Kreft, Markus; Brudermueller, Tobias; Fleisch, Elgar; Staake, Thorsten

Predictability of electric vehicle charging : Explaining extensive user behavior-specific heterogeneity

Date of secondary publication: 31.10.2025

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-111090x

Primary publication

Kreft, Markus; Brudermueller, Tobias; Fleisch, Elgar; Staake, Thorsten (2024): Predictability of electric vehicle charging : Explaining extensive user behavior-specific heterogeneity, in: Applied Energy, Amsterdam: Elsevier, Vol. 370, Nr. 123544, pp. 1–15, doi: 10.1016/j.apenergy.2024.123544.

Applied Energy 370 2024) 123544

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Predictability of electric vehicle charging: Explaining extensive user behavior-specific heterogeneity

Markus Kreft [a,*], Tobias Brudermueller [a], Elgar Fleisch [a,b], Thorsten Staake [a,c]

[a] ETH Zurich, Chair of Information Management, Weinbergstrasse 56/58, Zurich, 8092, Switzerland
[b] University of St. Gallen, Institute of Technology Management, Dufourstrasse 40a, St. Gallen, 9000, Switzerland
[c] University of Bamberg, Chair of Information Systems and Energy Efficient Systems, An der Weberei 5, Bamberg, 96047, Germany

ARTICLE INFO

Dataset link: https://github.com/markus-kreft/ev-charging-prediction, https://www.nationalgrid.co.uk/electric-nation-data

Keywords:
Electric vehicles
Smart charging
Demand response
Demand prediction
Real-world data

ABSTRACT

Smart charging systems can reduce the stress on the power grid from electric vehicles by coordinating the charging process. To meet user requirements, such systems need input on charging demand, i.e., departure time and desired state of charge. Deriving these parameters through predictions based on past mobility patterns allows the inference of realistic values that offer flexibility by charging vehicles until they are actually needed for departure. While previous studies have addressed the task of charging demand predictions, there is a lack of work investigating the heterogeneity of user behavior, which affects prediction performance. In this work we predict the duration and energy of residential charging sessions using a dataset with 59,520 real-world measurements from 267 electric vehicles. While replicating the results put forth in related work, we additionally find substantial differences in prediction performance between individual vehicles. An in-depth analysis shows that vehicles that on average start charging later in the day can be predicted better than others. Furthermore, we demonstrate how knowledge that a vehicles charges over night significantly increases prediction performance, reducing the mean absolute percentage error of plugged-in duration predictions from over 200 % to 15 %. Based on these insights, we propose that residential smart charging systems should focus on predictions of overnight charging to determine charging demand. These sessions are most relevant for smart charging as they offer most flexibility and need for coordinated charging and, as we show, they are also more predictable, increasing user acceptance.

1. Introduction

The share of electric vehicles (EVs) is rising rapidly. In 2020, less than 5% of all new cars sold globally were EVs. Two years later, in 2022, this number already reached 14% and it is projected to rise to 35% by 2030 as the first countries announce bans on new combustion engine vehicle sales [1]. While the increasing number of EVs may lead to reductions in greenhouse gas emissions [2], it also poses challenges for the electrical grid as EVs cause a significant increase in electricity demand. In particular, EV charging during peak periods can cause additional strain on the distribution grid, potentially necessitating costly infrastructure upgrades [3,4]. Consequently, managing demand efficiently becomes a critical concern for grid operators. In this context, smart charging is a promising solution as EVs hold the capacity to contribute to grid flexibility within a demand response (DR) framework [5,6]. The flexibility offered by EVs can serve various applications, including grid congestion management, utilization of local solar generation, and frequency response. Recent research suggests that EV batteries alone could soon meet short-term grid storage needs [7]. Thus, automatic and efficient management of EV charging becomes increasingly important while offering new opportunities for the grid.

Numerous studies have proposed algorithms for optimizing charging schedules (see [8–11] for reviews of algorithms and implementation strategies) and several real-wold smart charging projects already exist (e.g., [12–15]). Such optimization algorithms require information on charging requirements, namely the desired state of charge (SoC) and departure time. In the academic literature, smart charging algorithms are typically evaluated using stochastic charging demand based on surveys [12], specifically engineered distributions [3,16,17], or even a constant, predefined charging process [18]. Multiple software packages exist that can generate mobility data for the evaluation of smart charging algorithms [19,20]. Real driving behavior however, is highly individualized and varies with time and day [21,22], which means that in real-world scenarios, charging requirements need to be individually adapted. For this reason, existing smart charging solutions obtain requirements directly from drivers through smartphone or web

* Corresponding author.
E-mail addresses: mkreft@ethz.ch (M. Kreft), tbrudermuell@ethz.ch (T. Brudermueller), efleisch@ethz.ch (E. Fleisch), thorsten.staake@uni-bamberg.de (T. Staake).

https://doi.org/10.1016/j.apenergy.2024.123544
Received 15 December 2023; Received in revised form 25 April 2024; Accepted 22 May 2024
Available online 7 June 2024
0306-2619/

applications [12–15]. Such an approach, however, places the burden of configuring the smart charging system on end-users, who must balance their comfort with the flexibility they provide to the grid. In this context, maximum flexibility for the grid while still meeting the charging requirements is achieved when users set parameters to match the SoC with their actual driving demand and complete charging just before the next departure. Conversely, maximum comfort for users is achieved when they set the parameters to have the vehicle available as soon as possible.

To aid drivers in these choices, several decision support systems have been proposed [23–27]. In particular, historical driving behavior can inform future requirements. While prior research has addressed predicting EV charging demand in various contexts, there is a lack of comprehensive evaluation across a diverse vehicle set. Most studies focus on public charging stations [28–31], aggregate demand of multiple vehicles [30,32], or long-term forecasting [33,34]. While these predictions can help with grid planning and energy procurement, they exhibit charging patterns that differ from residential at-home charging. A few papers on residential charging have attempted to predict individual requirements [35–38]. However, they use short observation periods, manually collected travel logs, or data from a limited number of vehicles, focusing only on the predictions' impact on vehicle scheduling or battery health. While these predictions are generally difficult, results can differ substantially between different studies and there is a lack of in-depth investigation of the forecasting models themselves. Notably, prior work has not assessed differences in prediction performance across a large variety of vehicles. However, due to the highly individual nature of driving demand [21,22], an investigation of these differences is crucial. The knowledge of the individual prediction performance is essential for understanding and increasing acceptance by drivers and thus to leverage the full load shifting potential of smart charging.

