 20 kWh per 100 km), this level of consumption corresponds to charging approximately 30 EVs per day.

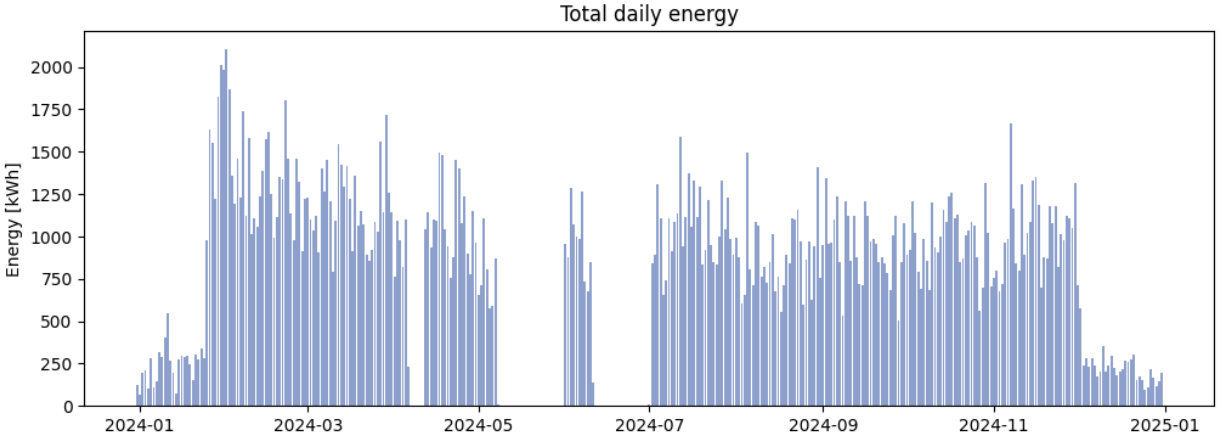

Figure 8: Total daily energy from all CSs.

Figure 9 depicts a representative week that best aligns with the average weekly power profile, selected based on the lowest squared error from the mean profile. The figure reveals a distinct consumption pattern during weekdays, where four out of five cases exhibit two peaks—one in the morning and another in the afternoon. In contrast, weekend consumption follows a different trend, with a single midday peak. Notably, on Sundays, an additional nighttime peak is observed, likely due to fleet charging in preparation for the upcoming workweek.

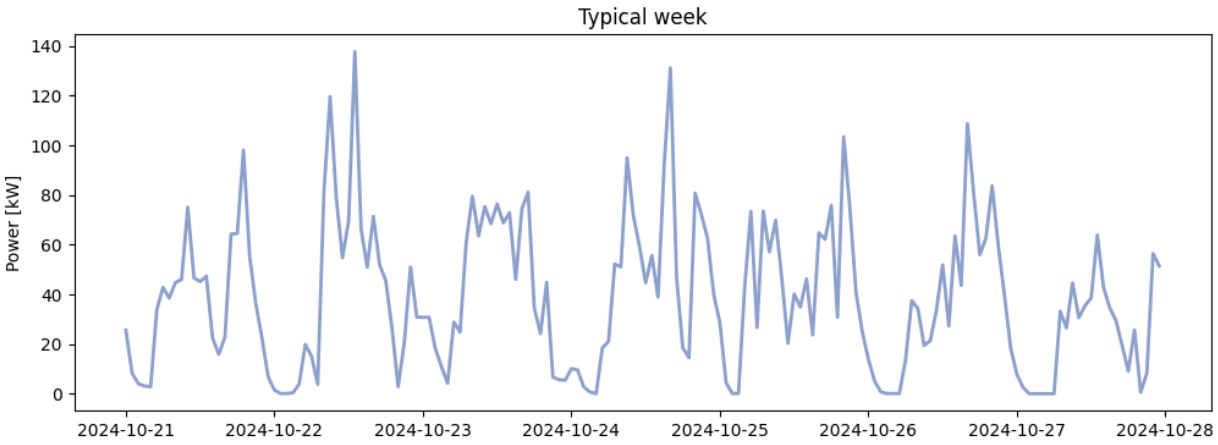

Figure 9: A typical week that best matches the average power profile.

Figure 10 compares the average power consumption on workdays and non-workdays. We can notice that power consumption is higher in the morning on workdays, while on non-workdays it is higher in the evening and during the night. This pattern reflects EV usage, with higher charging activity before or during the morning commutes, while shifting to later hours (presumably due to social events, such as dinners, concerts, etc.) on non-workdays. Furthermore, on non-workdays less consumption is needed for EV CSs fleet.

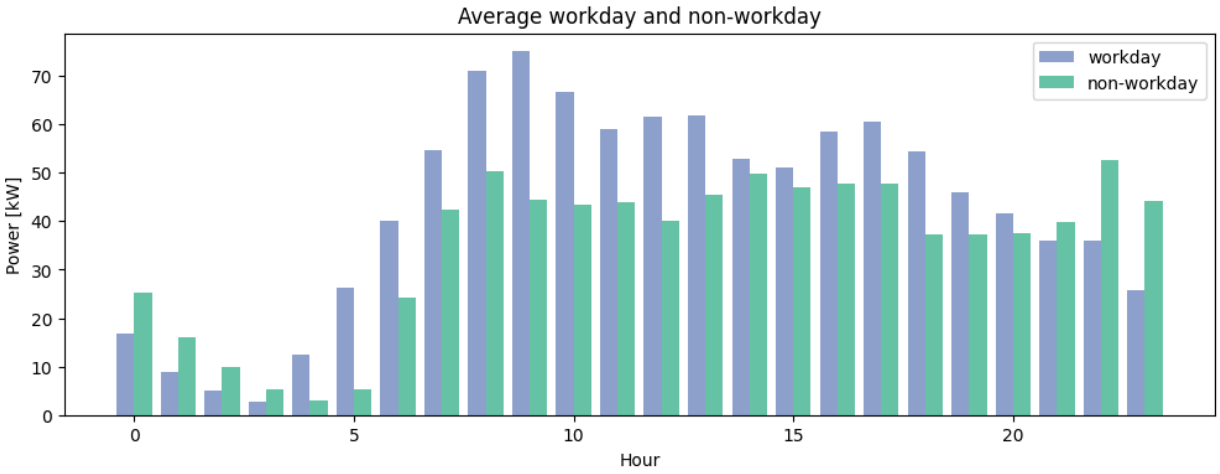

Figure 10: Average daily power profile on workdays and non-workdays.

Figure 11 confirms the previous assumptions, showing that the pronounced morning peak diminishes on weekends. Furthermore, we see that the afternoon peak on workdays is not as outstanding as the morning peak.

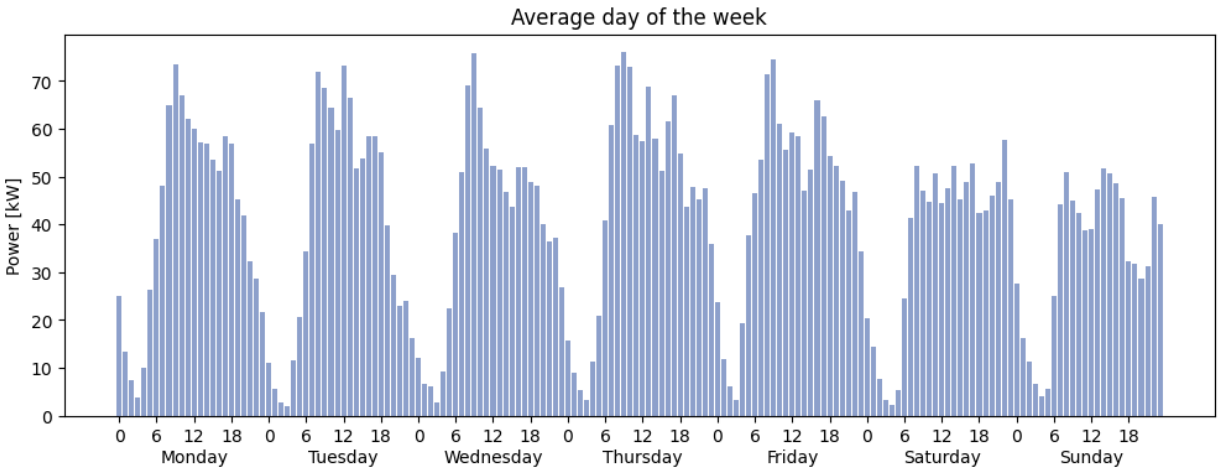

Figure 11: Average power consumption on each day of the week.

During data preparation for the forecasting model, we selected the following features: seasonal attributes such as month, weekday, hour, and work/non-workday indicators. Additionally, we incorporated temperature and precipitation data from the Open-Meteo [17]. The dataset, spanning five months (July to December), was split into a training dataset (four months) and a test set (one month).

To evaluate model accuracy, we used the weighted Mean Absolute Percentage Error (wMAPE):

$$\text{wMAPE} = \frac{\sum_i |A_i - F_i|}{\sum_i A_i}$$

where A_i is the actual value and F_i is the forecasted value in each time interval.

As mentioned in the previous section, we trained models using three different algorithms: RFR, LGBM and HGBRT. We provide an example of the performances during training, shown in Figure 12, Figure 13 and Figure 14. The accuracies achieved on the training set are collected in the Table 4.

Table 4: Performance of different algorithms in the training phase.

Algorithm	RFR	LGBM	HBBRT
wMAPE [%]	6.78	6.69	6.58

The table suggests that the performance of the proposed algorithms is similar in a training period. The assumption is confirmed on a visual level, by observing the three graphics. The main conclusion suggests that the base consumption without peaks can be forecasted very well (for example, 23.10.2024). On the other hand, consumption with huge peaks (for example, 15.10.2024) is extremely difficult to predict. This finding could be used in Avantcar's favour by shifting the peak manually and in parallel use it as flexibility. The idea will be expanded in the part for forecasting flexibility in the second part of Deliverable.

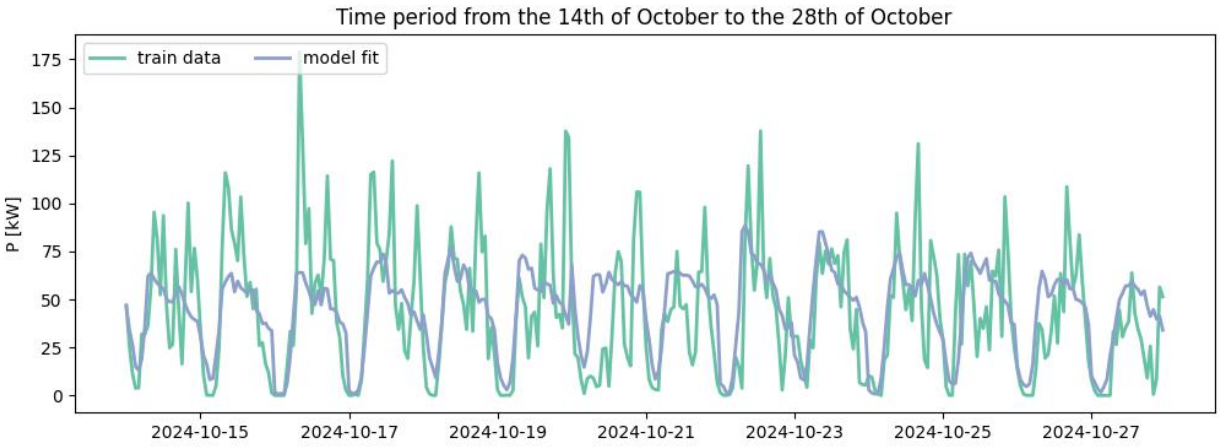

Figure 12: Performance of the RFR model during training.

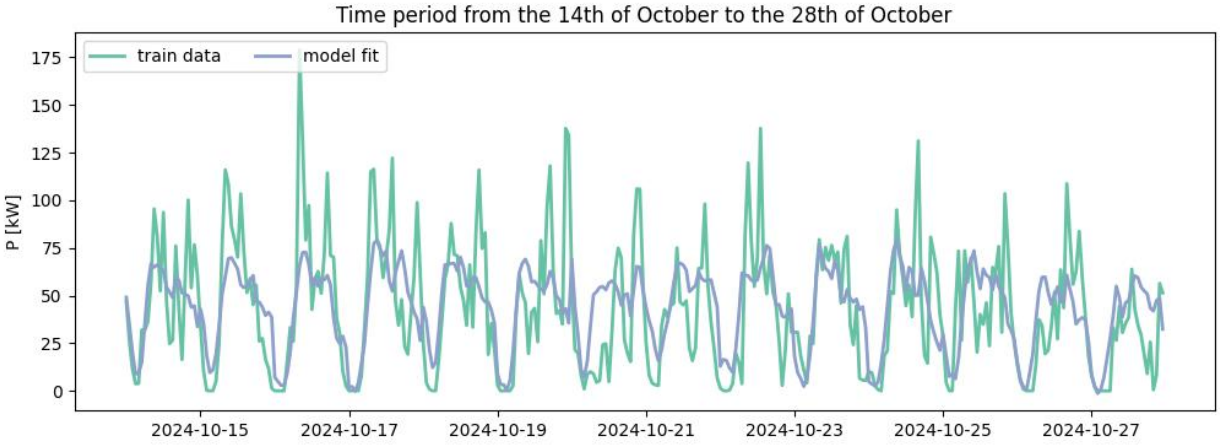

Figure 13: Performance of the LightGBM model during training.

