Electric vehicles as mobile batteries: automating charge and discharge using machine learning predictions

Victoria Van Rillaer[a], Mathias de Schietere de Lophem[b] and Hélène Verhaeghe[c]

[a] *UCLouvain, Louvain-La-Neuve, Belgium, vanrillaervictroia@gmail.com CA*
[b] *UCLouvain, Louvain-La-Neuve, Belgium, mathias.deschietere@uclouvain.be*
[c] *UCLouvain, Louvain-La-Neuve, Belgium, helene.verhaeghe@uclouvain.be*

Abstract:
The rapid growth of electric vehicle (EV) adoption, together with the emergence of Renewable Energy Communities (RECs), is reshaping local energy system optimization. Integrating EVs into RECs enables a Community-to-Vehicle-to-Community (C2V2C) paradigm, where EVs act as mobile batteries capable of storing, transporting, and returning energy within and between communities. Automating this process requires reliable knowledge of driver intent; however, user-declared inputs are unreliable, as drivers tend to overestimate their energy needs or provide inconsistent information. In this work, machine learning techniques are used to forecast EV mobility patterns, charging demand, and connection flexibility, enabling the design of smart (dis)charging strategies that balance user convenience with community self-consumption. This paper presents a comprehensive framework combining predictive modelling of charging-related variables with a smart C2V2C (dis)charging algorithm. To evaluate it, a one-year synthetic dataset was generated for 1000 EV users across five archetypes, delivery workers, remote workers, unemployed individuals, parents, and commuters, with mobility behavior derived from European passenger mobility statistics, ensuring realistic charging patterns. Five behavioral prediction approaches are compared: a user-based clustering method, two session-based clustering strategies, a tree-based method, and a cosine similarity approach, each estimating the next charging destination, trip energy requirement, and connection duration. Building on these forecasts, a recommendation algorithm maximizes REC self-consumption by aligning EV charging with renewable surplus and enabling controlled discharging during deficit periods. Results show that all smart charging strategies improve self-consumption and reduce grid dependency, with gains most pronounced in spring and summer when photovoltaic generation peaks. In contrast to most existing work focused on delivered energy, this framework predicts energy needs, reducing behavioral bias and extending the prediction scope to multiple charging locations and destination forecasting.

Keywords:
Electric vehicles, Machine learning, Community-to-Vehicle-to-community (C2V2C), Renewable energy communities, Charge-discharge optimization.

1. Introduction

The rapid electrification of road transport, together with the increasing diffusion of variable renewable energy sources, is profoundly transforming electricity systems across Europe [1]. Over the past decade, electric vehicles (EVs) have evolved from a minor technology into a growing component of electricity demand, while solar and wind power have become important contributors [2, 3] to the electricity mix. These create both challenges and opportunities for decentralized energy systems, as the temporal mismatch between demand and generation intensifies the need for flexible resources.
In this context, EVs represent more than an additional load. Their storage capability enables them to serve as flexible assets that can shift demand over time and, under bidirectional operation, supply

electricity back to energy systems [4]. This is particularly relevant for Renewable Energy Communities (RECs) [5], which the European Union promotes to enhance local renewable energy integration, self-consumption, and citizen participation in the energy transition. By aligning EV charging behavior with local renewable energy availability, RECs can improve overall system efficiency.

Bidirectional charging schemes enable Vehicle-to-Everything (V2X) services. In this work, we consider the concept of Community-to-Vehicle-to-Community (C2V2C), in which EVs operate as mobile storage units that can store a surplus of renewable energy and later discharge it. Despite its significant technical potential, its practical deployment remains limited. At the regulatory level, European directives such as the Renewable Energy Directive [1] and the Internal Electricity Market Directive [2] establish a legal framework for energy communities but impose constraints on electricity injection and storage usage. In several countries, like Belgium [6], Austria [7], Hungary [8], and Lithuania [9], only renewable electricity may be injected into a REC, restricting the use of EV batteries for reinjection. It must therefore be explicitly accounted for in smart charging and discharging strategies.

Recent research has investigated coordinated control and optimization frameworks for C2V2C energy exchange [10, 11]. Existing studies typically propose multistage approaches combining data acquisition from EVs and local energy systems, estimation of target states of charge, and optimization of charging and discharging schedules, aiming at maximizing renewable self-consumption or facilitating energy balance across communities. Reported results indicate substantial performance improvements compared to baseline scenarios without coordinated charging, particularly in contexts with high renewable generation at the workplace. However, these approaches generally rely on aggregated or simplified representations of EV availability and mobility, limiting their ability to capture heterogeneous usage patterns and realistic constraints at the individual vehicle level.

In parallel, a growing body of literature has focused on modeling and predicting EV charging behavior using data-driven methods [12]. Clustering-based approaches have been proposed to characterize typical charging sessions [13, 14] or user profiles [15]. Supervised learning and similarity-based methods have been applied to forecast charging duration, energy demand, or flexibility potential [16]. Although these methods demonstrate promising predictive performance, they are often restricted to a single charging context, such as residential or workplace charging, and frequently rely on user-declared charging demand as a proxy for mobility needs. Empirical evidence suggests that such assumptions are problematic, as EV users tend to systematically overestimate required charging due to driving range anxiety [4], leading to structural biases in forecasting and optimization processes.