In this paper, we address this research gap by employing several machine learning algorithms to predict individual EV charging requirements. We use a large dataset of real-world residential charging sessions from 300 EVs measured over a period of 22 months. The dataset reflects a variety of vehicles with diverse battery sizes, featuring 37 unique models from 16 different brands. The charging session duration and energy are predicted based on plug-in time and historical charging data. Importantly, we systematically assess the models' performance for each vehicle to understand how individual usage behavior affects charging session predictability.

We observe that the prediction performance varies substantially between different vehicles. An analysis of the characteristics of vehicles with good and poor prediction performance finds that, on average, vehicles that are plugged in later in the day can be predicted better. Based on these observations, we show that knowing whether a vehicle is charging overnight significantly increases performance to a range where it becomes usable in real-world applications. Here, the MAPE of the plugged-in duration predictions of overnight charging sessions is reduced from over 200% when it is unknown if a session lasts overnight, to 15% for sessions that are known to last overnight. Notably, both the greatest potential and the highest demand for smart charging lies precisely in these night-time hours [6]. In this case, vehicles are typically plugged in much longer than they need to charge, and the unmanaged charging times typically fall into the evening peak demand hours. Therefore, we propose to use overnight charging predictions as recommendations for realistic charging parameters in real-world smart charging approaches. Furthermore, our findings can be used to benchmark the performance of smart charging algorithms under uncertainty of future driving demand based on real-world conditions. Finally, the employed dataset and code are openly available, allowing future work to reproduce and extend our analysis.[1]

[1] https://github.com/markus-kreft/ev-charging-prediction

The remainder of this paper is structured as follows: Section 2 gives an overview of related work in the area of EV charging data analysis and demand forecasts. Section 3 introduces the dataset and the forecasting methods. Section 4 presents the results of our analysis of various forecasting algorithms which are discussed in Section 5. Finally, Section 6 concludes this paper with a summary of contributions.

2. Related work

This section provides an overview of relevant literature on predictions of EV charging demand. As this paper deals with smart charging applications, we focus on the short-term prediction of individual charging sessions and not on long-term forecasts, as for example done in [39] for infrastructure planning. Similarly, we do not include work on peak power predictions at transformer station level but only consider charging station demand. While we describe the most relevant studies in detail, Shahriar et al. [40] also review a selection of supervised machine learning methods. Table 1, lists existing literature on predictions in the context of EV charging demand. We categorize the work according to the following four characteristics:

1. Measured: Session data can stem from manually recorded travel logs or automatic measurements from chargers, vehicles, or smart electricity meters.
2. Next session: We distinguish between continuous forecasting of next session times and day-ahead forecasting (occurrence at each future time step).
3. Individual: Related work differs in the use of aggregated charging data from multiple vehicles and sessions from individual vehicles.
4. Residential: Studies can be divided according to whether the charging sessions are recorded in a residential setting or at public charging stations.

Table 1 shows that several studies focus on predicting day-ahead energy demand profiles, typically by using data from public charging stations. Common use cases of public charging station operators are energy procurement and grid management services [30,32]. Since these provide charging opportunities to multiple vehicles at the same time, knowledge about individual charging sessions in this context is not as relevant as aggregated information. Private residential charging, however, requires information about individual upcoming charging sessions instead of aggregated profiles.

Differences in public and residential charging are further present in terms of regularities and patterns within the data. For example, it is heavily discouraged for EVs to rest at public chargers longer than they need to charge to make space for other vehicles [54]. In fact, many station operators charge an idle fee to encourage immediate plug-out (e.g., [55]). Such an incentive is not present for residential charging, which results in different user behavior and charging patterns. For example, Frendo et al. [50] predict the time of day at which a vehicle is plugged out instead of the session duration. However, it assumes that a vehicle is plugged out on the same day it was plugged in. In contrast, such an assumption does not apply to residential charging, where overnight charging is common. While Lee et al. [49] predict the duration and energy consumption of individual sessions, the authors only evaluate the overall test error distribution and not the differences between individual vehicles. Majidpour et al. [28] observe large differences between charging stations prediction results, however only reports average error metrics and not variations.

Given the differences between public and residential charging, in this paper we focus on smart charging of residential customers. Hence, we predict next-session duration and energy demand for individual residential chargers. This task has also been addressed by a few other studies but with significant differences to our work, which are described in the following. Note that, while some of the publications listed below use real-world charging measurements, they also require

Table 1

Categorization of related work on predictions in the context of EV charging. Either data for the **next session** or the day ahead demand is predicted. Data is either **measured** or manually recorded for **individual** or aggregated, **residential** or public charging. Depending on what is reported in a study, the last column shows either the number of sessions (s) or the observation period (t).

	Measured	Next session	Individual	Residential	Chargers	Sessions/Time
Huber et al. [36]	✗	✓	✓	✓	6,465	s = 38,086
Jahangir et al. [35]	✗	✓	✓	✓	256,115	s = 611,342
Aguilar-Dominguez et al. [41]	✗	✓	✓	✓	–	s = 55,236
Graham and Teng [30]	✓	✗	✗	✗	–	s = 103,119
Saputra et al. [31]	✓	✗	✗	✗	58	s = 65,601
Perry et al. [42]	✓	✗	✗	✗	57	s = 47,269
El-Azab et al. [43]	✓	✗	✗	✗	–	–
Buzna et al. [32]	✓	✗	✗	✗	306	–
Majidpour et al. [28]	✓	✗	✓	✗	28	–
Bikcora et al. [29]	✓	✗	✓	✗	2	s = 3200
Ai et al. [44]	✓	✗	✓	✗	1	t = 20 days
Cai et al. [45]	✓	✗	✓	✗	113	t = 11 months
Schwenk et al. [46]	✓	✓	✓	✗	1	s = 2906
Straka et al. [47]	✓	✓	✓	✗	999	s = 189,000
Genov et al. [48]	✓	✓	✓	✗	–	s = 4957
Lee et al. [49]	✓	✓	✓	✗	2	s = 30,000
Frendo et al. [50]	✓	✓	✓	✗	–	s = 100,000
Xiong et al. [51]	✓	✓	✓	✗	30	–
Chung et al. [52]	✓	✓	✓	✓	252	s = 39,458
Chen et al. [37]	✓	✓	✓	✓	5	s = 1008
Phipps et al. [38]	✓	✓	✓	✓	1	s = 2906
Venticinque and Nacchia [53]	✓	✓	✓	✓	1	t = 1 year
Our work	✓	✓	✓	✓	267	s = 59,520

additional data like GPS records of vehicle and driver location tracked through a phone app [45]. Such information can be informative but is infeasible for widespread adoption in existing environments. Especially constant tracking of drivers' positions is unrealistic in most countries and can deter users who place a high value on data privacy. Therefore, our work only uses data measured directly at the charging stations.