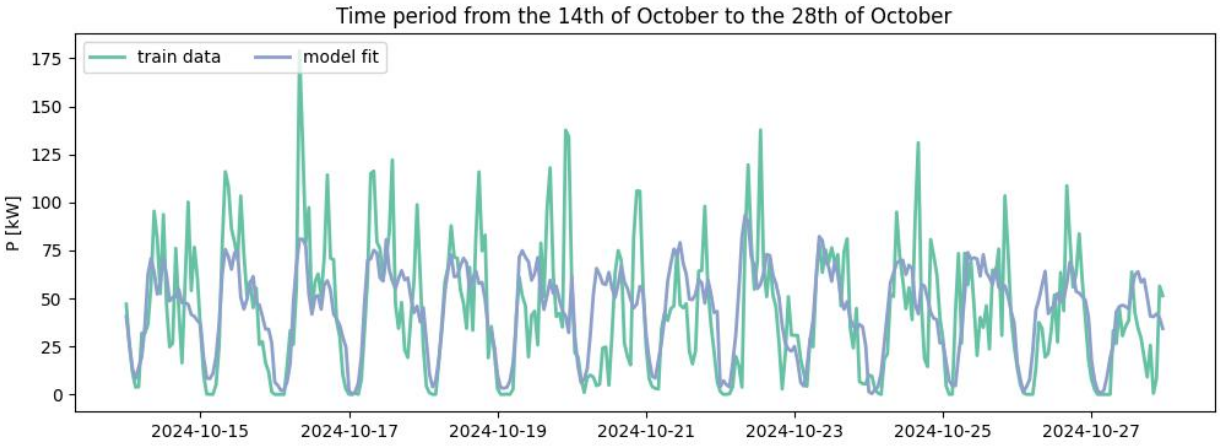

Figure 14: Performance of the HGBRT model during training.

Besides using the wMAPE as the selected metric for the accuracy of the model, we can also evaluate the cost of deviation between the forecast and the actual consumption. This can be written as:

$$c_1(t) = E_{\text{pred}}(t) \cdot p_{\text{DA}}(t) + \left(E_{\text{actual}}(t) - E_{\text{pred}}(t)\right) \cdot p_{\text{imb}}(t),$$

$$c_2(t) = E_{\text{actual}}(t) \cdot p_{\text{DA}}(t),$$

Where the total supplier's cost is $C_1 = \int c_1(t)\ \mathrm{d}t$ while the apparent user's cost is $C_2 = \int c_2(t)\ \mathrm{d}t$, which in our case equals the summation of all c_1^i, c_2^i calculated for each data point. E_{pred} is the prediction of energy consumption, E_{actual} is the actual realization, $p_{\text{DA}}$is the electricity price on the day-ahead (DA) market and p_{imb} is the price for imbalances. We calculate the difference between the supplier's and user's cost $\Delta = C_1 - C_2$, so we can output two ratios:

$$\sigma_C = \frac{\Delta}{C_1}, \qquad \sigma_E = \frac{\Delta}{\sum_i E^i_{\text{actual}}}$$

σ_C is the relative difference in percent and $\sigma_E$is the surcharge on the electricity price based on the inaccuracy of the forecasting. We calculated the proposed costs on a test dataset (data between 1st of November and 1st of December 2024). We evaluated the accuracies of chosen models on the test dataset and compared the results in Table 5.

Table 5: Performance of different algorithms in the test phase.

Algorithm	RFR	LGBM	HGBRT
wMAPE [%]	8.00	6.50	6.76
Consumer's cost [EUR]	4,603.33		
Supplier's cost [EUR]	4,783.04	4,658.94	4,711.83
Cost diff [EUR]	179.71	55.60	108.53
Relative diff [%]	5.84	1.81	3.53
Relative diff [EUR/MWh]	3.90	1.21	2.36

Ultimately, we found that while the choice of algorithm influences the results more than in the training phase, however, the impact remains relatively minor. The wMAPE is achieved with the LGBM model, followed closely by HGBRT, while the RFR performs the worst. LGBM demonstrates strong performance in both prediction accuracy and relative cost differences, with HGBRT yielding similar results. Given the inherent unpredictability of cost variations and the associated deviation costs, LGBM provides a competitive advantage over the other two algorithms.

Figure 15, Figure 16 and Figure 17 illustrate the achieved accuracies on the test dataset. However, the visual representation of algorithm performance offers less insight than the corresponding table. Based on the graphical analysis, the results remain consistent with those observed during the training phase: while the base consumption can be effectively forecasted, peak demand remains challenging to predict.

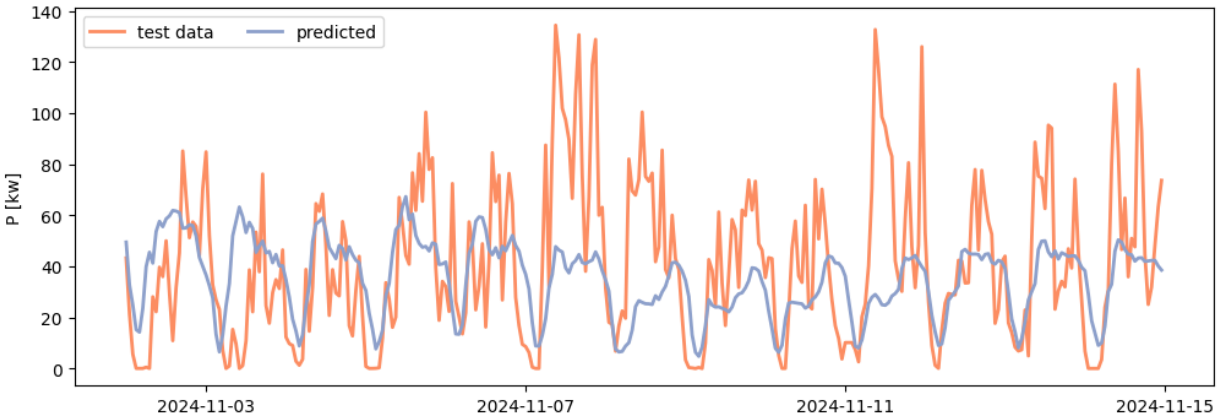

Figure 15: Performance of the RFR algorithm on the test dataset.

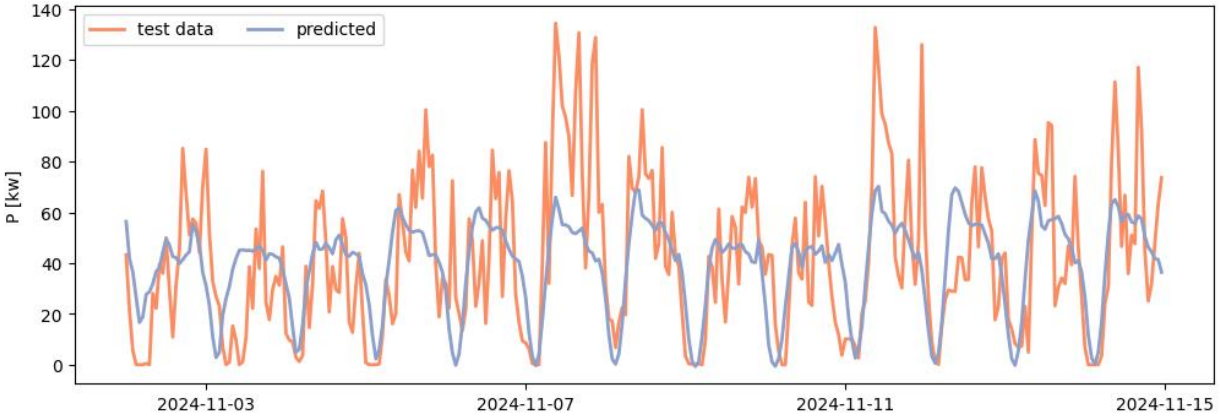

Figure 16: Performance of the LGBM algorithm on the test dataset.

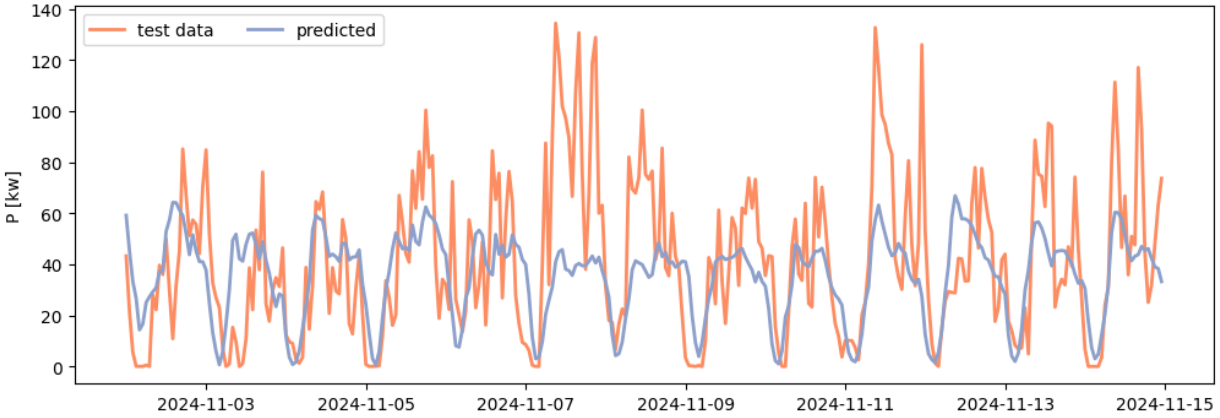

Figure 17: Performance of the HGBRT algorithm on the test dataset.

4.2.2 Individual charging station location level

The EV CS fleet integrated into the KOL aggregation platform spans multiple locations, with each site hosting between 2 and 10 CSs. Even at the aggregated level with 70 CSs, significant variability in charging demand is observed, highlighting the challenge of accurate forecasting—especially when external factors are not clearly correlated with usage. This challenge becomes even more pronounced at individual locations with fewer CSs. Charging patterns can vary widely across sites; some locations display relatively consistent and predictable behaviour, while others exhibit highly irregular usage. To illustrate these differences, we present two weekly charging profiles from locations with 9 and 4 CSs, respectively.

Figure 18 shows the weekly charging profile for a location with 9 CSs. A clear pattern emerges, with peaks occurring around midday. However, the peaks vary — exceeding 30 kW on April 17, 18, and 20 (Wednesday, Thursday, and Saturday), and around 20 kW on Monday, Tuesday, and Friday. There are no charging activities on Saturday. Some charging is also observed in the evening hours, though not consistently each day. This location shows a relatively stable charging pattern, which is key for reliable forecasting.

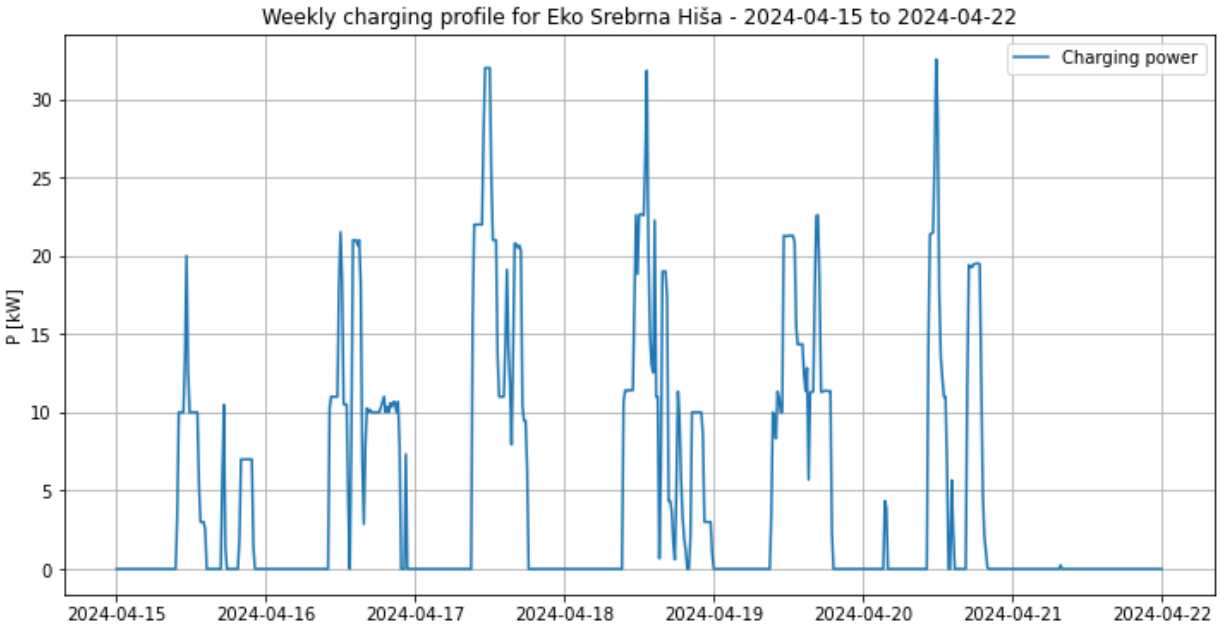

Figure 18: Weekly charging profile for location with 9 CS

On the other hand, several locations with fewer CSs (2 to 6) exhibit less consistent patterns. Figure 19 shows the weekly profile for a location with 4 CSs. The charging patterns vary significantly from day to day — some days have only a few charging events, while others have more frequent activity. Charging sessions also occur at different times of the day and with varying magnitudes, the charging events also being of very short length. This variability is not only evident on a daily basis but also changes from week to week. Such locations pose a significant challenge for accurate forecasting.