Existing work has integrated EV charging forecasts into energy management frameworks, notably through day-ahead optimization combined with rule-based real-time control in RECs [18]. However, EV behavior is typically represented via aggregated or synthetic demand profiles, without explicit user- or session-level modeling. Recent data-driven approaches improve short-term charging demand prediction at the station level using ensemble and deep learning methods [19], yet they still infer behavior implicitly and overlook individual mobility patterns and user heterogeneity.

Consequently, current energy management frameworks for RECs lack behaviorally grounded models capable of accurately predicting where vehicles will travel, how much energy they will require, and how long they will remain connected across multiple locations, which is necessary to optimize self-consumption and energy exchanges. This limitation is particularly critical for coordinated charging and discharging strategies operating under regulatory constraints. Inaccurate assumptions about EV availability or energy needs can significantly reduce flexibility and self-consumption gains.

To address these challenges, this work develops and evaluates data-driven methods for predicting EV mobility and charging behavior from the perspective of individual vehicles across trips and loca-

[1] Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources. `http://data.europa.eu/eli/dir/2018/2001/oj`

[2] Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU. `http://data.europa.eu/eli/dir/2019/944/oj`

tions. The proposed approach combines behavioral prediction with smart charging and discharging strategies designed to exploit EV flexibility within RECs while explicitly accounting for regulatory constraints, thereby enabling more efficient integration of electric vehicles, increasing local renewable energy utilization, and enhancing system-level flexibility and resilience. Finally, the proposed methods are evaluated through simulation using synthetic mobility data, enabling a systematic assessment of predictive performance and energy system impacts under controlled conditions.

2. Methodology

2.1. Datasets

A major limitation identified in the literature is the lack of accessible, open EV datasets that adopt a vehicle-centric perspective. Existing datasets focus on charging infrastructure, reporting isolated charging sessions at specific locations (e.g., workplace or public chargers) without capturing the vehicle's complete mobility. Such charger-based datasets do not provide information on where the EV comes from or goes to. However, when EVs act as flexible transport of local electricity surpluses, it is necessary to know each EV's full travel path. This includes trip sequences, parking locations, and charging opportunities throughout the day. To overcome these limitations, this work combines real REC energy data with synthetic EV mobility profiles that explicitly model EVs' full daily behavior.

2.1.1. REC datasets

The energy data used in this study originates from RECs managed by WeSmart[3], a company that facilitates energy sharing within communities in Belgium. These datasets provide realistic electricity consumption and surplus-injection profiles, measured at 15-minute resolution over a full year. They are anonymized real-world datasets collected from the operational management of a residential community and a workplace-based community in Brussels, Belgium. These datasets are used to characterize the temporal availability of local electricity surplus that could be absorbed by EVs.

2.1.2. Synthetic EV mobility and charging dataset

Due to the absence of open datasets providing complete EV mobility trajectories, synthetic EV profiles were generated using *emobpy* [20], an open-source Python-based tool that generates electric vehicle time series from empirical mobility statistics, vehicle physical characteristics, and user-defined behavioral assumptions. For each EV, emobpy produces four complementary time series: vehicle mobility, driving electricity consumption, EV grid availability, and EV grid electricity demand. The final time series generated was processed in order to create a charging sessions dataset.
Profiles are generated using a Monte Carlo approach with probabilistic input assumptions. This ensures variability. Charging infrastructure availability depends on the location (home, workplace, or public), each of which is associated with a specific nominal charging power.
To reflect heterogeneous mobility behaviors, five representative driver profiles were defined, each characterized by distinct travel patterns. The main assumptions underlying each profile are:

- *Delivery person*: high activity level during weekdays (average of 5 trips/day). It includes a daily commute of approximately 1 hour (to start and end the working day) and multiple work-related trips. Charging opportunities vary and include workplace and public infrastructure.
- *Commuter*: regular and structured weekday mobility (average of 2.3 trips/day), dominated by home-to-work and work-to-home travel. Charging is mainly available at home and workplace.
- *Parents*: moderate activity level (3 trips/day on average), characterized by trip chaining (e.g., escorting children, errands), leading to more variable schedules and intermediate stops. Charging opportunities are mixed across home, public locations, and workplace.
- *Remote worker*: low mobility (2.1 trips/day on average) and high presence at home compared to others. Less structured trips and more flexible in timing. Primary reliance on home charging.

[3] `https://wesmart.com/en`

- *Unemployed*: low and irregular mobility demand (2.1 trips/day on average), with no commuting. Remain at home for several consecutive days. Charging is predominantly home-based.

During weekends, mobility patterns across all profiles are assumed to converge toward similar behavior, with reduced differentiation compared to weekdays. The relative probabilities of these driver types were adapted based on European mobility studies [21, 22], ensuring consistency with observed travel patterns. For each category, specific constraints on trip frequency, destinations, and parking durations were applied, allowing the synthetic dataset to capture realistic inter-user diversity.
The final EV dataset comprises 1,000 users (200 per driver type). All time series are sampled at a 15-minute resolution and span 53 weeks for training data and 4 weeks for testing data (Monday-Sunday). The data generation process assumes a "typical" or steady-state week throughout the entire period, i.e., without accounting for seasonal effects such as holidays, school closures, or atypical events.
The generated datasets are publicly available on Dataverse [24]. All code used for dataset generation is openly accessible on GitHub[4], ensuring full reproducibility of the data used in this work.