A study by Chung et al. [52] investigates the performance of various machine learning algorithms to predict session duration and energy of a mixed dataset from public and residential charging. However, this dataset is dominated by public charging observations and short sessions (very few sessions longer than 10 h). While the study analyzes the correlation between prediction performance and the entropy and sparsity of vehicle charging data, it does not investigate more general, real-world characteristics of the vehicles.

Huber et al. [36] forecast single sessions' duration and subsequent driving demand. The authors use a single week of data based on manually collected travel logs from drivers of conventional internal combustion engine vehicles and assume that users' mobility behavior follows a weekly pattern. Additionally, the study makes use of meta data such as previous parking locations and driver type. Still, the performance of their best model is rather low, with a MAPE of 299% and a median absolute percentage error (MdAPE) of 28.4% for session duration predictions. The next trip distance is predicted with a MAPE of 103% and MdAPE of 45.4%. The large differences between MAPE and MdAPE point out that there may be substantial differences in performance across different sessions and the presence of extreme outliers, whose origins the authors however do not investigate in depth. Especially, they do not analyze the role of individual vehicles. In contrast to this study, we use real-world measurements of charging that stem directly from the at-home charging station and do not require any augmentation.

The work of Aguilar-Dominguez et al. [41] also uses seven day travel diary data, recorded in the UK national travel survey [56]. The study only predicts trip distance, and start and end location, but not duration or departure time, which are crucial for smart charging systems. While the authors report an R^2 value of 0.79 for predicting trip distances, they do not provide any metrics on the prediction error. However, they show the distribution of residual errors which indicates that the errors in predicting trip distance are mostly in the range of tens of miles, but also up to hundreds of miles for some outliers. Similarly, Jahangir et al. [35] use a simple neural network approach to forecast travel behavior from the National Household Travel Survey 2018 [57] but since this survey only collects one day of data per participant, it does not reflect realistic individual charging behavior.

Making use of real-world measurement data, Venticinque and Nacchia [53] extract sessions from smart electricity meter data and predict time of arrival, time of departure, and charged energy. While their dataset contains recordings of more than one year, which may be able to reflect individual charging patterns, it contains only a single vehicle. Thus, the authors do not study the heterogeneity of driving behavior and how it relates to prediction performance. Furthermore, since the method is based on smart meter recordings, it is not possible to observe when the vehicle is unplugged, but only whether it charges or not. However, this differentiation is essential for smart charging because an EV only provides flexibility for the grid when its parking duration exceeds its charging time.

Phipps et al. [38] define a custom performance metric for the prediction of the session duration that quantifies critical undercharging. While their metric is not directly comparable to more common metrics, task performance can be considered rather low. The best model has a MAE outside the 10% prediction interval of 3.27 h at a mean interval width of 1.12 h. Outside the 90% interval, the MAE is only 0.22 h, but the interval has an average width of 14.7 h, which is a rather vague prediction. They evaluate their models on real-world recordings of plug-in and plug-out times over a two year period, but again only for a single vehicle, which does not reflect heterogeneous driving behavior. Furthermore, the study only includes parking durations between two and twenty-four hours. The authors argue that only these sessions are of interest as shorter sessions have low flexibility, while longer sessions have low constraints on flexibility. This, however, ignores how it can at all be determined if a session is particularly long or short.

Lastly, Chen et al. [37] predict EV parking durations to improve battery health by reducing the time the battery is at full capacity. To our knowledge, it is the only study that presents individual prediction performances for different vehicles. The authors use data from the UK's office for low emission vehicles [58] and train individual models for five different charging stations with 150 to 350 sessions each. Even within this small set of distinct charging points, differences in MSE ranging from $0.75\,\mathrm{h}^2$ to $2.35\,\mathrm{h}^2$ can be observed. This confirms the observation from previous studies that the performance between vehicles differs substantially, which means that the charging of some EVs is better predicable than that of others. However, the authors do not examine the details of these variations and instead focus on the impact of charging on battery health. Furthermore, they exclude sessions with durations longer than 40 h without giving a justification how to identify such sessions beforehand.

To conclude, the prediction of session duration and charged energy of EVs is generally a difficult task and differs between individual vehicles, but no work investigates these differences in detail. To the best of our knowledge, no publication systematically evaluates the differences in performance for a large variety of vehicles. Furthermore, few papers use publicly available real-world charging measurements and many studies rely on additional context information being provided. To address these gaps in research, this study relies on a large dataset of real-world recordings to predict the duration and charged energy of individual vehicle sessions. Moreover, we perform a detailed analysis of the differences in predictability by examining the behavior of individual EVs. This way, we contribute to a deeper understanding of EV charging behavior and under which conditions predictions are feasible for real-world applications. Specifically, we investigate the relationship between session flexibility for smart charging and predictability.

3. Methods

To analyze the predictability of charging demand, we train and test a variety of models, and evaluate them in a unified framework. While neural networks and other deep learning methods are used in some studies, simpler models such as Random Forests and Gradient Boosting have shown comparable performance [37,41,44,47]. Since our work focuses on analyzing the properties and variations of predictions, we use these rather simple models instead of more advanced models with higher complexity. Similarly, this paper focuses on point forecasts instead of probabilistic forecasts because they provide more intuitive interpretability. The following subsections explain the details of the dataset, models, and training procedure.

3.1. Dataset

The analyses require a large and heterogeneous dataset of long-term driving behavior with real-world measurements. The dataset provided by the Electric Nation Smart Charging project [12,59], which is publicly available, fulfills these requirements [60]. The dataset contains measurements of 71,264 charging processes from 300 EVs collected at several residential charging stations in the UK over a period of 22 months in the years 2017 and 2018. It reflects a variety of vehicles, featuring 37 unique models from 16 brands. Each session contains data on when a vehicle started and finished charging (charging session) and, most importantly, when it was plugged in and unplugged (plugged-in session). This allows for a clear differentiation between charging and plugged-in duration and thus to estimate flexibility. In addition, the total charged energy of each session and metadata about each vehicle's battery size and charging speed are included.