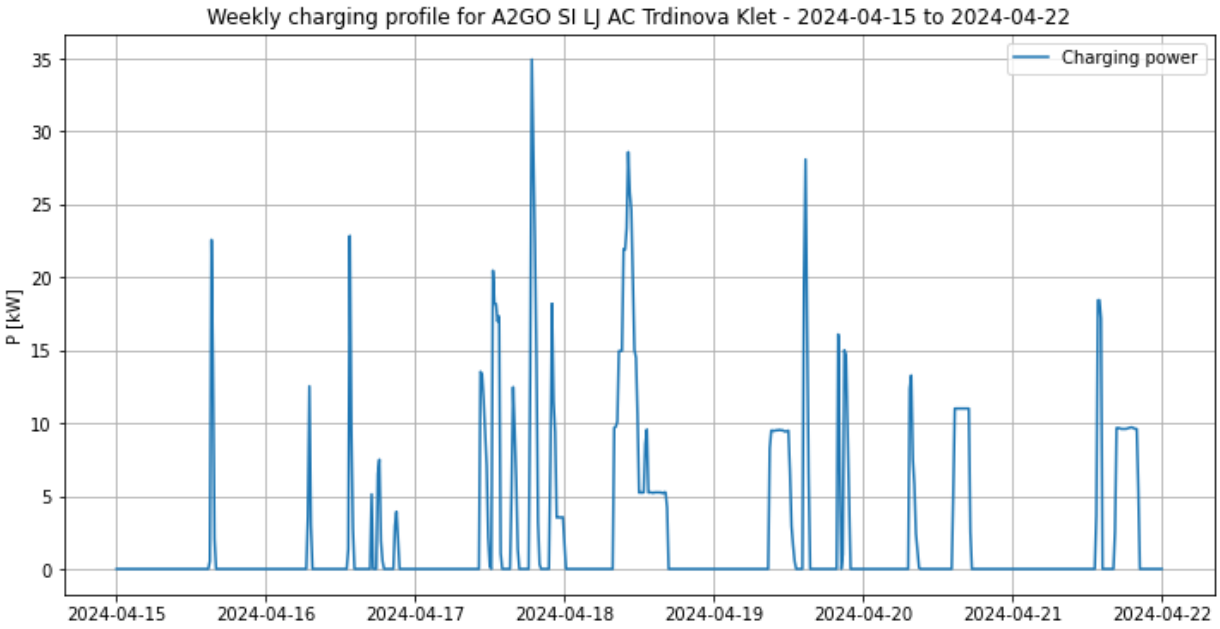

Figure 19: Weekly charging profile for location with 4 CS

Even before starting the forecasting process, it was clear that predicting charging profiles, especially for locations with fewer CSs, would be challenging. However, assessing how accurately we can forecast these profiles are key for determining the feasibility of extracting flexibility from these locations and whether this flexibility could be offered on a local flexibility market.

Baseline forecasts for individual locations were generated using LSTM and XGBoost models, as described in Section 3.3.2. Each location was modelled separately due to differences in the number of CSs and the unique charging patterns observed at each site. As noted previously, the CS data is incomplete, with large gaps in May and June. For this analysis, we used data from January 26, 2024, to May 8, 2024, for model training, while one month of data from July 2024 was used for testing and validation.

The features used for these two models were:

- **Time of day** (ToD): The time of day is a critical feature, as energy usage patterns often follow daily cycles. However, the cyclical nature of time must be accounted for to avoid artificial discontinuities in the model (e.g., the transition from 23:59 to 00:00). To address this, the ToD feature was transformed into sine and cosine components, which allows the model to understand the cyclical nature of time:

$$ToD_{sin} = \sin\left(\frac{2\pi \cdot hour}{24}\right) \quad ; \quad ToD_{cos} = \cos\left(\frac{2\pi \cdot hour}{24}\right)$$

- **Type of day**: Energy consumption differs between workdays and non-workdays (weekends and holidays). This binary feature helps the model account for variations in user behavior and device usage on different types of days.
- **Lag feature (24h)**: For the smart meter baseline forecast, a lag feature was introduced to capture the consumption patterns from the same hour of the previous day, enabling the models to forecast the charging profile for the following day.

$$Lag_{24h} = Consumption(t - 24h)$$

Two key metrics selected for evaluating the EV charging forecasting results are NMAE (Normalized Mean Absolute Error) and R^2 (R-squared). These metrics were chosen due to the unique challenges posed by the data (different numbers of CS per location, high frequency of zero values in the dataset).

NMAE provides a scale-independent measure of error by normalizing the mean absolute error by the average consumption, allowing fair comparison across stations with different power levels. It is calculated as:

$$NMAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n \times \bar{y}}$$

Where n is the number of observations, y_i is the actual value and $\hat{y}_i$ is the predicted value and $\bar{y}$ is the mean of the observed values

R^2 (R-squared) indicates how well the model explains the variability of the target variable. R^2 value closer to 1 signifies that a large proportion of the variance in the observed data is captured by the model. It is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

NMAE was chosen over metrics like nMAPE due to its robustness against zero values in EV charging data, which often distort evaluations. Similarly, RMSE, while commonly used, was avoided due to its scale dependence, making it less suitable for comparing locations with significantly different maximum power capacities.

The forecasting results for the test dataset are presented in Table 6, showing the evaluation for each individual CS location.

Table 6: Forecasting results for individual EV CS locations

Location name	CS number	LSTM NMAE	LSTM R^2	XGBoost NMAE	XGBoost R^2
AC Trdinova klet	4	1.1710	0.0273	4.5119	-0.0345
Eko Srebrna Hiša	9	1.0003	0.2896	1.0301	0.2010
Cesta Dveh Cesarjev	2	1.8736	-0.0208	2.3419	-0.5928
FP by Sheraton LJ	2	1.8191	0.0133	1.7999	-0.0864
AC Prva Osebna	3	1.4279	0.0031	1.8357	-0.0974
AC Tehnološki Park	4	1.8741	-0.0397	2.0958	-0.3096
BTC Decathlon	4	1.7301	-0.0137	1.8930	-0.1380

BTC Kristalna Palača	6	1.4094	0.0045	1.6617	-0.0827
BTC Street Food	6	1.5414	0.0281	1.6861	-0.1555
BTC Trg Mladih	6	1.4105	0.0571	1.3879	-0.1085
GH Kongresni Trg	6	1.1227	0.0524	1.2512	-0.1998
Supernova Rudnik	4	1.4792	0.0433	1.5870	-0.1147
LJAPT	10	0.9417	0.1261	1.0336	-0.0853
Aggregated	70	0.5644	0.2449	0.5654	0.1995

Locations with a higher number of CSs, such as Eko Srebrna Hiša (9 CSs) and LJAPT (10 CSs), demonstrate relatively good forecasting performance, with NMAE values close to 1.0 and positive R^2 scores. These results suggest that the models, particularly LSTM, were able to capture the overall charging patterns despite day-to-day variations. For instance, Eko Srebrna Hiša achieved an NMAE of 1.0003 with LSTM and 1.0301 with XGBoost, while LJAPT recorded NMAE values of 0.9417 and 1.0336, respectively. Their positive R^2 values indicate that the models explained some of the variance in the charging data, but still not being great

Figure 20 for Eko Srebrna Hiša illustrates this point, showing that both LSTM and XGBoost models capture general charging trends, especially during peak periods around midday. However, discrepancies remain, particularly in predicting the magnitude of peaks and sudden charging events. XGBoost tends to overestimate peak values, while LSTM provides smoother predictions but sometimes underestimates rapid fluctuations. This only applies to weekday, as there is much less activity during the weekend.

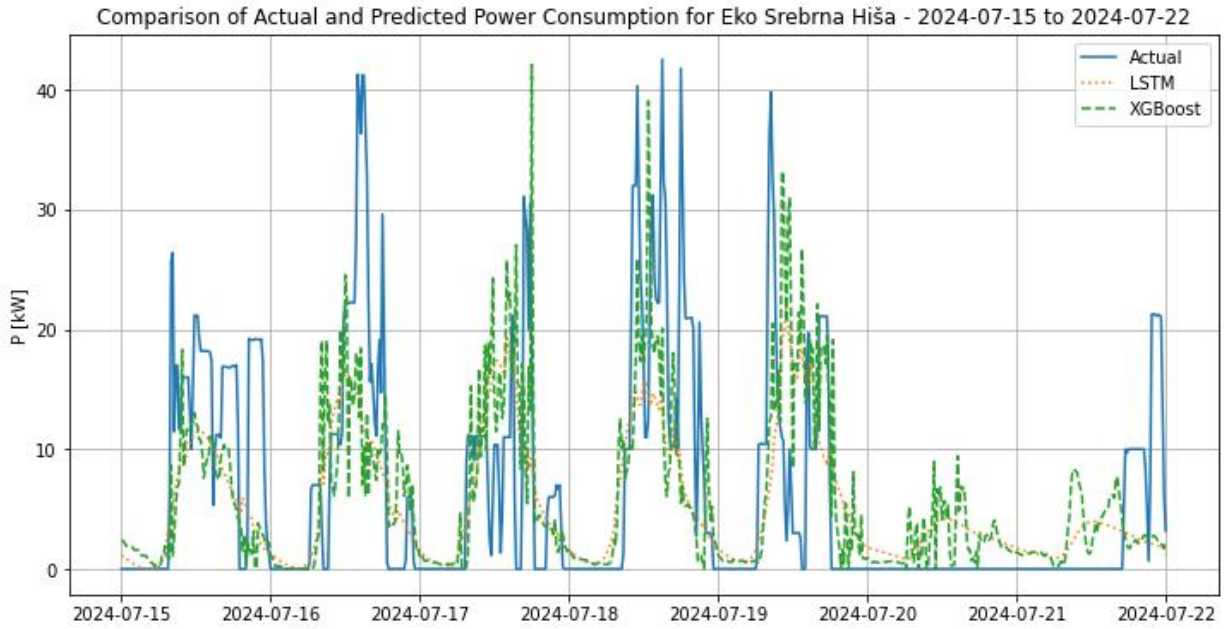

Figure 20: Forecasting results for location with 9 CS

On the other hand, locations with fewer CSs, such as Cesta Dveh Cesarjev (2 CSs) and AC Trdinova klet (4 CSs), display significantly higher NMAE values and negative R^2 scores, particularly with XGBoost. This suggests that the high variability and randomness of charging sessions at these smaller locations pose significant challenges for accurate forecasting which resulted in these bad forecast results.

This is clearly illustrated in below, where both models fail to capture any charging events in their forecasts. The charging activities occur sporadically at different times, are often short in duration, and some days have no charging activity at all.

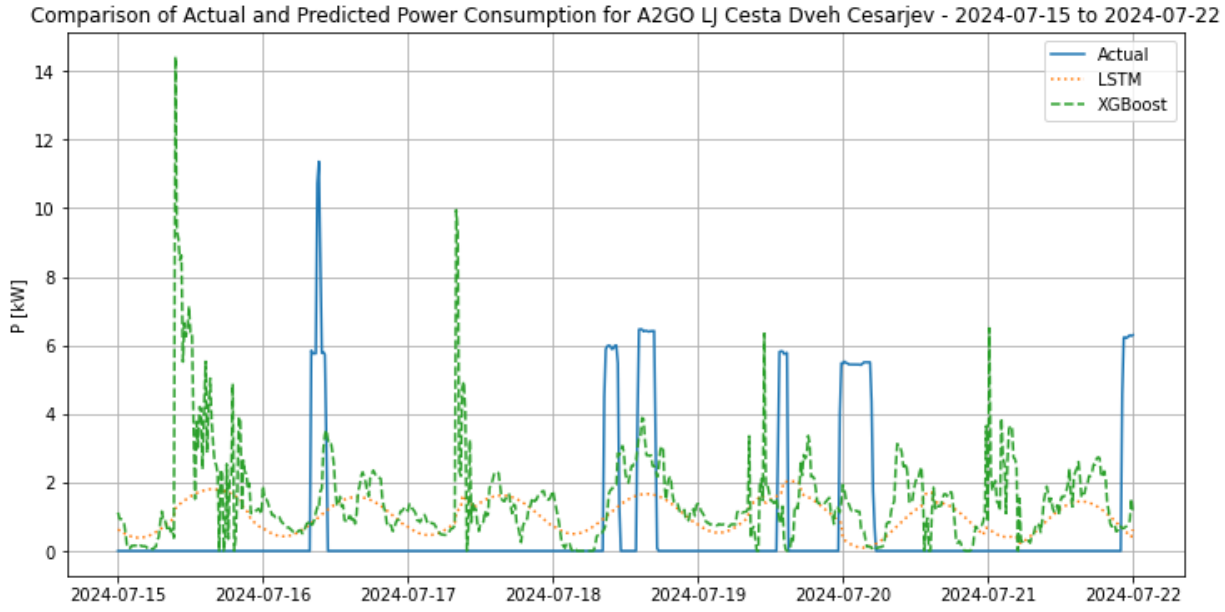

Figure 21: Forecasting results for location with 2 CS

In terms of model comparison, LSTM generally outperforms XGBoost across most locations, achieving lower NMAE values and less negative R^2 scores. For instance, at BTC Kristalna Palača (6 CSs), LSTM's NMAE is 1.4094 with a near-zero R^2 (0.0045), while XGBoost produces higher NMAE and negative R^2 values, indicating that LSTM's ability to capture temporal dependencies provides an slight advantage in modelling time-series data with moderate variability.