2.2. Simulation setup and data completion

All simulations were conducted at the individual EV level for 1000 profiles. To enable charging and discharging decisions, several variables were initialized, computed, or predicted. The initial state of charge (SoC) was set to 50% for all EVs. During the simulation, the following variables are completed for each EV: 1) battery energy level (kWh), 2) energy charged from the REC (kWh), 3) energy charged from the grid (kWh), 4) predicted next destination, 5) predicted plug-out time (hours), and 6) predicted energy required for the next trip (kWh).

2.3. Prediction of the charging patterns

The prediction of EV mobility patterns is addressed from an EV-centric perspective. The prediction problem is decomposed into three tasks: 1) the prediction of the next destination (multiclass classification); 2) the prediction of the energy required for the upcoming trips (regression); and 3) the prediction of the duration of the current charging session (regression).
Five models drawn from two methodological families are adapted to the EV-centric perspective. The adaptations consist of augmenting each model with a next-destination prediction component and replacing the energy-charged variable with the net energy required to cover the consumption anticipated before the subsequent charging session. First, the family of *behavioral models* aims to capture charging habits and mobility patterns by either identifying latent sessions or driver types through clustering, or by identifying historically similar sessions. The models used are the following: Population-level Gaussian Mixture Model (BEH-GMM-P), Individual-level Gaussian Mixture Model (BEH-GMM-I), 2-step sessions and users clustering (BEH-2Step-Clust), and Similarity-based predictions (BEH-SimS). Second, the family of *predictive models* uses supervised learning to maximize predictive accuracy, here via a tree-based model: Light Gradient Boosting Machine (TREE-LGBM).

2.3.1. Behavioral - Population-level Gaussian Mixture Model (BEH-GMM-P)

BEH-GMM-P [13] models EV charging behavior using a Gaussian mixture fitted on the set of charging sessions aggregated across all EVs. A Gaussian Mixture Model (GMM) [25] represents a probability distribution as a weighted sum of Gaussian components, where each can be intuitively interpreted as capturing a distinct underlying subpopulation within the data. Each component represents a latent charging session type characterized by typical arrival time, charging duration, and consumption for the next trip. By learning these shared patterns across the population, the model captures common charging and mobility behaviors to generate predictions even with limited or no historical data.
Predictions are obtained by conditioning the learned mixture on the observed arrival time of the current session. The predicted session duration and energy demand are then computed as conditional

[4] `https://github.com/vicvrl/EV_profiles_generation.git`

expectations across the components. In addition, the model predicts the EV's next destination. During training, each mixture component is associated with a categorical distribution over destinations learned from the historical sessions assigned to that component. Given the arrival time, the probability of each destination is obtained by marginalizing over the mixture components. The predicted destination is then taken as the most probable class under this distribution.

2.3.2. Behavioral - Individual-level Gaussian Mixture Models (BEH-GMM-I)

While the BEH-GMM-P captures common charging patterns across the population, EV users often exhibit regular individual charging habits. To capture these personalized behaviors, an Individual-Level Gaussian Mixture Model (BEH-GMM-I) [13] is trained on each EV's historical charging sessions.
Predictions are obtained using the same conditioning procedure as in BEH-GMM-P, but using mixture parameters learned from the historical sessions of the corresponding EV. This exploits correlations between arrival time, session duration, energy consumption, and destination within each EV's historical charging behavior, enabling personalized predictions.

2.3.3. Behavioral - Two-step user clustering and predictions (BEH-2Step-clust)

To acknowledge the intra-user variability in behavior across time of day and charging location, a two-step clustering approach is considered [15]. First, a population-level GMM is used to define a typology of charging sessions based on a feature set capturing session characteristics and near-term mobility demand. Sessions are described using four features: connected duration, the next distance between two sessions (DBS), hours between two sessions (HBS), and the plug-in time. The number of clusters is selected using the Bayesian Information Criterion (BIC) and the elbow method.
Second, each driver is represented by a portfolio describing the proportions of their sessions by session type. Users are then clustered into archetypes using a hierarchical clustering algorithm. These work by grouping similar data points based on their similarity into trees of clusters. The idea is to begin with each data point as its own separate cluster and then merge or split them based on their similarity.
For prediction, the model adopts a hybrid approach combining the learned clustering structure with supervised learning. At inference time, the session cluster of a new observation is not directly inferred from the GMM, as several input features (e.g., DBS or HBS) are not observable at plug-in time. Instead, a lookup-based procedure assigns the most likely session cluster by matching the current observation to historical sessions with similar characteristics. This matching is performed hierarchically using the user cluster, charging location, and plug-in time within a configurable temporal window. If no exact match is found, the procedure progressively relaxes these constraints to ensure a robust assignment. The inferred session cluster, together with observable features such as user identifier, user cluster, plug-in time, charging location, and arrival state of charge, forms the input to a set of supervised learning models. Gradient boosting models are trained to predict three key outcomes: the next destination (classification task), the next charging state (regression), and the connected duration (regression). It was stated that the gradient-boosted decision tree implementations [16] demonstrated particular effectiveness in EV charging applications while efficiently handling heterogeneous feature types and nonlinear relationships inherent to user charging behavior.
Although connected duration is both a GMM clustering feature and a downstream prediction target, no circularity arises: the GMM is fitted exclusively on completed historical sessions where all four features are fully observed, while at inference time the session cluster is assigned via the lookup procedure using only plug-in-time observables (user cluster, location, and plug-in time), so the duration of the incoming session is never required nor used.