3.1.1. Preprocessing and data cleaning

To prepare the data for analyses, it is necessary to clean it by removing outlier observations. In a first step, we exclude 511 sessions in which an EV was plugged in less than five minutes. These observations can be considered outliers or erroneous measurements, for which no accurate forecasts are of practical interest. In a second step, we also apply an upper limit by clipping the duration of 23 sessions where a vehicle was plugged in longer than two weeks. Such long sessions can again be considered anomalous events (e.g., holidays or prolonged sickness). For such, there is no practical value in determining the exact plugged-in duration, because scheduling of smart charging takes place over the course of hours, or at most days. However, instead of removing these sessions, we clip their duration at the threshold. This approach emphasizes the importance of predicting that a session is of long duration, while ignoring the exact length beyond the threshold. Next, we exclude 132 erroneous sessions that have an overlap in time with other sessions from the same vehicle, which is practically impossible as a vehicle can only charge once in a single moment. Hence, for each vehicle, the start time of a session must be after the end time of the previous session. To account for this, we additionally exclude the first session of each vehicle such that every session has a previous session. Lastly, a minimum number of sessions is required to correctly infer the underlying charging behavior of an EV. Therefore, we completely remove sessions from 12 vehicles that have less than 10 sessions in total and 20 vehicles with less than one session per week on average. After performing all preprocessing steps, the dataset contains 59,520 sessions from 267 distinct vehicles.

3.1.2. Features

From the cleaned dataset, we calculate 13 features that are listed in Table 2. The choice of this approach and in particular, the types of features are based on related work [36,37]. Importantly, all features are purely derived from charging station measurements and the vehicles metadata, and do not depend on any personal information of the drivers like position tracking or calendar schedule. This makes the predictions suitable for real-world applications. The features can be grouped into three different categories. The *temporal* features are directly calculated from the plug-in time of a vehicle and thus provide temporal context. Note that the time of day (in UTC) of the session start is encoded in a cyclic manner (i.e., as cosine and sine functions of the normalized start time) to better reflect temporal periodicity as is typically done with times of day. This coding ensures a uniform distance between time values, e.g., the distance between 23:00 and 1:00 is equal to the distance between 6:00 and 8:00 [47,61]. The *meta* features describe a vehicle's characteristics and charger type, and thus are constant across all sessions of a vehicle. Lastly, the *historic* features depend on the previous session of a vehicle and thus require knowledge of a vehicles charging history.

3.2. Model selection

We train and test a selection of traditional machine learning models commonly used in literature and compare their performance in predicting duration and charged energy of EV plugged-in sessions. As mentioned earlier, several studies use neural networks for similar tasks but have not achieved a substantial improvement in performance compared to traditional methods [32,36,47]. Therefore, to demonstrate and analyze the performance differences between vehicles we focus on simpler models, as they are easier to train and interpret.

For evaluation, three types of baseline models are used, which can be seen as naive and cheap predictors that the other models have to outperform. The first baseline model always predicts the previously observed value of the same vehicle and thus, in the following, is referred to as *LastValue*. The second baseline model serves as a historically derived estimate and always predicts the mean of the training set. Throughout the remainder of the paper, this model is called *MeanValue*. The third and last baseline model uses the median instead of the mean to put less emphasis on outliers and is called *MedianValue* accordingly.

One category of trained machine learning models are classical regression models, namely *linear regression* and *quantile regression*. Linear regression is one of the most used machine learning methods and commonly applied to all types of point forecasts because it is cheap to train and evaluate and can make use of a variety of features. As

Table 2
Features used for predicting the plugged-in duration and charged energy of vehicle sessions.

Feature name	Category	Description
`weekday`	Temporal	Boolean indicating if plug in is on a weekday
`weekend`	Temporal	Boolean indicating if plug in is on a weekend day
`holiday`	Temporal	Boolean indicating if plug in is on a holiday
`month`	Temporal	Month of the plug-in time
`startHourNorm`	Temporal	Normalized hour of the plug-in time
`startCos`	Temporal	Sine function of the plug-in time (cyclic encoding)
`startSin`	Temporal	Cosine function of the plug-in time (cyclic encoding)
`carKWh`	Meta	Capacity of the vehicle's battery (constant for all session of a vehicle)
`carKW`	Meta	Power of the charger (constant for all session of a vehicle)
`timeSinceLastStop`	Historic	Hours since the end of the previous session
`sessionsToday`	Historic	Number of previous sessions on the same calendar day
`lastDuration`	Historic	Duration of the previous session
`lastConsumedKWh`	Historic	Energy charged in the previous session

linear regression, however, assumes linearity to be present in the data, it may struggle if this is not the case. Quantile regression is an extension of linear regression typically used when its conditions are not met but comes at the cost of higher complexity. The main advantage of quantile regression lies in its improved robustness to extreme outliers by estimating the median of the response variable instead of the mean.

The other category of employed methods are tree models in the form of *Histogram-based Gradient Boosting Regression (HGBR)* [62], which is a tree based regression method inspired by Light GBM [63], a highly efficient gradient boosting decision tree method. Tree models in general have been among the best performing methods with respect to EV session predictions [37,41,44,47].

3.2.1. Training and testing

As typical for any machine learning task, the performance of a model needs to be evaluated on observations that were not seen during training. This requires that the dataset is split into a subset used for training and another one for testing. In the case of charging predictions, however, there is a temporal relationship between the individual charging processes. For example, an EV does not require much charging when its battery is full because it was fully charged in the previous session. Therefore, the split of this data must be chosen more carefully than for conventional regression tasks to prevent biases in the results. Some related studies such as [30,36] randomly shuffle and split the data after showing that the time series of sessions is stationary. Our dataset is stationary across all sessions as given by the augmented Dickey–Fuller test (H_0: Process contains unit root, rejected at $p < 0.0005$ ($\mathrm{DF}_\tau = -43$)), whereas the Kwiatkowski–Phillips–Schmidt–Shin test shows a constant and linear time trend (H_0: Process is weakly stationary, rejected at $p < 0.0005$ ($\mathrm{LM} = 1.3$)). Therefore, we split the data after sorting it temporally across all vehicles. Hence, we use the first 80% of sessions for training and parameter tuning and test the model performance on the last 20% of the data. For the training set, this results in an average number of 178 sessions per vehicles over a period of 299 days. In the test set, there are on average 48 sessions per vehicle over a span of 70 days. This also reflects how data would be available in a practical application scenario, where the order of observed events is chronological.[2] For hyper-parameter tuning, a grid search with 10-fold cross validation is applied to the training set. All experiments are implemented in Python using models from the scikit-learn library [62] and the code to reproduce the experiments is available online.

[2] Note that we also evaluated the performance when leaving individual vehicles completely out of the training set. However, the overall performance was not substantially different from the results presented here and for the sake of readability and simplicity, we do not to show both evaluation strategies.