When compared to the aggregated fleet of 72 CSs, which achieved the lowest NMAE values (0.5644 for LSTM and 0.5654 for XGBoost) and positive R^2 scores, it becomes evident that increasing the number of CSs reduces variability. The aggregated fleet results serve as a benchmark, demonstrating that pooling data from multiple stations smooths out randomness and enhances forecasting performance. While Eko Srebrna Hiša and LJAPT show relatively good accuracy for individual locations, their NMAE values remain higher than the aggregated fleet, reflecting the benefits of averaging multiple charging behaviours across a larger portfolio.

These findings have important implications for the viability of such assets in local flexibility markets. For an aggregator to effectively participate, reliable predictions of charging patterns are key to determine the viability of a location. Locations with a larger number of CSs are more likely to yield more reliable forecasts and are thus better suited for flexibility services. Moreover, the potential for flexibility will depend not only on the inherent predictability of the charging demand but also on the specific design of local market products (such as peak shaving, which is more suited to the operation of EV fleets) and regulations.

4.3 Amibit HEMS fleet

The HEMS fleet poses unique challenges for baseline forecasting due to the diverse range of connected devices—some known, others unknown—and the variability in their energy usage. While fleet-level forecasting might yield better immediate results, it often lacks the granularity required for meaningful flexibility assessments. To address this, we opted to focus on the household level, forecasting both the overall load curve as observed through the smart meter and modelling the specific connected devices, such as HPs and PV systems.

Certain asset types have been excluded from the baseline forecasting. For example, BESS were not considered because their operation is typically driven by explicit instructions or schedules rather than intrinsic consumption patterns. Similarly, although some HEMS units include EV CSs, we did not attempt to forecast their baseline operation. EV charging behaviour is highly variable and user-dependent—factors like driving schedules and individual charging preferences introduce significant unpredictability, making it extremely challenging to develop a reliable baseline using the available data.

For the forecasting process, we employed LSTM and XGBoost models implemented using Python libraries (TensorFlow [18], Keras [18], and XGBoost [19]). The models were trained using a combination of historical power consumption or production data and relevant external influencing factors. The key features used in the training process include:

- **Time of day** (ToD): The time of day is a critical feature, as energy usage patterns often follow daily cycles. However, the cyclical nature of time must be accounted for to avoid artificial discontinuities in the model (e.g., the transition from 23:59 to 00:00). To address this, the ToD feature was transformed into sine and cosine components, which allows the model to understand the cyclical nature of time:

$$ToD_{sin} = \sin\left(\frac{2\pi \cdot hour}{24}\right)$$

$$ToD_{cos} = \cos\left(\frac{2\pi \cdot hour}{24}\right)$$

- **Type of day**: Energy consumption differs between workdays and non-workdays (weekends and holidays). This binary feature helps the model account for variations in user behavior and device usage on different types of days.
- **Lag feature (24h)**: For the smart meter baseline forecast, a lag feature was introduced to capture the consumption patterns from the same hour of the previous day:

$$Lag_{24h} = Consumption(t - 24h)$$

- **Ambient temperature** (°C) and **Global solar irradiance** ($^{W}/_{m^2}$): These external variables, sourced from ARSO Meteo [20], influence device operation (e.g., affecting HP demand and PV production). The original weather data (recorded at 10-minute intervals) was resampled to match the 15-minute resolution of the power data.

It has to be noted that due to the anonymization of data received from Reduxi HEMS systems, precise location information was not available. As a result, we assumed uniform weather conditions

(temperature and global irradiance) across all HEMS devices, even though weather can vary across different regions of Slovenia. This simplification may affect the forecast accuracy for some units.

For each individual HEMS unit—and its associated assets—the forecasting models were trained using only its historical data. This device-specific approach offers several advantages:

- It allows for device-specific forecasts without requiring detailed information on connected assets.
- It eliminates the need for detailed metadata (e.g., device-rated power, orientation) that may not always be available (as in our case).
- The forecasting models can be directly deployed to other similar HEMS devices.

However, this approach assumes that sufficient historical data is available for each household. To address scenarios where individual historical data is sparse or unavailable, we also developed a generic forecasting model. This alternative approach leverages training data pooled from multiple HEMS units and incorporates an additional feature—the rated power of the asset (PV or HP). Rated power was estimated based on the maximum value observed in the timeseries data, adjusted by a scaling factor and rounded. This generic model allows us to forecast for any new device by adjusting for its rated power and applying location-specific weather forecasts.

The primary accuracy metric used for evaluating our forecasting models is the normalized Mean Absolute Percentage Error (nMAPE), defined as:

$$nMAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{\max(y) - \min(y)}$$

Training and test data for these models were obtained from March to August, with a standard 80-20 train-test split. The device-specific models use only the historical data from the corresponding HEMS unit, while the generic model aggregates data across multiple units to account for variability when individual device data is insufficient.

4.3.1 Household baseline consumption forecast

The forecasting results for household-level consumption are shown in two figures. The first figure provides a view of the forecasts over the entire test period, while the second figure focuses on a single week, allowing for a closer examination of model performance.

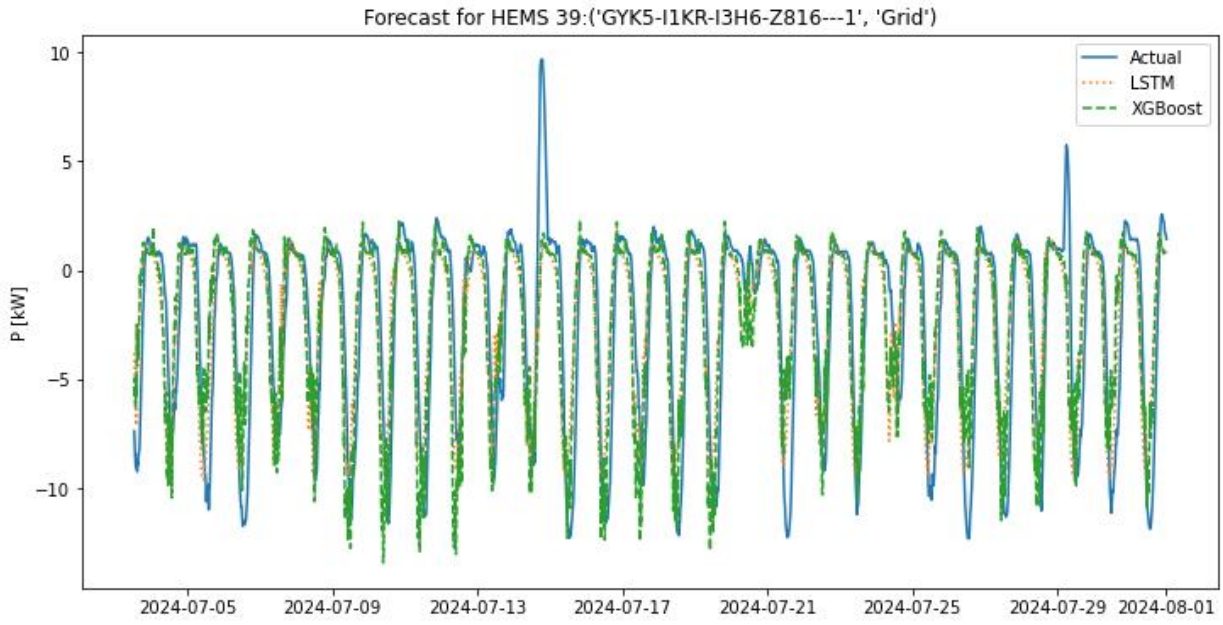

Figure 22: Household baseline forecast results for test dataset, HEMS 39

From Figure 22, it can be seen that both the LSTM and XGBoost models effectively capture the general patterns of household consumption over time. The periodic nature of consumption, driven by daily and weekly usage patterns, is well-represented in both models. However, XGBoost appears to have a slight advantage in capturing peak values and responding to sharp fluctuations, particularly during high-consumption periods. On the other hand, LSTM exhibits a smoother response, which, while stable, occasionally results in missed sharp changes.

However, both models tend to underestimate negative load (production) on certain days during the daytime—a discrepancy that may be attributed to the simplified weather data used (i.e., the assumption of uniform weather conditions across all HEMS devices). Additionally, the models fail to accurately capture two sharp consumption peaks observed on July 15 and July 30.

Table 7 shows the nMAPE values for both models for the household consumption.

Table 7: Household consumption model evaluation

	LSTM	XGBoost
nMAPE [%]	9.50	9.38

For the test period, the nMAPE values are 9.80% for LSTM and 9.38% for XGBoost. These values indicate that XGBoost achieves slightly better overall accuracy than LSTM. The difference in nMAPE aligns with what we saw in the figure, where XGBoost's ability to handle sharp consumption changes gives it a slight edge.

In Figure 23, which focuses on a one-week period, both models perform well in aligning with the actual data during low-consumption (night-time) periods, as well as negative consumption (generation) during the day. However, XGBoost demonstrates slightly superior accuracy in capturing short-term peaks, although it occasionally overestimates during rapid transitions. LSTM, in contrast, struggles slightly more with sharp transitions, tending to smooth out peaks and lows, which can lead to underestimations or delayed responses during quick changes in consumption. Notably, neither model was able to forecast a sudden increase in consumption during the evening hours of September 27.

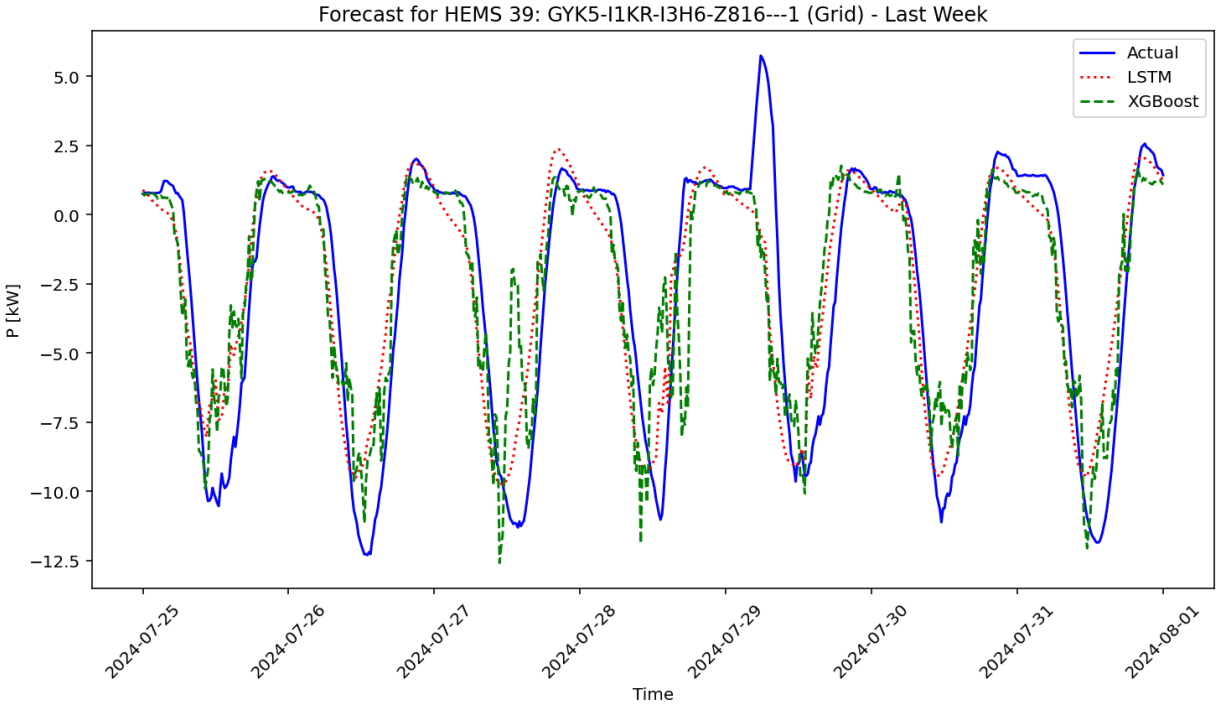

Figure 23: Household load profile forecasting results for HEMS 39 - one week

Overall, the forecast results for household consumption prove to be quite good. Although some improvements could be made—particularly by incorporating more accurate, location-specific weather data—the current accuracy is promising. This level of performance is highly encouraging for further applications, whether for validating flexibility potential (where we expect similar or even better accuracy when forecasting for an aggregated portfolio of multiple HEMS units) or for use in subsequent optimization and planning stages.

4.3.2 PV baseline forecast

Figure 24 shows the actual PV production alongside the predictions of LSTM and XGBoost models for the entire test period. Both models successfully replicate the PV production patterns overall, including the consistent peaks during midday and the drop to zero production during nighttime, but struggle a bit with reaching the peak values sometimes. However, there are some differences in performance: on certain days (e.g., July 20), both models tend to over-predict the peak values, and the LSTM model shows a tendency to begin increasing its forecast before the actual PV production begins.