2.3.4. Behavioral - Similarity-Based prediction model (BEH-SimS)

The SimS model, adapted from [16], predicts charging session outcomes by retrieving historical sessions most similar to the current charging state. Unlike parametric models, SimS is an instance-based method that does not require training. Each session is represented by a numerical feature vector composed of temporal and contextual features. Temporal information is encoded using the plug-in

timestamp (hour of day, month, and day of week). To preserve the periodic nature of these variables, each temporal feature is transformed using sine and cosine representations. In addition, categorical variables, including the user identifier and the charging location, are encoded using one-hot encoding to capture user-specific charging habits and contextual differences across charging locations.
For a new session, similarity with each historical session is computed using cosine similarity.Only sessions occurring prior to the current timestamp are considered in the similarity search to ensure a strictly causal prediction setting. The m most similar sessions are selected, and predictions are derived from their observed outcomes. Specifically, the predicted next destination is obtained as the mode of the retrieved sessions, while the predicted energy consumption for the next trip and the charging session duration are computed as the mean values across the selected sessions.
By incorporating user identifiers as features, SimS naturally exploits both within-user historical patterns and cross-user behavioral similarities when retrieving relevant past sessions.

2.3.5. Predictive - Tree-Based predictive model (TREE-LGBM)

Light Gradient Boosting Machine (LGBM) models are particularly well-suited for tabular data, and tree-based methods have demonstrated strong predictive performance in modeling EV behavior [16]. LGBM sequentially builds an ensemble of decision trees, where each tree partitions the feature space into regions with similar target values. At each iteration, a new tree is trained to correct the ensemble's residual errors, allowing the model to capture complex and non-linear relationships [17].
As the three tasks correspond to distinct outputs and learning objectives, separate single-output LGBM models are trained per task. Categorical variables are encoded as one-hot vectors, enabling the model to capture place- and user-specific effects. Key hyperparameters considered include the number of leaves, maximum tree depth, learning rate, minimum data in a leaf, and regularization terms. These parameters control the complexity of the trees and the learning dynamics of the boosting process. Hyperparameters are tuned using the Tree-structured Parzen Estimator (TPE) algorithm.

2.3.6. Evaluation

Model performance is evaluated separately for regression and classification tasks. For charging duration and next consumption between sessions (CBS), formulated as regression problems, accuracy is quantified using the Mean Absolute Error (MAE), reported in the natural units of the target variable (hours for duration and kWh for CBS), and the Symmetric Mean Absolute Percentage Error (SMAPE), which captures relative prediction error. For next-destination prediction, formulated as a multiclass classification problem, performance is measured by the success rate, defined as the proportion of charging sessions in which the predicted destination matches the observed destination.

2.4. Charging and discharging recommendation algorithms

The proposed recommendation algorithm, illustrated in Figure 1, determines charging and discharging actions for EVs connected to charging infrastructure within RECs. Its primary objective is to maximize local renewable energy self-consumption while ensuring that each EV satisfies its expected mobility needs. The algorithm operates at each plug-in event and provides charging or discharging recommendations over the entire predicted connection duration.
Unlike centralized scheduling approaches, the algorithm follows a rule-based decision logic that integrates predicted mobility behavior, local energy availability, and battery constraints.

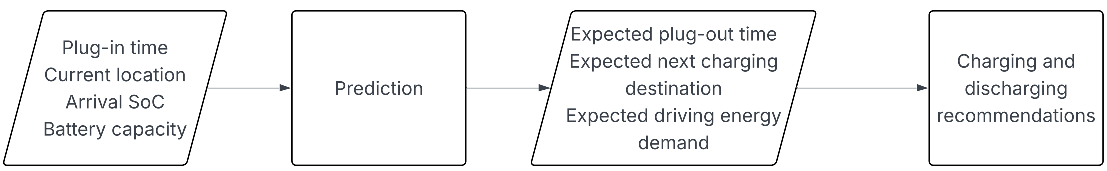

Figure 1: Flowchart of the recommendation algorithm

2.4.1. Inputs and Prediction

Upon plug-in, the algorithm receives the following inputs: 1) the plug-in time and location, 2) the arrival state of charge (SoC), and 3) the battery capacity. Based on these inputs, the prediction module provides 1) the expected plug-out time, 2) the next charging destination, and 3) the expected driving energy demand until the next charging session.
Locations classified as home or workplace are assumed to belong to RECs with local renewable production, while public locations are treated as grid-only access points.

2.4.2. Energy requirement formulation

The energy required to guarantee the next mobility period (E_{needed}) is defined as the sum of the predicted energy (i.e., expected driving energy demand until next charging event), and an unpredicted amount of energy accounting for safety margins related to driving range anxiety factor. As in [27], we use a factor of 1.5 (i.e., the unpredicted energy is 1.5 times the predicted one). Minimum and maximum energy thresholds are enforced to avoid excessive battery degradation by maintaining the SoC above a predefined lower bound (20%) and a predefined higher bound (80%) [28].