3.3. Performance evaluation

To make the results comparable to other studies, we apply a set of error metrics commonly used to assess regression tasks of any type. Beyond these common metrics we specifically analyze the distributions of prediction errors and do not only provide average scores (as it will be shown in the Results). This allows to gain a deeper understanding of the methods' performance and to investigate specific predictions. As stated above, a key contribution of our study is the specific investigation of variations in prediction errors to understand how predictions differ across different vehicles and charging behaviors. This way, we show under which circumstances EV session duration and energy can be predicted well and under which not.

One of the reported metrics is the AE, which quantifies the absolute deviation of a prediction $\hat{y}_i$ from the true value y_i. It can be written as follows:

$$\mathrm{AE}_i = \left| y_i - \hat{y}_i \right| . \tag{1}$$

Calculated from the AE, a common metric is the mean absolute error (MAE), which is defined as the sum over all observations i normed by the number of observations n:

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{AE}_i . \tag{2}$$

In the case of highly skewed distributions with large outliers, the median absolute error (MdAE) is a more robust metric than the MAE as the mean is more sensitive to outliers. The MdAE can be written as follows:

$$\mathrm{MdAE} = \mathrm{med}(\mathrm{AE}_i). \tag{3}$$

A main drawback of using metrics related to the absolute error is that they depend on the magnitude of y and thus need to be interpreted in the context of the distribution of y. The APE addresses this shortcoming by normalizing the deviation by y. It is given by:

$$\mathrm{APE}_i = \left| \frac{y_i - \hat{y}_i}{y} \right| . \tag{4}$$

The MAPE and MdAPE are common metrics that can be defined as above. However, it needs to be noted that also the APE has a shortcoming. In case of small y, the denominator becomes close to zero and thus, the APE can become arbitrarily large. Therefore, the symmetric mean absolute percentage error (SMAPE) is another widely used metric that mitigates this problem, as it is limited between 0% and 200%:

$$\mathrm{SMAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{2|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|} . \tag{5}$$

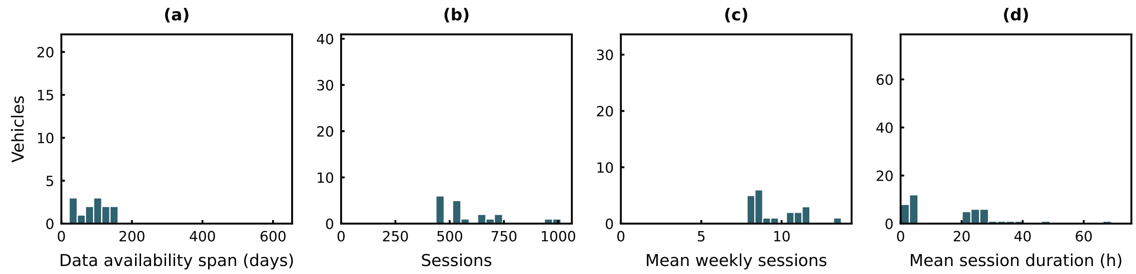
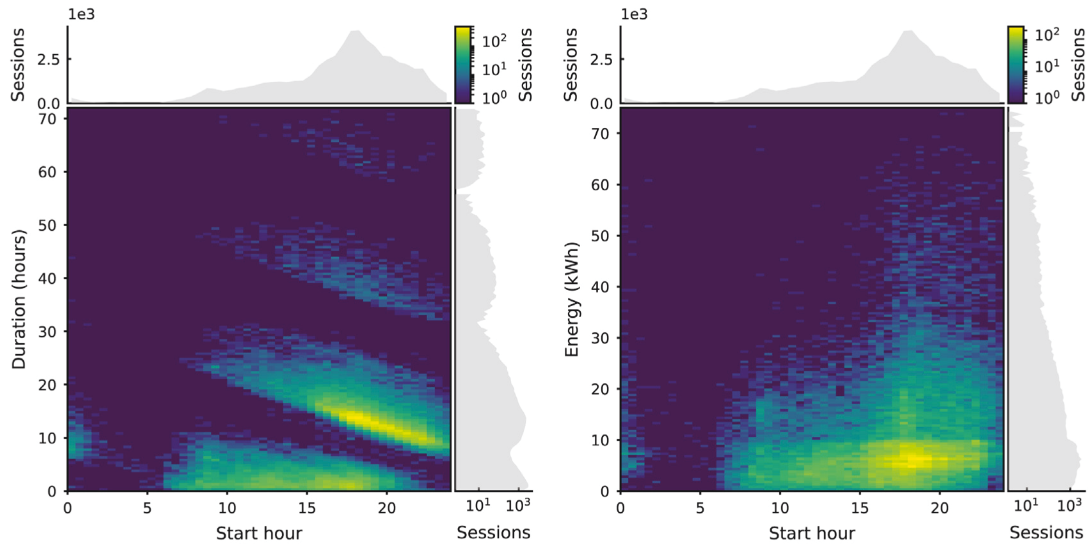


Fig. 1. Distributions of per-vehicle data availability and characteristics. For lack of a significant number of data points, in (d) the duration is cut off at 72 h.



(a) Distribution of session duration over session start. For lack of a significant number of data points, durations are cut off at 72 h.

(b) Distribution of charged energy over session start. For lack of a significant number of data points, energy is cut off at 75 kWh.

Fig. 2. Distributions of the dependent variables over time of day of session start. The x-insert shows the marginal distribution of session start times. The y-insert shows the marginal distribution of the dependent variable. The color map and the y-insert are scaled logarithmically.

4. Results

As explained in Section 2, related work has shown that charging behavior can vary substantially between different EVs. This observation has however not been further investigated. Analyzing these variations in charging and how they affect machine learning models' ability to predict session duration and charged energy is one of the main objectives and contributions of our study. Therefore, we start with a purely descriptive analysis of the variations in "true" charging observations, i.e., how the observed charging sessions differ across EVs. Only afterwards, we evaluate the prediction results of the plugged-in duration and charged energy and how these differ across vehicles and models.

4.1. Descriptive analyses

This section investigates the characteristics of the Electric Nation dataset across different EVs. In this context, we describe variations in data availability, the distribution of sessions' duration and charged energy, and the flexibility potential of charging sessions.