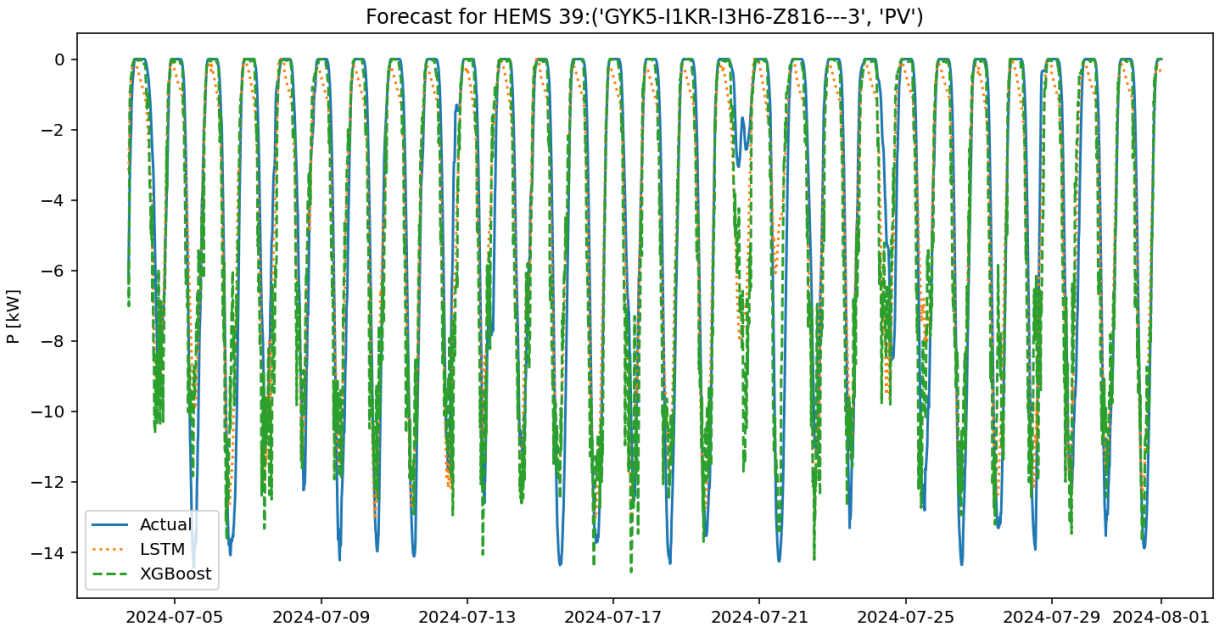

Figure 24: PV baseline forecast results for test dataset, HEMS 39

The evaluation metric nMAPE (presented in Table 8) shows that LSTM achieves an nMAPE of 9.18, indicating a slightly better alignment with the actual PV production values compared to XGBoost's nMAPE of 10.06.

Table 8: PV model evaluation

	LSTM	XGBoost
nMAPE [%]	9.18	10.06

Figure 25 shows the actual PV production alongside predictions generated by the LSTM and XGBoost models for the last week of the test period. Both models successfully follow the daily production patterns, including peak production during midday and zero values at night.

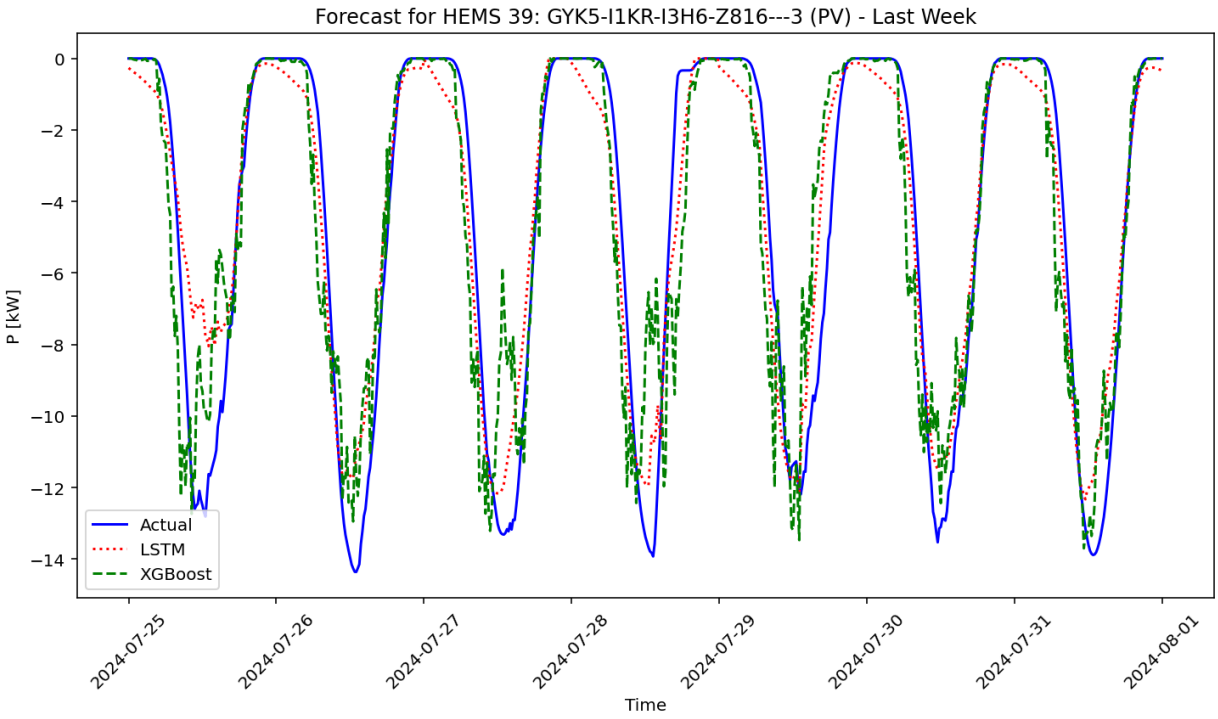

Figure 25: PV generation forecasting results for HEMS 39 - one week

The LSTM model demonstrates a slightly better fit to the actual data, particularly during the transitional periods of sunrise and sunset, as well as at the peaks of daily production. XGBoost, while reasonably accurate, occasionally overestimates the peaks and introduces oscillations during these transitions. Notably, XGBoost handles the nighttime period very well by accurately predicting zero production, whereas the LSTM model sometimes outputs non-zero values during these periods, likely due to its smoothing effect and difficulty handling abrupt changes.

Similarly to the household consumption forecasts, the PV production forecasts are very promising. In our analysis, the LSTM model provided slightly better results than XGBoost. However, performance could be further improved by combining the strengths of both methods or by explicitly post-processing to enforce a zero output during nighttime. This level of forecast reliability enables the aggregator to confidently use PV systems as a dependable source of downward flexibility, by curtailing production when necessary to support grid stability and optimize market participation.

4.3.2.1Generic approach

A generic forecasting model for PV production was developed using data collected from multiple PV units with varying rated powers and production patterns. This model was then validated using data from seven different PV units—including one that was excluded from the training dataset to ensure an unbiased evaluation. The validation was conducted on the test set for each channel time series using both LSTM and XGBoost models.

Table 9 presents the nMAPE values for each of the seven PV units:

Table 9: Generic PV model evaluation results

HEMS channel	Rated power [kW]	LSTM - nMAPE [%]	XGBoost - nMAPE [%]
HEMS 39	20	13.29	14.77
HEMS 9	9	14.39	15.08
HEMS 16	15	19.29	15.68
HEMS 36	10	14.28	13.21
HEMS 22	16	9.72	10.00
HEMS 27	16	18.08	18.85
HEMS 21	111	15.89	15.72

The PV units used for validation range in rated power from 9 kW to 111 kW, providing a diverse set of systems that help evaluate model performance across different scales.

The nMAPE values show considerable variability, with LSTM models achieving error between 9.72% and 19.29%, while XGBoost models range from 10.00% to 18.85%. The best performance was observed for HEMS 22, with a rated power of 16 kW, where LSTM achieved a nMAPE of 9.72%. This performance is comparable to the model presented in the previous section, which was trained using only the historical data of that unit.

On average, LSTM models achieved an accuracy of approximately 14.99%, while XGBoost models averaged slightly better at 14.76%. In four out of the seven cases, LSTM outperformed XGBoost, suggesting its greater ability to generalize for smaller or medium-sized PV units. However, for HEMS 16 (rated power of 15 kW), XGBoost performed notably better, achieving a nMAPE of 15.68% compared to LSTM's 19.29%.

For HEMS 39, the generic model resulted in nMAPE values of 13.29% for LSTM and 14.77% for XGBoost, which are both higher than the results obtained with the model specifically trained for that unit. The visual representation of the forecasting results for this unit (for one week) is provided in Figure 26:

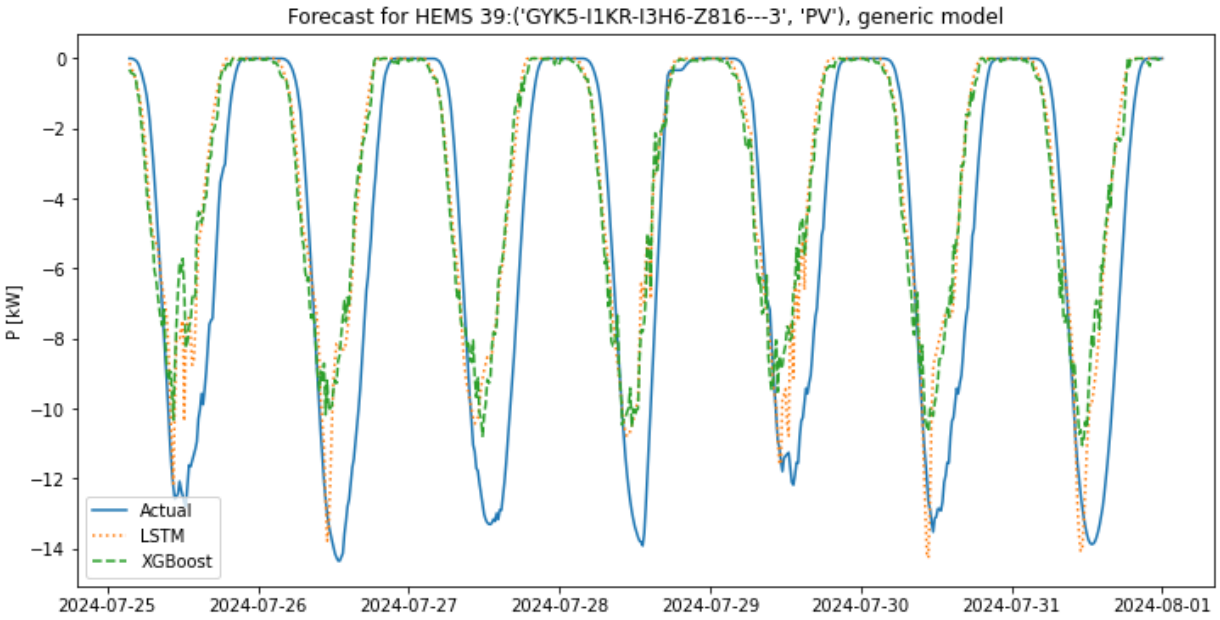

Figure 26: PV generation forecasting results for HEMS 39 - one week, generic model

From the figure, it can be observed that both models, particularly XGBoost, tend to underestimate PV production during peak periods, with deviations of approximately 20–40%. This underestimation is most pronounced in the latter part of the day when the models predict a decline in production earlier than observed in the actual data. A likely cause for this discrepancy is an underestimated rated power, which was calculated based on available timeseries data. Additionally, the use of generalized weather data for all units, rather than location-specific forecasts, may contribute to the observed forecasting errors.

Despite these limitations, the models performed reasonably well, demonstrating the potential of this approach for forecasting PV production without access to historical data for individual units. Further improvements could be achieved by refining the rated power estimation process and integrating more precise weather data.

4.3.3 HP baseline forecast

The baseline forecasting of HP operation primarily depends on external temperature, which influences the HP's operation by determining the heating or cooling demand required to maintain indoor comfort. The performance of the developed forecasting models was evaluated using test data for a selected HP unit, with the results presented below.

Figure 27 shows the forecasted load curve for the entire test dataset. The models generally capture the trends observed in the actual data, including the increase in power consumption between the 9th and 19th of July. During this period, both models follow the upward trend; however, XGBoost consistently overestimates the peak power values, particularly during high-demand periods.