2.4.3. Charging and discharging decision logic

When the EV plugs in, a first decision is made based on the state, predicted future action (E_{needed}, next destination), and the current charging location.
Non-REC locations (public, fast charging): The EV only charge until E_{needed} is reached. The goal is to minimize the charging at those places.
REC-connected locations (home, workplace): Figure 2 illustrates the workflow of charging and discharging recommendations when plugged at a REC.

- If the predicted renewable surplus during the connection exceeds E_{needed}, charging is supplied exclusively from the local surplus until the predefined upper SoC limit is reached.
- If the surplus is insufficient, the algorithm prioritizes renewable charging and complements the remaining energy demand with grid supply as early as possible.

At a REC, discharging may also be recommended to support renewable consumption during periods of renewable production deficit. The discharged energy is constrained by the available battery energy beyond E_{needed}, the renewable-origin energy stored in the battery and the maximum discharge power.

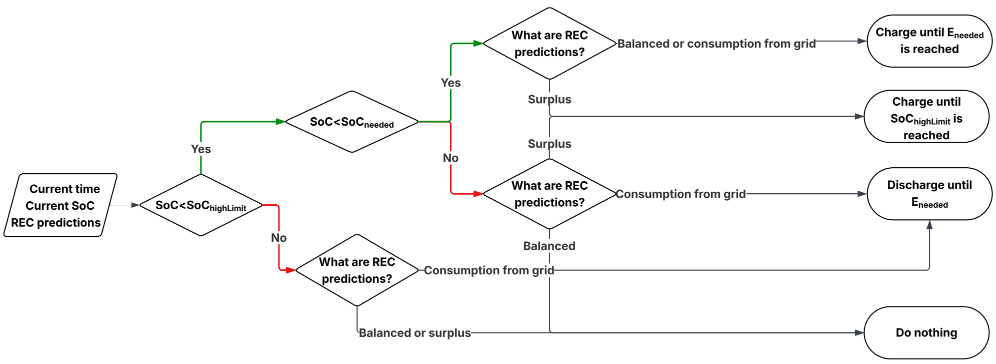

Figure 2: Flowchart of the decision algorithm. Case: the EV is connected to a REC.

2.5. Implementation and constraints

The recommendation process is evaluated in discrete 15-minute intervals over the entire connection duration. At each time step, the EV may charge, discharge, or remain idle. All actions respect battery capacity limits, charging efficiency, and power constraints. The algorithm ensures that mobility requirements are preserved while enabling flexible interaction between EVs and local renewable generation. The REC production and consumption predictions are assumed to be reliable over the considered horizon, allowing the focus to remain on the charging decision logic.

3. Results and discussion

This section presents the predictive performance of the proposed models and evaluates their impact on smart charging strategies through simulation. The code can be found in the GitHub repository[5].

3.1. Prediction Results

Table 1 presents the predictive performance of the proposed models for the three target variables: next destination, session duration, and energy consumption of the next trip. Destination prediction is evaluated using accuracy, while session duration and energy consumption are assessed using mean absolute error (MAE). The corresponding SMAPE values are reported in parentheses.
For destination prediction, the TREE-LGBM model achieves the highest accuracy, reaching 68%. The BEH-2Step-Clust and BEH-SimS approaches follow, while the BEH-GMM-P and BEH-GMM-I obtain slightly lower performance. Regarding session duration prediction, BEH-2Step-Clust provides the lowest MAE (3.81 h), indicating the best performance for this task. TREE-LGBM and BEH-SimS yield intermediate errors, while BEH-GMM-I and BEH-GMM-P show higher errors. In terms of relative error, BEH-SimS achieves the lowest SMAPE (65%), although its MAE remains higher than that of BEH-2Step-Clust. For energy consumption prediction, all models exhibit similar performance. The BEH-GMM-P approach achieves the lowest MAE (0.11 kWh) and SMAPE (61%). Overall, the results indicate that performance depends on the prediction task. The SMAPE values are consistent with MAE results and do not modify the ranking of the models. This indicates that absolute and relative error metrics convey similar information in this context, providing limited additional insight.

Table 1: Prediction performance of the different models for the three tasks

Model	Destination Acc. ↑	Duration MAE (h) (SMAPE) ↓	Consumption MAE (kWh) (SMAPE) ↓
BEH-GMM-P	63%	7.67 (74%)	**0.11** (**61%**)
BEH-GMM-I	63%	7.24 (72%)	**0.11** (62%)
BEH-2Step-Clust	67%	**3.81** (70%)	0.12 (64%)
BEH-SimS (m=50)	67%	6.47 (**65%**)	0.12 (76%)
TREE-LGBM	**68%**	5.91 (72%)	0.12 (63%)

A more detailed analysis reveals heterogeneity across location contexts. Session duration prediction is notably more challenging in residential settings, likely due to higher variability in user behavior compared to workplace or public charging environments.
In contrast, energy consumption prediction exhibits relatively stable performance across locations and models. The observed differences are marginal and correspond to a small fraction of total battery capacity, suggesting limited practical impact in the studied dataset. However, this observation may not generalize to scenarios with higher variability in charging demand.
Finally, destination prediction accuracy is systematically higher in public charging contexts, indicating more regular and predictable mobility patterns. Across all tasks, LGBM demonstrates the most consistent and robust performance, while SimS remains a competitive alternative specifically for destination prediction.