4.1.1. Data availability

The dataset spans a total duration of nearly two years and contains charging sessions of 300 EVs. As vehicles became part of the data collection on a rolling basis, the data availability varies between individual vehicles (Fig. 1a). On average for each vehicle in the dataset, there are 375 days of data and 223 charging sessions available. However, there is a large variety in the number of sessions available per vehicle (Fig. 1b), which also results in a wide distribution of average weekly sessions (Fig. 1c). The mean and standard deviation per vehicle of the latter are 4.2 ± 2.4 sessions/week. Lastly, while the distribution of the mean session duration per vehicle is rather narrow as shown in Fig. 1d (with the mean being 13 h), it features a few extreme outliers. In fact, 90% of the EVs have a mean session duration shorter than 19 h hours and observations beyond this value are very rare. (As described in Section 3.1, we clip session durations at 2 weeks.) To conclude, the Electric Nation dataset used in this study represents heterogeneous charging behavior that differs strongly between vehicles, which is the basis for our work investigating these variations.

4.1.2. Duration and energy of plugged-in sessions

Fig. 2 visualizes the distributions of duration and charged energy of all sessions in the dataset with respect to the start hour of the

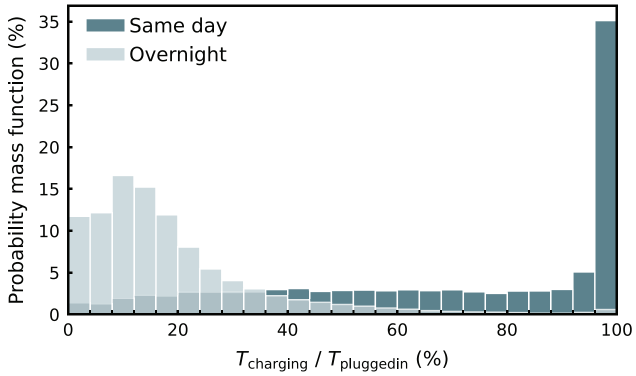


Fig. 3. Distribution of ratios of charging time to plugged-in time for same-day and overnight sessions.

sessions. The color maps are scaled logarithmically to enhance less dominant patterns while preserving visual trends. The histograms on top of the graphs show that the plugged-in sessions rarely start at night (i.e., 23:00–24:00 and 0:00–5:00) and most dominantly occur at times people typically return home from work (around 18:00–19:00). In terms of session duration over start hour (Fig. 2(a)), the color map shows two major clusters. Short sessions typically start during daytime hours, with an especially distinct cluster in the early afternoon. Most longer sessions last around 10 h to 15 h and start in the late afternoon, whereas sessions longer than 20 h are rare. These two clusters correspond to the two main modes of charging: Short "top-up" sessions during the day and long sessions over night. Even more importantly, the graph shows prominent gaps for durations at periods of 24 h. These gaps correspond to the late night hours given the respective start time. Vehicles are rarely plugged out at late night.

For the charged energy with respect to the plug-in time of each session, there are less distinct clusters (Fig. 2(b)). The majority of sessions charge between 5 kWh to 10 kWh, with a drop present at 10 kWh. Sessions with more than 30 kWh charged energy are very rare and predominantly start in the afternoon.

4.1.3. Flexibility of charging sessions

As explained in the introduction, the flexibility potential of charging EVs is closely related to the difference in their plugged-in and charging duration. Therefore, we calculate the flexibility of each charging session as the ratio of the duration that a vehicle charges to the duration it is plugged in [36,64]. Sessions with a low ratio have high flexibility, as charging could be shifted to another point in time while they are plugged in. The distribution of these ratios differs widely between sessions as shown in Fig. 3, which differentiates between "same-day" and "overnight" sessions. The first refer to sessions where plug in and plug out occur on the same day, whereas for the second, vehicles are plugged out on a different day than when they were plugged in. The majority of sessions last overnight (64%) and 79% of vehicles charge more often overnight than within a day. Fig. 3 shows that 35% of same-day sessions require 100% of the time they are plugged in to be charged ("hot unplug"), but most overnight sessions only require 10% to 20% of the plugged-in time for charging. This shows that overnight charging is more prevalent and provides substantially more flexibility. Thus it is more relevant for load shifting.

4.2. Model predictions

This section provides the prediction results of the machine learning models. Table 3 lists the error metrics of the performance on the test set as defined in Section 3.3. Note that the columns differentiate between two sets of performance scores: one set calculated over all sessions and one as the mean of each metric calculated per vehicle. This allows us to investigate the variation of performance across vehicles. The rows refer to the performance of different models with respect to predicting the duration and charged energy. As expected, all trained models (LinearRegression, QuantileRegressor, and HGBR) perform better than the naive baseline models (MedianValue, MeanValue, LastValue), which means that they have learned to generalize. The HGBR model generally achieves the best results. When referring to the performance of all sessions, it achieves a SMAPE of 52.8% for predicted duration and 39.4% for predicted energy. Accordingly, for per-vehicle performance it has a duration-related SMAPE of 52.0% and an energy-related SMAPE of 40.7%. Additionally, Table 3 shows large differences between the mean and median scores, indicating a highly skewed distribution. Therefore, Fig. 4 exemplarily depicts the distribution of AE values with respect to plugged-in duration across sessions (dark color) and MAE scores across vehicles (light color) as given by the best performing model (HGBR). The distribution clearly shows large performance differences between individual sessions and between vehicles. The difference of the MAE between the best and worst vehicle spans a range of 65 h, meaning that the predictability of plugged-in durations for individual vehicles differs widely.

To assess the temporal stability of predictions, we chronologically order each vehicle's sessions in the test set. We then calculate the difference in MAE between the first and the second half of the sessions, normalized by the MAE of all test sessions from the vehicle. The distribution of these deviations has a Gaussian shape (not shown) with a mean of −0.067 and a standard deviation of 0.525. This finding confirms that on average there is no degradation of prediction performance. For individual vehicles, the performance varies over time on the order of half of the AE. Furthermore, the prediction performance does not show any clear seasonal behavior on either the test or the training set. However, the observation period of below two years is too short to apply common statistical tests.

An important part of any machine learning application is the selection and analysis of features. As described in Section 3.1.2, this paper does not focus on feature selection and instead employs features commonly used in literature 2. Nonetheless, Fig. 9 in the Appendix showcases the feature importance for the HGBR using SHapley Additive exPlanations (SHAP) [65]. SHAP values draw on the game theoretic approach of optimal credit allocation to explain the output of any machine learning model. For duration predictions the temporal features (`startSin` and `lastDuration`) are the most important. Conversely, for energy predictions the energy features (`carKWh` and `lastConsumedKWh`) as well as the `timeSinceLastStop` have the highest SHAP values. Importantly, similar to the prediction performance itself, the mean feature importance varies significantly between different vehicles. Hence, while some features may be less important on average, they can still have a substantial impact on the performance for selected vehicles.