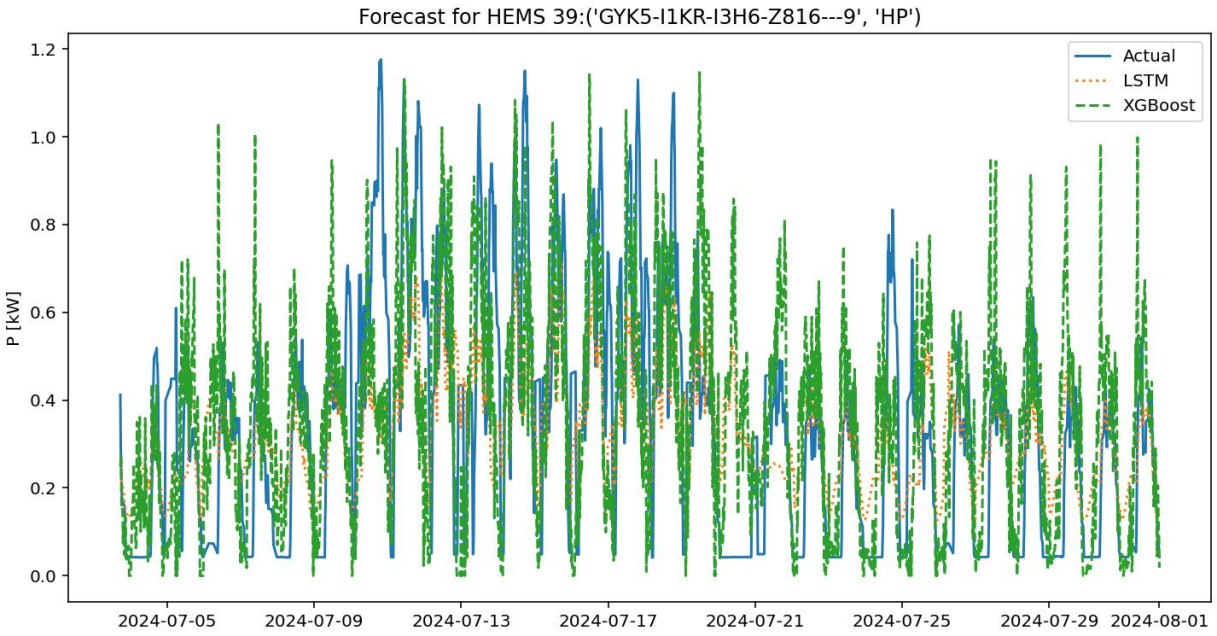

Figure 27: HP baseline forecast results for test dataset, HEMS 39

The performance of the two models is quantified using the nMAPE metric, as shown in Table 10:

Table 10: HP model evaluation

	LSTM	XGBoost
nMAPE [%]	14.98	16.27

The results indicate that the LSTM model achieves better overall accuracy, with an nMAPE of 14.98%, compared to 16.27% for XGBoost, which suggests that LSTM is more effective at capturing the underlying consumption patterns of the HP.

Figure 28 provides a more detailed view, focusing on one week of test data. This closer examination reveals several key observations:

- **Daytime Behavior**: During daytime hours, when HP activity increases, both models accurately follow the observed patterns. However, XGBoost tends to overestimate power consumption peaks, which is consistent with the trend seen in the full-period results.
- **Nighttime Behavior**: During nighttime hours, when the HP is mostly inactive, the LSTM model occasionally predicts higher-than-actual power values, resulting in small overestimations. Despite these discrepancies, LSTM maintains better overall accuracy.

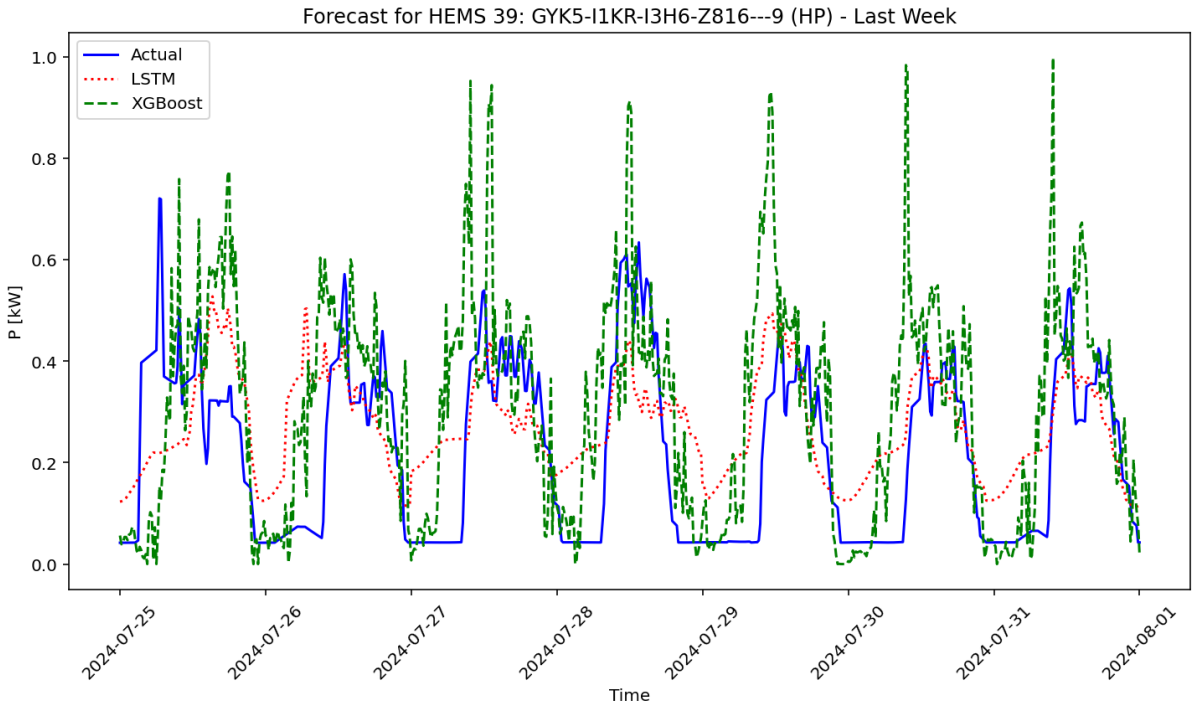

Figure 28: HP consumption forecasting results for HEMS 39 - one week

Although the accuracy of forecasting HP consumption is lower than that achieved for PV generation or aggregated household consumption, it still delivers a valuable baseline—especially when using LSTM, which could be further enhanced through additional post-processing or other methods. This baseline enables the aggregator to effectively calculate the device's flexibility potential.

4.3.3.1 Generic approach

To develop a generic forecasting model for HPs, we pooled three separate HP channels into a single training dataset, with each channel keeping its own time-series structure and corresponding features. Since only three usable HP channels were available, the total quantity of training data was notably smaller compared to what was used in the generic PV model.

As before with generic approach for PV forecasting, rated power served as an additional input feature. However, due to limited data for extremely low temperature periods, we estimated each HP's maximum rated power by taking its highest observed consumption and applying a 0.7 factor. Furthermore, we lacked information on the specific HP type—such as air-source versus ground-source—which can exhibit distinct operational patterns, making this a further simplification. Both LSTM and XGBoost models were then trained on this combined dataset, learning to generalize across different HP behaviors.

Results for the test period are shown in Table 11.

Table 11: Generic HP model evaluation results

HEMS channel	Rated power [kW]	LSTM - nMAPE [%]	XGBoost - nMAPE [%]
HEMS 22	5	26.22	14.10

HEMS 39	6	22.86	15.97
HEMS 21	26	15.21	14.07

Despite the smaller training sample, XGBoost showed relatively stable performance across all three channels, consistently outperforming LSTM in terms of nMAPE. Conversely, LSTM's accuracy declined significantly—particularly for HEMS 22 (26.22%) and HEMS 39 (22.86%)—when compared to the results obtained from a dedicated single-asset model (around 15%).

Furthermore, HEMS 39 displayed a notable difference between its dedicated and generic forecasting performance. Under the generic approach, LSTM accuracy dropped from 14.98% to 22.86%, whereas XGBoost actually improved slightly—from 16.27% (dedicated) to 15.97% (generic).

Figure 29 shows the forecasting results for one week using both generic models. We can see that LSTM model behaves in a similar as with the dedicated model – overestimating the consumption during night-time, but during the peaks it captures them quite closely. XGBoost on the other hand most days overestimates the consumption.

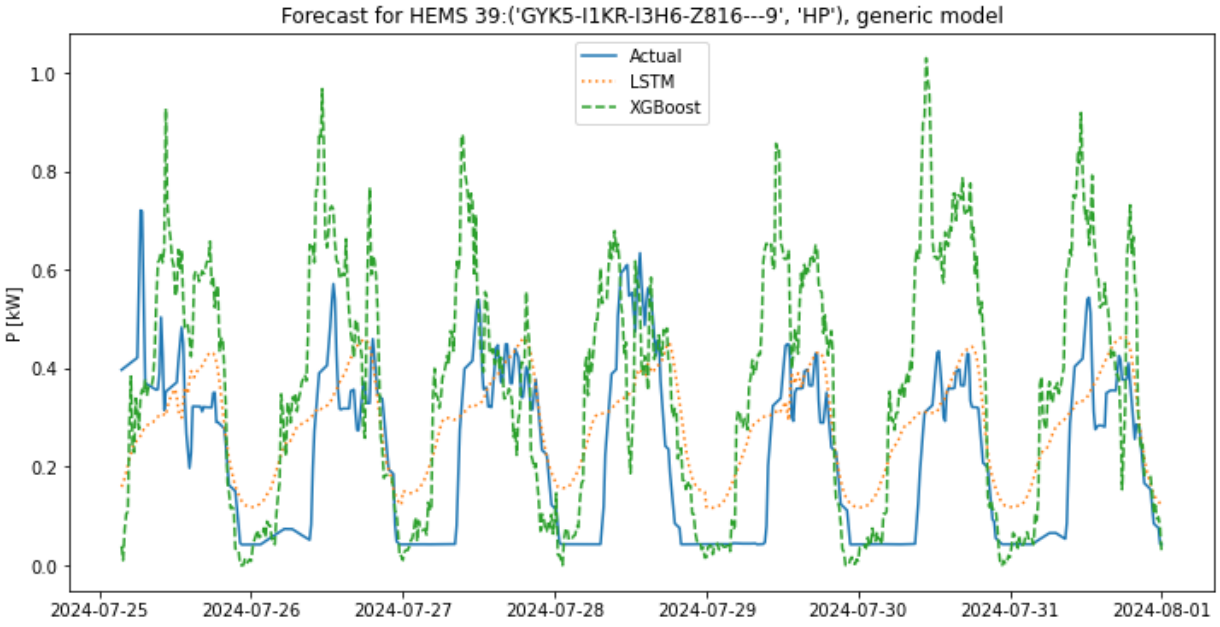

Figure 29: consumption forecasting results for HEMS 39 - one week, generic approach

Interestingly, a closer examination of HEMS 39's final week revealed that LSTM produced a lower nMAPE (16.28%) than XGBoost (27.24%), underscoring how model performance can vary significantly based on the specific timeframe or data segment. This illustrates the importance of robust testing over multiple periods when selecting a model. Compared to the dedicated model, the forecasts from the generic model are somewhat less accurate for HPs, suggesting that a dedicated modelling approach may be more appropriate for HP forecasting. Nevertheless, the model's performance is quite impressive given all the simplifications. It would be very interesting to see how performance improves if we had access to a larger dataset that includes more HP assets with detailed type information and precise location data for more accurate weather inputs.

5 Flexibility Estimation Definition

Following the baseline forecast modelling presented in the previous chapter, this section focuses on flexibility estimation, which builds directly on the insights and results obtained from the baseline forecasts. While the baseline forecast provides an estimate of normal operation for the assets, flexibility forecasting goes a step further by quantifying the potential deviations from this baseline that can be utilized for grid services or market participation.

Flexibility forecast modelling aims to estimate the ability of DERs to adjust their consumption or generation in response to external signals, such as price incentives or grid constraints. This involves identifying the range of possible deviations from the baseline operation while respecting the technical and operational constraints of the assets. The primary objective of flexibility forecasting is to enable aggregators to determine the availability and magnitude of flexibility across their portfolio, providing a foundation for optimal utilization in energy markets.

The main considerations flexibility modelling incorporates are the following:

- **Asset constraints**: these includes power limits, ramp-up and ramp-down times, response time and minimum on/off durations, which define the operational limits of each type of asset.
- **User preferences**: these are important to ensure that the provision of flexibility does not impact the user in a negative way. For example, while using HP or other HVAC devices for flexibility we have to ensure that the temperature levels do not cross the temperature comfort levels.
- **External influences**: weather and market conditions can significantly impact both the availability and value of flexibility.

The starting point for flexibility definition is the baseline forecast, which provides a time-series estimate of normal energy consumption or generation under non-flexibility conditions. The baseline serves as a reference against which potential deviations (upward or downward flexibility) are measured. The accuracy of the baseline forecast directly influences the reliability of the flexibility estimation.

5.1 Flexibility estimation for EVs and EV charging stations

This section first outlines a general EV flexibility model and then details how our forecasted baseline can be used to derive the flexibility available from the EV charging fleet.

5.1.1 EV flexibility model

In a general model, each EV is treated similarly to a BESS but with two key differences:

1. An EV is only available for flexibility when it is plugged in (i.e., during the charging period).
2. In the absence of Vehicle-to-Grid (V2G) capabilities, an EV can only draw power from the grid, meaning that flexibility is provided solely by curtailing its charging load.

The EV is described by the following parameters and constraints:

- **Battery capacity** $C_{EV,bat}$ (in kWh)
- Battery's **state of charge (SoC)**, which at any time equals:

$$0 \leq SoC(t) \leq 1$$

- **Charging power limit**, which is limited by:

$$0 \leq P(t) \leq P_{max}$$

Where P_{max} is the maximum charging power available at the CS.