3.2. Simulation results

This section presents simulation results for the different smart charging strategies and compares them with the non-smart baseline (*noSM*) and the ideal predictive benchmark, i.e., an oracle case with an

[5] `https://github.com/vicvrl/EV-as-mobile-batteries-predictions-and-simulations.git`

hypothetic 100%-accurate predictor (*SM_ORACLE*). A second non-smart strategy (*noSM_noPublic*) is simulated, which considers a potential incentive for the driver to charge only at home or work. The analysis is structured around three main questions: 1) the improvement in EV self-consumption relative to *noSM*, 2) the reduction in grid use enabled by the EV support at both RECs, and 3) the impact of prediction errors, assessed by comparison with the oracle strategy.

3.2.1. Improvement in renewable energy self-consumption

The mean monthly self-consumption per EV, averaged across all simulation runs and separated by plugin location, is shown in Table 2. As expected, *noSM* and *noSM_noPublic* (not in the table) exhibit values close to zero. This reflects typical user behavior, in which charging predominantly occurs in the evening after returning home, i.e., outside local photovoltaic production. The *noSM_noPublic* configuration shows slightly higher values, but the difference remains negligible, confirming that infrastructure alone does not significantly enhance self-consumption in the absence of coordinated control.

Table 2: Mean monthly self-consumption per EV (kWh) by strategy and location (home and workplace). Values are averaged over all simulation runs. In parentheses are the self-consumption gap to oracle (%).

	Home (kWh)				Workplace (kWh)			
Strategy	Jan	Mar	Jun	Oct	Jan	Mar	Jun	Oct
SM_BEH-GMM-I	18.80 (48.7)	93.50 (44.8)	213.15 (32.7)	79.48 (49.9)	**0.46** (**43.9**)	2.78 (53.7)	10.67 (48.5)	1.24 (53.5)
SM_BEH-GMM-P	19.47 (46.8)	94.75 (44.0)	211.86 (33.1)	79.92 (49.6)	**0.46** (**43.9**)	2.78 (53.7)	10.66 (48.6)	1.25 (53.0)
SM_BEH-SIMS	**21.60** (**41.0**)	**112.71** (**33.4**)	**238.73** (**24.6**)	**97.56** (**38.5**)	0.30 (63.4)	2.20 (63.4)	10.18 (50.9)	0.99 (62.8)
SM_BEH-2Step-Clust	4.80 (86.9)	51.21 (69.8)	138.37 (56.3)	46.94 (70.4)	0.31 (61.9)	**3.56** (**40.8**)	**13.43** (**35.2**)	**1.63** (**38.5**)
SM_TREE-LGBM	18.12 (50.5)	94.94 (43.9)	205.85 (35.0)	84.18 (46.9)	0.24 (70.3)	1.88 (68.7)	8.91 (57.0)	0.69 (74.0)
SM_ORACLE	36.63	169.32	316.75	158.57	0.82	6.01	20.72	2.66

Overall, all smart charging strategies significantly improve self-consumption, though the magnitude of the improvement varies by month and location. At home, substantial gains are observed for all smart strategies. The highest improvements occur in June, which is consistent with increased PV generation during summer and higher availability of surplus energy. This highlights the strong dependency of smart charging performance on renewable generation.

A ranking of the strategies can be observed. While, *SM_BEH-SIMS* achieves the highest performance, *SM_BEH-GMM-I*, *SM_BEH-GMM-P*, and *SM_TREE-LGBM* achieve comparable, consistently high performance. In contrast, *SM_BEH-2Step-Clust* is worse at home. This can be attributed to its lower ability to accurately predict long charging sessions happening essentially at home, which are critical for aligning charging with local PV production.

At the workplace, self-consumption remains significantly lower across all strategies. This can be explained by a combination of factors: (1) the workplace REC already exhibits a high level of intrinsic self-consumption, leaving limited surplus available for EV charging, and (2) EV often arrive with relatively high SoC, reducing charging needs during working hours. As a result, most energy exchanges occur at home rather than at the workplace. Interestingly, the *SM_BEH-2Step-Clust* strategy performs comparatively better at the workplace than at home, suggesting that it is more effective at capturing workplace charging patterns, which tend to be more regular and predictable.

Finally, the results reveal a seasonal effect. Gains are highest in June, intermediate in March and October, and lowest in January. This confirms that smart charging strategies are strongly influenced by PV availability, with higher generation leading to more frequent surplus periods and more favorable charging opportunities. These findings underscore the importance of evaluating control strategies across contrasting seasonal conditions rather than relying on a single representative period.

3.2.2. Impact of prediction errors: comparison with the oracle scenario

The predictive strategies were compared with an oracle scenario (perfect foresight) to isolate the loss attributable to imperfect forecasts. Table 2 reports, in parentheses, the self-consumption gap to the oracle benchmark at both locations. At home, the best method (*SM_BEH-SIMS*) remains 25-41% below the oracle, depending on the month, with the smallest gap observed in June when higher PV availability facilitates alignment. At the workplace, *SM_BEH-2Step-Clust* narrows its gap to 35-62%. *SM_BEH-GMM-I* and *SM_BEH-GMM-P* remain quasi-identical across all configurations, indicating that individualizing the GMM yields no measurable gain for the dataset used.