Lastly, we visualize the performance of the different models. Fig. 5(a) shows the distributions of AE and APE for individual sessions and per-vehicle MAE as box plots. When comparing the absolute error of the HGBR with that of the best performing baseline model (MedianValue) with respect to plugged-in duration, an independent T-test reveals a significant difference in the mean per-vehicle performance of the models (H_0: Distributions have the same mean, $p = 0.0062$). This finding is confirmed by a Wilcoxon signed-rank test (H_0: Values follow the same distribution, $p < 0.0001$).

The charged energy predictions similarly exhibit substantial differences in performance as reported in Table 3 and Fig. 5(b). Especially the APE shows a highly skewed distribution with the means falling well above the 75% percentile. Since this effect is not as pronounced in the distribution of the AE, it is likely caused by outliers with extremely low charged energy, which results in extremely high relative errors (see Section 3.3). These extreme outliers may be the reason for the baseline MedianModel outperforming the HGBR in terms of MAPE for the energy predictions, as the former is much more robust towards outliers.

Table 3
Performance of the models on the prediction tasks. Error metrics are calculated across all sessions and as the mean of the metrics calculated for the sessions of each individual vehicle.

Task	Model	All sessions					Per-vehicle mean				
		MAE	MdAE	MAPE	MdAPE	SMAPE	MAE	MdAE	MAPE	MdAPE	SMAPE
Duration	LinearRegression	6.45 h	4.05 h	279%	34.5%	61.2%	7.01 h	4.78 h	253%	65.8%	59.2%
	QuantileRegressor	6.08 h	3.06 h	224%	31.3%	59.2%	6.74 h	4.27 h	196%	52.5%	57.8%
	HistGradientBoosting	5.53 h	1.91 h	202%	23.4%	52.8%	6.21 h	3.30 h	181%	37.6%	52.0%
	MedianValue	7.08 h	5.30 h	315%	36.0%	66.5%	7.64 h	5.68 h	280%	78.9%	64.3%
	MeanValue	7.08 h	5.07 h	329%	34.5%	65.8%	7.61 h	5.61 h	292%	81.3%	63.5%
	LastValue	8.93 h	5.02 h	294%	57.1%	78.3%	9.54 h	6.66 h	285%	55.5%	73.5%
Energy	LinearRegression	4.44 kWh	2.56 kWh	300%	32.5%	46.4%	5.45 kWh	4.95 kWh	302%	40.9%	46.8%
	QuantileRegressor	4.34 kWh	2.46 kWh	287%	31.6%	45.4%	5.43 kWh	4.86 kWh	291%	38.4%	46.4%
	HistGradientBoosting	3.83 kWh	1.97 kWh	281%	24.8%	39.4%	4.83 kWh	4.15 kWh	290%	32.6%	40.7%
	MedianValue	6.00 kWh	3.24 kWh	236%	49.3%	59.4%	7.78 kWh	7.29 kWh	219%	50.6%	64.3%
	MeanValue	6.91 kWh	5.40 kWh	355%	64.4%	67.4%	8.02 kWh	7.52 kWh	320%	69.3%	66.4%
	LastValue	5.10 kWh	2.73 kWh	330%	37.2%	54.4%	6.32 kWh	5.26 kWh	343%	40.6%	54.3%

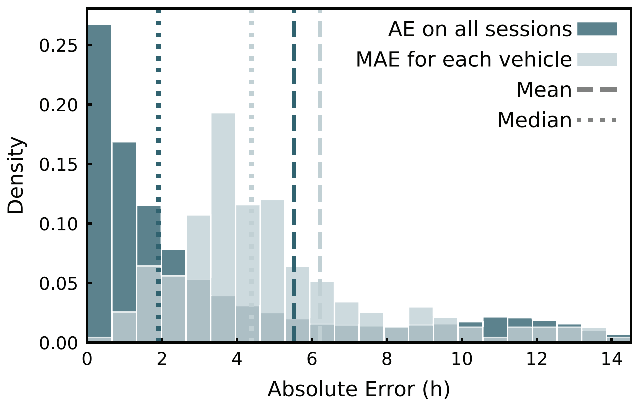


Fig. 4. Distribution of the plugged-in duration AE values for all sessions (dark) and of the per-vehicle MAE (light).

4.3. Analysis of prediction results

In the following, we further investigate the variations in performance between vehicles and the underlying reasons.

4.3.1. Comparison of models

The results presented in the previous section and in the related work show that predicting session duration and charged energy is generally a difficult task. Notably, while the best MAE of 5.53 h for predicting session durations is in the range of multiple hours, the MdAPE of 23.4% slightly outperforms the best results of Huber et al. [36], the only other study that reports an MdAPE. Here it is however important to keep in mind that a direct comparison is not possible due to the use of different datasets. Furthermore, our per-vehicle evaluation demonstrates that for half of the vehicles the MAE is less than 3.3 h, which is similar to what Phipps et al. [38] report for an individual vehicle (3.27 h outside the 1.12 h prediction interval). For the best-performing vehicle, we achieve an MAE below 0.7 h, which is low, even compared to the mean squared error of the best vehicle from Chen et al. [37] (0.75 h^2).