- **Battery energy equation**, in which the evolution of SoC is given by;

$$SoC(t + \Delta t) = SoC(t) + \frac{\eta \cdot P(t) \cdot \Delta t}{C_{EV,bat}}$$

Where η is the charging efficiency and Δt the time interval.

- **Availability window**, in which the EV is available for charging (and therefore flexibility). This window is between the EV's arrival t_{arr} and departure time t_{dep}. Thus, the charging (and any flexibility actions) are limited to:

$$t \in [t_{arr}, t_{dep}]$$

If the EV user provides explicit arrival (tt_{arr}) and departure (t_{dep}) times (via user input or scheduling data), these times are fixed. Otherwise, if they are not provided, they can be modelled as stochastic variables using statistical distributions fitted to historical data. This stochastic modelling allows the aggregator to estimate the probability that an EV is available at any given moment and adjust flexibility bids accordingly.

- **User comfort constraint** – ensuring that the battery has to reach the sufficient SoC by the time of departure:

$$SoC(t_{dep}) \geq SoC_{target}$$

5.1.2 Flexibility based on forecasted baselines

The forecasting approach we used for the baseline operation of the EV fleet provides an expected charging demand for each time interval. However, accurately capturing this demand is challenging due to the unpredictable behaviour of Avantcar's car-sharing users and their varying arrival and departure times. In addition, our preliminary analysis revealed that EV parking durations—especially during daytime hours—can be quite short. As a result, the baseline forecasting approach is not ideally suited for every charging location within the fleet.

However, we can still identify some charging locations that show relatively predictable demand patterns. For these locations, upward flexibility can be achieved by limiting the total charging power across several CSs. In practice, the aggregator would send charging limitation setpoints to the active CSs, thereby capping the overall charging power. The effectiveness of this flexibility provision would be verified by comparing the actual charging data with the calculated baseline; it is crucial that the forecasted baseline closely matches the actual demand to avoid significantly disrupting the charging process.

For example, suppose the aggregator forecasts a charging demand of 15 kW for the period between 10:00 and 11:00. Based on this forecast, the aggregator might decide to offer 3 kW of upward flexibility

by reducing the charging power, so that the charging load is capped at 12 kW during that period. Two scenarios may occur at activation:

- If the forecast overestimates demand, the imposed limitation will not hinder the charging process significantly and may even reduce overall consumption further.
- If the forecast underestimates demand, the limitation could restrict charging excessively, leading to lower-than-required SoC levels, which would negatively impact user comfort.

Therefore, it is essential for the aggregator to forecast demand as accurately as possible and to determine the appropriate magnitude of flexibility to minimize any adverse effects on user comfort.

It is worth noting that the current framework is designed for the existing market structure, which may not fully reflect the operational characteristics of EV fleets. In the future, new market services could be developed that better accommodate these fleets. For instance, a DSO might compensate aggregators for controlled charging reductions during critical periods. As an example, if full-power charging during peak hours requires 50 kW, an 80% cap could reduce charging to 40 kW, with the DSO compensating the aggregator for the curtailed 10 kW.

The utilization of EV CSs for flexibility will be further explored in the next deliverable, analysing their viability for participating in various energy markets—each with its own mechanisms for leveraging the flexibility of the fleet.

5.2 Flexibility estimation for HEMS assets

Building on the baseline forecasts presented in the previous chapter, this section focuses on quantifying flexibility for the assets connected to the HEMS system (such as Amibit's Reduxi). The flexibility of these assets is what ultimately enables aggregators to offer demand-side services in energy markets. Each type of asset—such as PV systems, HPs, and BESS—presents unique opportunities and constraints for flexibility provision.

In the following sub-sections, we will detail:

- **PV Systems**: Their ability to curtail generation by reducing output below the baseline forecast.
- **HPs**: How they can adjust energy consumption within thermal comfort constraints by modulating their operation.
- **BESS**: The bidirectional flexibility they offer, allowing for both charging (consumption) and discharging (injection) based on the state-of-charge and operational limits.

By quantifying the deviations from the established baseline (i.e., the "normal" operating conditions), we derive the flexibility potential for each asset type. This flexibility is then constrained by technical factors (such as rated power and response times) and user preferences (such as maintaining comfort levels in the case of HPs)

5.2.1 Photovoltaic

PV systems play a crucial role in renewable energy generation, contributing significantly to decarbonization and the green transition. However, despite their many benefits, the rapid penetration of PV systems, especially in low-voltage distribution networks, can lead to challenges during periods of high production. These challenges include grid congestion, voltage limit violations, and negative electricity prices during midday hours—outcomes that are far from ideal for electricity producers.

Fortunately, PV systems can adjust their production by modifying inverter settings to limit output power. This process, known as curtailment, represents a valuable source of flexibility, although it may reduce revenue for producers if not managed properly.

The flexibility of a PV system is defined as its capacity to reduce electricity production below its baseline generation under normal operating conditions. In this context, baseline generation is forecasted using the models detailed in Section 4.3. Flexibility is expressed as downward flexibility, where the PV system can reduce its electricity generation from the baseline to zero, constrained by the technical limits of the PV system.

The flexibility limits can be mathematically expressed as:

$$0 \leq flex_{PV,down}(t_i) \leq P_{PV,base}(t_i)$$

Where:

- $flex_{PV,down}(t_i)$: downward flexibility value at timestep t_i, representing the amount of curtailed power
- $P_{PV,base}(t_i)$: baseline power generation of the PV system at timestep t_i.

5.2.2 Heat pumps

HPs are widely used in households for heating and cooling, and their adoption has grown steadily in recent years as part of the green transition. This shift away from heating systems reliant on gas, oil, or biomass toward electricity-based solutions has highlighted HPs' potential for providing flexibility (especially during colder months). Flexibility arises from their ability to adjust energy consumption within thermal comfort and operational constraints, enabling households to reduce electricity costs or participate in flexibility markets.

The flexibility of a HP is quantified as the deviation from its baseline power consumption, which represents the expected operation under normal conditions. Flexibility can be expressed as:

- **Downward Flexibility**: The ability to increase power consumption, constrained by the HP's maximum rated power and the upper comfort temperature limit.
- **Upward Flexibility**: The ability to decrease power consumption, including temporarily switching off the HP, provided the indoor temperature does not drop below the lower comfort level.

The flexibility limits for a HP at a given timestep are defined as:

$$0 < flex_{HP,down}(t_i) \leq P_{HP,rated} - P_{HP,base}(t_i)$$

$$0 < flex_{HP,up}(t_i) \leq P_{HP,base}(t_i)$$

Where:

- $flex_{HP,up}$ and $flex_{HP,down}$: upward and downward flexibility values at timestep t_i.
- $P_{HP,rated}$: HP rated power.
- $P_{base}(t_i)$: baseline power consumption of the HP at timestep t_i

During flexibility provision, it is essential to maintain indoor temperature within the defined thermal comfort range:

$$T_{min} \leq T_{indoor}(t_i) \leq T_{max}$$

Where:

- T_{min} and T_{max}: lower and upper comfort temperature limits for
- $T_{indoor}(t_i)$: indoor temperature at timestep t_i

Since we are focusing exclusively on the heating mode—when the HP operates at higher power during colder months (meaning more flexibility potential) —the relationship between HP operation and indoor temperature can be modelled using the following thermal dynamics equation:

$$T_{indoor}(t_{i+1}) = T_{indoor}(t_i) + \frac{\eta_{HP} \cdot P_{HP}(t) \cdot \Delta t}{C_{thermal}} - \frac{\left(T_{indoor}(t_i) - T_{outdoor}(t_i)\right) \cdot \Delta t}{R_{thermal}\, C_{thermal}}$$

$$T_{indoor}(t_{i+1}) = T_{indoor}(t_i) + \Delta T_{HP} - \Delta T_{loss}$$

Where:

- η_{HP} Coefficient of performance (COP) of the HP, representing its efficiency in converting electrical energy into thermal energy. This value typically ranges from 2.5 to 4.
- $C_{thermal}$: Thermal capacity of the building, representing its ability to store heat, ranging from 8-15 $\left(kWh/_{°C}\right)$ for typical households
- $R_{thermal}$: Thermal resistance of the building, representing its insulation effectiveness, ranging from 2 to 4 $\left(°C/_{kW}\right)$ depending on the insulation quality.
- $T_{outdoor}(t_i)$: Outdoor temperature at timestep t_i

5.2.3 Battery energy storage system

BESS are unique among household flexibility assets due to their ability to provide services in both directions—discharging power (injection) and charging power (consumption). The flexibility of a BESS is calculated at discrete time steps, where the flexibility at each step is determined based on the state of the system in the previous time steps. This ensures that the constraints imposed by the SoC and operational limits are respected.

BESS operate in two different modes (charging and discharging), which are detailed below:

1. **Discharging mode**

   When discharging, the **available energy** is determined by the current SoC. The energy at each timestep is determined by:

$$E^{dis}_{avail}(t_i) = SoC(t_i) \times C_{BESS}$$

Where $SoC(t_i)$ is the state of charge at time t_i and C_{BESS} the total energy capacity of the BESS.

The **maximum power available for discharging** at time t_i is then defined as:

$$P^{dis}_{BESS,max}(t_i) = min\left\{P^{dis}_{BESS,rated}, \frac{E^{dis}_{avail}(t_i)}{\Delta t}\right\}$$

Where $P^{dis}_{BESS,rated}$ is the rated discharging power of the BESS and Δt the duration of the timestep.

Accordingly, the **actual discharging power** $P^{dis}_{BESS}(t_i)$ must satisfy:

$$0 \leq P^{dis}_{BESS}(t_i) \leq P^{dis}_{BESS,max}(t_i)$$

2. **Charging mode**

During charging, the energy capacity available to store additional energy is:

$$E^{ch}_{avail}(t_i) = [1 - SoC(t_i)] \times C_{BESS}$$

The maximum power that can be applied for charging is then given by:

$$P^{ch}_{BESS,max}(t_i) = min\left\{P^{ch}_{BESS,rated}, \frac{E^{ch}_{avail}(t_i)}{\Delta t}\right\}$$

Where $P^{ch}_{BESS,rated}$ is the rated charging power.

Thus, the actual charging power $P^{ch}_{BESS}(t_i)$ is limited by:

$$0 \leq P^{ch}_{BESS}(t_i) \leq P^{ch}_{BESS,max}(t_i)$$

The evolution of the SoC is affected by the mode of operation and is updated at each time step t_{i+1} as follows:

$$SoC(t_{i+1}) = \begin{cases} SoC(t_i) + \dfrac{\mu^{ch}_{BESS} \cdot P^{ch}_{BESS}(t_i) \cdot \Delta t}{C_{BESS}}, & during\ charging \\ SoC(t_i) - \dfrac{P^{dis}_{BESS}(t_i) \cdot \Delta t}{\mu^{dis}_{BESS} \cdot C_{BESS}}, & during\ discharging \end{cases}$$

6 Optimal selection for the determined flexibility

As we transition into the next phase of Task T4.4, this section serves as a bridge to the second part of deliverable (due in M34), focusing on the optimal selection of available flexibility. The term "optimal selection" encompasses two key dimensions that we will explore in greater depth in the coming months. Namely, the optimal selection of the most appropriate market for the flexibility asset to participate in and the optimal selection of flexibility to activate if the opportunity/activation in the selected market arises. The following paragraphs in this section establish the foundation for that work, outlining the two perspectives on optimal selection and how they shape the aggregator's decision-making process.

1. **Optimal selection based on market applicability and technical suitability**

    - This dimension of optimal selection refers to the identification of the market/s, where the participation of the flexibility asset is most lucrative economically while considering all technical and operational limitations of the flexibility assets and technical boundaries of the market. Flexibility assets, such as EV chargers and vehicles, as well as devices aggregated under HEMS controllers, must meet technical requirements to be considered viable in different flexibility markets. Beyond technical feasibility, the goal is to determine how these assets can generate the highest financial returns by selecting the most profitable participation strategy.
    - This process requires analysing the forecasted baseline and flexibility, availability of the assets, identifying patterns in their operation, considering technical limitations of the asset and market and ultimately determination of flexibility assets suitability for various markets based on economic gain.

2. **Optimal selection algorithm (disaggregation process)**

    - The second part of optimal selection refers to the aggregator's process of selecting which assets to activate when a flexibility bid is accepted in the case of ancillary services markets or when a market opportunity arises, for example in the ID market for decreasing the portfolio deviation. This is often referred to as disaggregation, in which the aggregator determines which flexibility sources to activate and at what setpoints to fulfil bid requirements.
    - The effectiveness of this selection process is directly tied to the accuracy of the forecasting models and the availability of flexibility at any given time.

Both of these aspects of optimal selection will be thoroughly examined in the upcoming deliverable. However, before diving into those specifics, it is important to first establish a clear understanding of the available flexibility utilization options and their implications for aggregators. The following section provides an overview of these options, outlining the potential markets where flexibility can be monetized, the requirements for participation, and the key challenges associated with each.

6.1 Flexibility utilization options

Flexibility services can be provided in various market settings, each with its own technical, regulatory, and economic requirements. Aggregators or FSPs must carefully assess the characteristics of the flexibility assets they manage to determine where they can be most effectively deployed. This process involves two steps:

1. **Identifying technical feasibility of flexibility assets**: Understanding which markets have demand for the flexibility services that the aggregator's or FSP's assets can provide, considering aspects like response time, bid size, availability constraints, etc.
2. **Determining the most profitable use of flexibility**: Beyond technical suitability, aggregators must analyse where flexibility will yield the highest financial return based on price signals, expected activations, and market structure.