3.2.3. Reduction in grid use and REC support

The second objective of the analysis is to evaluate the reduction of electricity imported from the grid. This is assessed by the reduction in residual grid draw after EV support relative to *noSM* (no discharging). Table 3 shows the relative reduction in grid use at home achieved by each strategy compared to the *noSM*. Overall, smart charging strategies reduce residual grid draw, confirming their ability to better align EV charging with locally available energy.
The magnitude of the reduction varies across both strategies and months. The strongest improvements are observed in June, reaching up to 1.21% for the *SM_ORACLE* strategy, followed by March and October, while January exhibits the lowest gains. This seasonal trend is consistent with self-consumption results. It reflects higher PV availability during spring and summer, creating more opportunities to overcharge EVs, allowing them to later discharge RE into the REC.
Among the predictive strategies, *SM_BEH-GMM-I*, *SM_BEH-GMM-P*, and *SM_TREE-LGBM* exhibit similar performance, achieving reductions close to 0.7-1.0% during favorable months while *SM_BEH-SIMS* achieves the best performance. In contrast, the *SM_BEH-2Step-Clust* approach systematically underperforms, with reductions approximately half those of the best-performing methods. This confirms its lower ability to capture charging flexibility, particularly for longer and less regular home charging sessions.
It should be noted that the reductions reported in Table 3 remain below 1.5%, as the metric is expressed relative to *total* REC consumption, which includes all non-EV loads beyond the reach of smart charging. Additionally, the present study models a single EV operating across two renewable energy communities, which represents a conservative setting: the absolute grid benefit per community is inherently limited by the energy volume of a single charging session (EV battery capacity and charger power). These results should therefore be understood as a per-vehicle lower bound, with greater absolute reductions expected as community EV penetration increases.

Table 3: Relative reduction in residual grid draw at the residential REC compared to *noSM*, reported by month (%).

Residual grid draw - Home (%)				
Strategy	Jan	Mar	Jun	Oct
SM_BEH-GMM-I	0.236	0.777	0.980	0.713
SM_BEH-GMM-P	0.300	0.775	0.905	0.817
SM_BEH-SIMS	**0.285**	**1.001**	**1.075**	**0.918**
SM_BEH-2Step-Clust	0.089	0.428	0.695	0.370
SM_TREE-LGBM	0.222	0.757	0.868	0.702
SM_ORACLE	0.323	1.180	1.209	1.046

4. Conclusion

This paper investigated the prediction of EV charging behavior using complementary modeling approaches, including probabilistic clustering methods and a tree-based model, addressing three key tasks: next destination, session duration, and energy consumption. The resulting predictions were integrated into a simulation framework spanning two RECs and public charging infrastructures through a rule-based smart-charging and discharging recommendation algorithm.
Results confirm that smart charging strategies effectively improve self-consumption and reduce grid dependency, with gains most pronounced when PV availability is highest. Predictive strategies recover a substantial share of the oracle scenario, demonstrating that accurate behavioral forecasting translates into measurable energy community benefits, even under imperfect predictions.

Several limitations remain. The framework models a single EV, and extending it to multi-vehicle communities represents a natural and promising direction, where fleet-level coordination could yield substantially larger grid benefits. The synthetic dataset used for training assumes temporally homogeneous charging behavior, with no seasonal variation in EV usage patterns; including the seasonal structure of the charging session with national and school holidays, e.g., represents an important direction for future work. More broadly, EV usage is inherently stochastic and shaped by factors difficult to capture from historical data alone, such as days off and atypical patterns. Incorporating user input directly into the prediction pipeline, for instance through calendar integration or contextual queries regarding planned departures or upcoming trips, could significantly reduce this uncertainty and improve scheduling decisions. Finally, real-world deployment would require accounting for electricity price signals, battery degradation, and driver feedback, all of which are left for future work.

Acknowledgments

The authors would like to sincerely thank WeSmart for providing access to the datasets from their RECs, which made this research possible. Computational resources have been provided by the *Consortium des Équipements de Calcul Intensif* (CÉCI), funded by the *Fonds de la Recherche Scientifique de Belgique* (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

References

[1] Blumberg G., Broll R., Weber C. *The Impact of Electric Vehicles on the Future European Electricity System—A Scenario Analysis*. Energy Policy 2022;161:112751.

[2] Ritchie H., Rosado P. *Electricity Mix*. Our World in Data 2020. Available at: `https://ourworldindata.org/electricity-mix` [accessed 19.1.2026].

[3] Ritchie H. *Tracking Global Data on Electric Vehicles*. Our World in Data 2024. Available at: `https://ourworldindata.org/electric-car-sales` [accessed 19.1.2026].

[4] Li X., Wang Z., Zhang L., Sun F., Cui D., Hecht C., Figgener J., Sauer DU. *Electric Vehicle Behavior Modeling and Applications in Vehicle-Grid Integration: An Overview*. Energy 2023;268:126647.

[5] Menyhart J. *Overview of Sustainable Mobility: The Role of Electric Vehicles in Energy Communities*. World Electr Veh J 2024;15(6):275.