The following sub-sections provide an overview of the primary flexibility utilization options available to aggregators. These include TSO-level ancillary services for frequency regulation, DSO-operated local flexibility markets and optimization strategies in the day-ahead and intraday markets, as well as network tariff-based flexibility.

At this stage, our analysis remains theoretical, setting the stage for the next phase of the task, where a detailed assessment of flexibility participation will be conducted for both EV CS fleet (Avantcar fleet) and HEMS (Amibit Reduxi fleet).

The following sub-sections will now present each flexibility utilization option in detail, examining their market mechanisms, participation requirements, and potential role in the OPENTUNITY project.

6.1.1 TSO AS market

The TSO ancillary services market is a specialized segment of the energy market where TSOs procure services to maintain the stability, reliability, and efficiency of the electricity grid. Ancillary services are crucial for balancing grid frequency and addressing fluctuations in electricity supply and demand.

Avantcar's EV charging points and Amibit's HEMS device portfolios could offer flexibility in the ancillary services market for two distinct products: Automatic Frequency Restoration Reserve (aFRR) and Manual Frequency Restoration Reserve (mFRR). Each product involves two types of bids: capacity and energy. While both portfolios could participate in these products, energy bids are more suitable because they allow flexibility to be offered during specific hours of the day when it is available. In contrast, capacity bids require flexibility to be available during the product slot if called upon. The main difference is that the flexibility provider gets paid for the capacity bid even if not activated, while in the case of the energy bid, monetization is gained only in the case of the activation.

In the Slovenian ancillary services market, the minimum bid for both products is 1 MW, which poses a significant limitation for the OPENTUNITY project portfolios since their available flexibility power is currently far below this threshold. To meet the minimum bid requirement, several assets must be aggregated into a combined portfolio.

At present, individual participation of Avantcar's CSs or Amibit's HEMS devices in the ancillary services market is not foreseen due to their limited flexibility. However, combining the several assets in the aggregator's portfolio increases the likelihood of participation. Looking ahead, the potential for

ancillary services market participation should be evaluated, as the expected growth in CSs and HEMS devices could make this market an attractive opportunity for both portfolios.

6.1.2 DA/ID market participation – price optimization

The electricity price optimization for a final user (e.g., Avantcar's CSs and Amibit's HEMS devices) can be done by optimal participation in Day-Ahead (DA) and Intraday (ID) markets.

The DA market is a wholesale electricity market where participants trade electricity for delivery during the next day. Bids and offers are submitted based on expected demand and supply, and market clearing determines prices and quantities for each hour of the upcoming day.

The ID market allows participants to trade electricity closer to real-time (in Slovenia one hour before delivery), typically to adjust for changes in demand or supply forecasts after the day-ahead market has closed.

The primary objective is to optimally calculate the operating schedules of energy assets within the portfolio based on predicted DA market prices, with the goal to minimize electricity costs. This optimization leverages the flexibility of individual assets, such as operating CSs, which are typically more flexible during nighttime and partially during daytime hours. Electricity costs can be further reduced in the ID market by balancing differences between actual consumption and the quantities purchased in the DA market, thereby avoiding deviation costs. Additionally, analyzing the power system's position—whether it has a surplus or deficit of power—can provide opportunities for strategic trading in the ID market. For example, if the system is predicted to be short (lacking sufficient power), positive deviations are likely to be priced higher than DA market rates. In such a scenario, instead of selling surplus electricity when actual consumption is lower than the DA market purchases, it could be more profitable to wait and sell the excess in the ID market, where better prices are anticipated.

6.1.3 Tariff optimization

Electricity costs for consumers are made up of three main components: supply costs, network fees (grid tariffs), and other regulatory charges, where grid tariffs specifically cover the cost of using the transmission and distribution network.

As of October 2024, Slovenia has introduced a new network tariff model that replaces the previous fixed capacity charges with a system where power is charged dynamically based on the highest 15-minute average power demand from the previous billing period. These new rules introduce five time-based tariffs, covering both energy consumption and active power demand, which means that users who exceed their predefined limits may face significantly higher costs.

This new tariff model has proven highly controversial, as many consumers fear that their electricity bills could increase, especially for those who are unable to shift their consumption away from peak hours. The shift from a fixed to a dynamic power charge means that users who experience occasional high loads (e.g., EV charging, HPs) may see unpredictable increases in their network fees, leading to concerns about affordability and fairness.

With the introduction of time-based tariffs, flexibility in energy consumption offers a way to reduce costs and manage bill increases. By strategically adjusting consumption patterns, users can optimize their electricity usage and avoid penalties for exceeding predefined power limits. HEMS, such as Amibit's Reduxi device, is able to support these strategies by providing real-time energy monitoring and device control for tariff optimization. By continuously tracking household electricity consumption and managing connected devices, such systems could help users stay within active power limits and shift consumption to lower-cost periods. For instance, if a household is nearing a critical power threshold, a HEMS system could temporarily adjust or delay non-essential loads, such as EV charging or HP operation, to prevent additional costs.

6.1.4 DSO local flexibility market participation

The concept of local flexibility markets has gained significant traction in recent years due to the increasing penetration of DERs in low-voltage distribution networks. The growth in renewable generation, such as PVs, alongside the electrification of heating (mainly HPs) and mobility sectors (EVs), has introduced new challenges for DSOs such as voltage limit violations and network congestion.

Traditionally, DSOs would address these issues through grid infrastructure upgrades, such as replacing transformers or expanding line capacities. However, this approach is costly and time-consuming. As a more cost-efficient and flexible alternative, DSOs are increasingly exploring the use of local flexibility markets to mitigate these grid challenges. This concept is strongly supported by EU energy policy, particularly Article 32 of Directive 2019/944 [1], which emphasizes the role of DSOs in utilizing demand-side flexibility. Furthermore, the new network code on demand response, expected to be finalized by March 2025, will provide a standardized framework for implementing flexibility markets across Europe.

Local flexibility markets enable DSOs to procure flexibility services from prosumers, aggregators or FSPs. These services involve adjusting energy consumption or generation to alleviate local grid issues. Unlike other markets, local flexibility markets focus on grid constraints in specific geographic areas, making the location of flexibility assets a crucial factor. For this reason, the aggregator's portfolio for market bids must include assets located in the defined flexibility zone.

Across Europe, various projects have explored this concept, including X-FLEX [21], STREAM [22], and CoordiNet [23], while commercial solutions like Piclo Flex [24] and NODES [25] demonstrate real-world applications. Yet local flexibility markets remain underutilized due to regulatory uncertainty, limited liquidity, and inconsistent technical standards. As these challenges are gradually addressed—through more cohesive regulations, standardization, and increased stakeholder engagement—local flexibility markets are expected to see wider adoption.

While TSO-level ancillary services typically require flexibility volumes of 1 MW or more and impose stricter technical requirements, local flexibility markets often have lower entry thresholds, creating opportunities for smaller DERs—such as household PV systems, HPs, or EV chargers—to participate and monetize their flexibility. However, even with these lower thresholds, local markets still demand that participating devices can measure and report consumption or generation in near real-time, adhere to standard communication protocols, and respond to grid signals within the timeframes defined by the market operator.

7 Conclusion

This document presents the groundwork for the realization of two aspects of the optimal selection of available flexibility, a crucial step in supporting aggregators and FSPs in their role of integrating DERs into various flexibility markets. By leveraging advanced forecasting methods and analysing different flexibility utilization options, we have established a solid foundation for the next phase of Task T4.4.

A central aspect of this work was the development of baseline forecasts for both EV CSs and HEMS devices. For EV fleets, two forecasting approaches were implemented:

- **Aggregated fleet forecasting**, which accurately captured base load but struggled with unpredictable peaks due to user behaviour.
- **Individual station-level forecasting**, which proved more challenging at locations with fewer CSs, as sporadic usage patterns made prediction uncertain. Locations with a higher number of CS showed more stable patterns, making them more predictable.

For HEMS devices, forecasting was conducted at the household level, modelling smart meter consumption, PV generation, and HP consumption separately. Generic models for PV and HP were also explored, achieving lower accuracy than household-specific models but proving useful for generalizing across different households.

Flexibility estimation models were developed to quantify the potential deviations from normal operation that can be leveraged for grid services or market participation. These models incorporate key considerations, including technical constraints and user comfort levels, enabling the aggregator to transform the forecasted baseline into actionable flexibility bids.

In evaluating flexibility utilization options, we identified key market opportunities, including TSO ancillary services, DA/ID market participation, network tariff optimization, and DSO local flexibility markets. The feasibility of each depends on technical constraints of both flexibility assets and markets, minimum bid sizes, and market access conditions. This analysis sets the stage for further refinement, leading towards the second phase of T4.4, where these options will be evaluated for the two fleets. Additionally, the focus will shift toward the development of an optimal selection algorithm, which will assist aggregators in determining which assets to activate when a bid is accepted.

8 References

[1] European Comission, *Directive (EU) 2019/944.*

[2] CoordiNet project, "D2.3 – Aggregation of large-scale and small-scale assets connected to the electricity network," 2021.

[3] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice (3rd ed), 2021.

[4] C. Develver, N. Sadeghianpourhamami, M. Strobbe and N. Refa, "Quantifying flexibility in EV charging as DR potential: Analysis of two real-world data sets," *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm),* 2016 .

[5] V. Barthel, J. Schlund, P. Landes, V. Brandmeier and M. Pruckner, "Analyzing the Charging Flexibility Potential of Different Electric Vehicle Fleets Using Real-World Charging Data," *Energies,* 2021.

[6] K. G. M. A. M. C. a. L. S. X. M. Zhang, "Forecasting Residential Energy Consumption: Single Household Perspective," *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA),* 2018.

[7] J. G. N. G. H. S. Oliver Frendo, "Data-driven smart charging for heterogeneous electric vehicle fleets," *Energy and AI,* 2020.

[8] S. S. X. C. X. Z. Y. K. J. C. Y. G. T. W. Yuanyuan Wang, "Short-term load forecasting of industrial customers based on SVMD and XGBoost," *International Journal of Electrical Power & Energy Systems,* 2021.

[9] G. Van Kriekinge, C. De Cauwer, N. Sapountzoglou, T. Coosemans and M. Messagie, "Day-Ahead Forecast of Electric Vehicle Charging Demand with Deep Neural Networks," *World Electr. Veh. J.,* 2021.

[10] H. Z. Y. Z. H. Z. L. N. G. Y. Junjie Hu, "Flexibility Prediction of Aggregated Electric Vehicles and Domestic Hot Water Systems in Smart Grids," *Engineering,,* vol. 7, no. 8, pp. 1101-1114, 2021.

[11] M. Chang, S. Bae, G. Cha and J. Yoo, "Aggregated Electric Vehicle Fast-Charging Power Demand Analysis and Forecast Based on LSTM Neural Network," *Sustainability* , 2021.

[12] "scikit learn," [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html.

[13] "github," [Online]. Available: https://github.com/Microsoft/LightGBM.

[14] "scikit learn," [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.HistGradientBoostingRegressor.html.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation,* 1997.

[16] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[17] "Open-meteo," [Online]. Available: https://open-meteo.com.

[18] "Keras," [Online]. Available: https://keras.io.

[19] "xgboost," [Online]. Available: https://xgboost.readthedocs.io/en/stable/python/index.html.

[20] "ARSO METEO," [Online]. Available: https://meteo.arso.gov.si.

[21] "X-FLEX project," [Online]. Available: https://xflexproject.eu.

[22] "STREAM project," [Online]. Available: https://stream-he-project.eu.

[23] "CoordiNet project," [Online]. Available: https://www.edsoforsmartgrids.eu/eu-projects/coordinet/.

[24] "Piclo Flex," [Online]. Available: https://www.piclo.energy/flex.

[25] "NODES," [Online]. Available: https://nodesmarket.com.

[26] Opentunity Grant Agreement. EC, 2022.

8.1 Acronyms

Table 12: Acronym meaning

Acronym	Meaning
aFRR	Automatic Frequency Restoration Reserve
API	Application Programming Interface
BESS	Battery Energy Storage Systems
CS	Charging Station
DA	Day-Ahead
DER	Distributed Energy Resource
DL	Deep Learning
DSO	Distribution System Operator

EV	Electric Vehicle
FSP	Flexibility Service Provider
GBMs	Gradient Boosting Machines
HEMS	Home Energy Management System
HGBRT	Histogram-Based Gradient Boosting Regression Tree
HP	Heat Pump
HVAC	Heating, Ventilation, And Air Conditioning
ID	Intraday
LGBM	Light GBM
LSTM	Long Short-Term Memory
mFRR	Manual Frequency Restoration Reserve
ML	Machine Learning
NMAE	normalized Mean Absolute Error
nMAPE	normalized Mean Absolute Percentage Error
PV	Photovoltaic
RFR	Random Forest Regressor
RNNs	Recurrent Neural Networks
SoC	State Of Charge
TSO	Transmission System Operator
V2G	Vehicle-To-Grid
wMAPE	Weighted Mean Absolute Percentage Error