[6] Ecorys España SL. *Energy Communities Policy Framework: Belgium—Region of Flanders*. European Energy Communities Facility (EECF); 2025. Technical Report. Grant agreement No. 101167230.

[7] Ecorys España SL. *Energy Communities Policy Framework: Austria*. European Energy Communities Facility (EECF); 2025. Technical Report. Grant agreement No. 101167230.

[8] Ecorys España SL. *Energy Communities Policy Framework: Hungary*. European Energy Communities Facility (EECF); 2025. Technical Report. Grant agreement No. 101167230.

[9] Ecorys España SL. *Energy Communities Policy Framework: Lithuania*. European Energy Communities Facility (EECF); 2025. Technical Report. Grant agreement No. 101167230.

[10] Zhang L., Alvarez EMO., Huang P. *Optimizing Coordinated Spatio-Temporal Control of Electric Vehicles for Enhanced Energy Sharing and Performance across Building Communities*. Energy Build 2024;312:114167.

[11] Board A., Sun Y., Huang P., Xu T. *Community-to-Vehicle-to-Community (C2V2C) for Inter-Community Electricity Delivery and Sharing via Electric Vehicle: Performance Evaluation and Robustness Analysis*. Appl Energy 2024;363:123054.

[12] Yaghoubi E., Yaghoubi E., Khamees A., Razmi D., Lu T. *A Systematic Review and Meta-Analysis of Machine Learning, Deep Learning, and Ensemble Learning Approaches in Pre-*

dicting EV Charging Behavior. Eng Appl Artif Intell 2024;135:108789.

[13] Lee ZJ., Li T., Low SH. *ACN-Data: Analysis and Applications of an Open EV Charging Dataset*. In: Proceedings of the Tenth ACM International Conference on Future Energy Systems (E-Energy '19); 2019 Jun 15; New York, USA. ACM:139–49.

[14] Van Kriekinge G., De Cauwer C., Sapountzoglou N., Coosemans T., Messagie M. *Electric Vehicle Charging Sessions Generator Based on Clustered Driver Behaviors*. World Electr Veh J 2023;14(2):37.

[15] Helmus JR., Lees MH., Van Den Hoed R. *A Data Driven Typology of Electric Vehicle User Types and Charging Sessions*. Transp Res Part C Emerg Technol 2020;115:102637.

[16] Genov E., De Cauwer C., Van Kriekinge G., Coosemans T., Messagie M. *Forecasting Flexibility of Charging of Electric Vehicles: Tree and Cluster-Based Methods*. Appl Energy 2024;353:121969.

[17] Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., and Liu T.-Y., *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. Advances in Neural Information Processing Systems 30 (NeurIPS 2017), 2017.

[18] Pezzati N., Innocenti E., Berzi L., Delogu M. *Scalable Energy Management Model for Integrating V2G Capabilities into Renewable Energy Communities*. World Electr Veh J 2025;16(8):450.

[19] Mao Y., Yu X., Wang F., Zhu J. *Electric Vehicle Charging Demand Forecasting: A Data-Driven Integrated Learning Approach*. Renew Energy 2026;256:124141.

[20] Gaete-Morales C., Kramer H., Schill W-P., Zerrahn A. *An Open Tool for Creating Battery-Electric Vehicle Time Series from Empirical Data, emobpy*. Sci Data 2021;8(1):152.

[21] Eurostat. *Passenger Mobility Statistics*. Available at: `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Passenger_mobility_statistics` [accessed 5.1.2026].

[22] European Alternative Fuels Observatory. *Electric Vehicle Model Statistics*. Available at: `https://alternative-fuels-observatory.ec.europa.eu/policymakers-and-public-authorities/electric-vehicle-model-statistics` [accessed 9.1.2026].

[23] Baden KB. *More than Just Miles: Range Anxiety Unplugged*. Transp Res Part A Policy Pract 2025;200:104661.

[24] Van Rillaer V., de Schietere de Lophem M., Verhaeghe H. *Synthetic Dataset of Electric Vehicles Charging Sessions*. Open Data @ UCLouvain 2026. `https://doi.org/10.14428/DVN/6GUFM9`.

[25] Reynolds D. *Gaussian Mixture Models*. MIT Lincoln Laboratory; Encyclopedia of Biometrics. Available at: `http://leap.ee.iisc.ac.in/sriram/teaching/MLSP_16/refs/GMM_Tutorial_Reynolds.pdf`

[26] Scikit-learn developers. *Histogram-Based Gradient Boosting*. scikit-learn User Guide. Available at: `https://scikit-learn.org/stable/modules/ensemble.html#histogram-based-gradient-boosting` [accessed 24.3.2026].

[27] Gonzalez Venegas F., Petit M., Perez Y. *Plug-in Behavior of Electric Vehicles Users: Insights from a Large-Scale Trial and Impacts for Grid Integration Studies*. eTransportation 2021;10:100131.

[28] Kostopoulos ED., Spyropoulos GC., Kaldellis JK. *Real-World Study for the Optimal Charging of Electric Vehicles*. Energy Rep 2020;6:418-26.

[29] Shahriar S., Al-Ali AR., Osman AH., Dhou S., Nijim M. *Machine Learning Approaches for EV Charging Behavior: A Review*. IEEE Access 2020;8:168980–93.