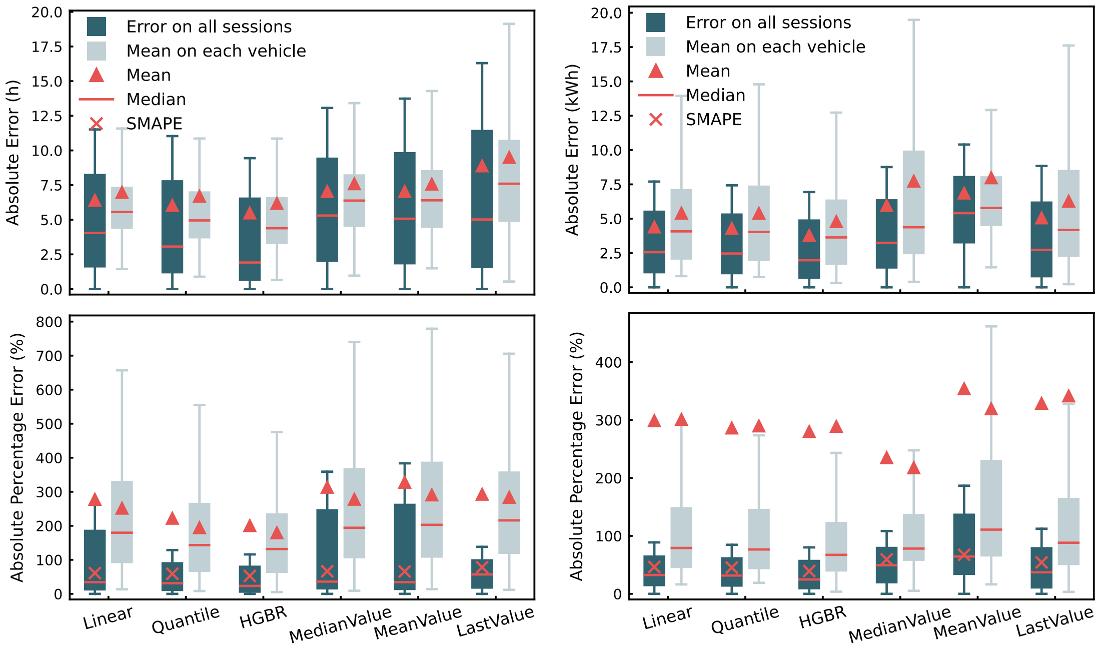
Therefore, our results also confirm the overall difficulty of the prediction task. However, the performance strongly depends on individual vehicles and while the HGBR model performs best, the differences between models and the baseline are smaller than those between vehicles with good and bad prediction performance. This finding is in line with related work, in which the use of more advanced and complex models like deep regression forests or neural networks does not improve performance much [36,47]. When comparing the per-vehicle MAE scores of each model in Fig. 6, we observe generally good agreement between the models with respect to good and bad vehicles. There are no vehicles for which one model performs very well and another model very poorly. All points fall close to the diagonal line, which indicates consensus between the models. This showcases that not the type of model used for prediction leads to variations in performance but the characteristic charging behavior of the individual vehicles.

4.3.2. Comparison of vehicles

The descriptions of the previous section highlight that instead of only focusing on improving model performance, it is necessary to understand the characteristics of vehicles with good and bad predictability. Fig. 10 in the Appendix shows individual time series of the sessions of the six vehicles with best and worst predictability. These examples clearly show that the regularity of the charging behavior has an impact on the prediction performance. To analyze this observation quantitatively, we investigate the prediction performance per-vehicle as a function of vehicle characteristics derived from all sessions of an individual vehicle. A linear fit to this data can determine whether the relationship between a characteristic and the predictability is statistically significant by testing the deviation of the fitted slope from zero (Wald Test). In Fig. 7, scatter plots of the per-vehicle MdAPE values show three types of characteristics: the variation of session duration, the battery size and the plug-in time. The first row refers to the predictions of session duration, whereas the second row refers to the predictions of charged energy. The straight lines in the graphs represent the previously mentioned linear fits, while the additional thin lines show kernel density estimations of the point clouds. The resulting parameters of the linear models are given in the legend, where a refers to the slope, b is the intercept, p the p-value, and r^2 the correlation coefficient. If the slope of the fit is significantly different from zero, the line is colored in red, otherwise in grey. In the following sections, we use these graphs to describe how the three characteristics of a vehicle influence the predictability.

4.3.3. Influence of variation of session duration

To quantify the variation between session durations of each individual vehicle, we calculate their 10 %–90 % inter-percentile ranges (IPRs) and divide them by their medians. This leaves a single metric per vehicle that indicates how high the variation of session durations is — the higher the score, the higher the variation. This type of score can be considered equivalent to the coefficient of variation [66] (the



(a) Session duration **(b)** Charged energy

Fig. 5. Distribution of the AE for all sessions (dark) and of the per-vehicle MAE (light) for the different models. The bottom plots additionally show the distribution of the APE and the per-vehicle MAPE.

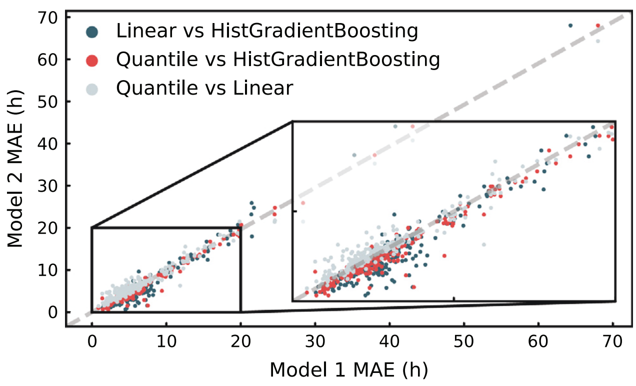


Fig. 6. Comparison of per-vehicle predictability between the different models.

ratio of standard deviation to mean) while being less sensitive to outliers similar to the quartile coefficient of dispersion [67]. The graphs referring to this calculated variation are shown in the first column of Fig. 7. For the prediction of charged energy (Fig. 7d), there is no significant influence of the session variation on performance. This is expected, since EVs being unplugged before reaching the desired SoC ("hot unplugs") is rare and thus we expect no direct dependence of the energy requirements and plugged-in duration. On the other hand, the variation of session durations has a significant impact on its duration predictability (Fig. 7a), which again is to be expected. The larger the variation of session durations, the worse the prediction performance. Hence, if the session duration of an EV is very regular, its predictability improves independent of the plug-in time.

4.3.4. Influence of battery size

Fig. 7e shows that the charged energy of vehicles with smaller batteries can be better predicted. This can be understood as vehicles with smaller batteries generally charging more predictable amounts of energy. A possible reason could be that smaller batteries, which equate to a lower driving range, have to be fully charged more frequently, which in turn is more predictable. In contrast, large battery vehicles have more leeway in how much they need to charge at a given point in time to satisfy the same driving demand. While the dependency is significant, we note that smaller batteries show a higher variation in charged energy. Conversely, in Fig. 7b, we do not observe a significant impact of the battery size on the duration prediction performance. Thus, when comparing vehicles with larger batteries to vehicles with smaller batteries, the amount of charged energy becomes less predictable, while this is not the case for the session duration.

4.3.5. Influence of plug-in time

Most importantly, we find a significant impact of the average hour at which a vehicle is plugged in on the performance of session duration predictions (Fig. 7c). The later a session starts, the lower the MdAPE, as indicated by the slope of the fit smaller than zero ($a = -6.692$, $p = 0.002$).

Further, as discussed in Section 4.1, sessions which start in the evening typically last over night (Fig. 2(a)). Therefore, we hypothesize that overnight sessions are more predictable and test this hypothesis

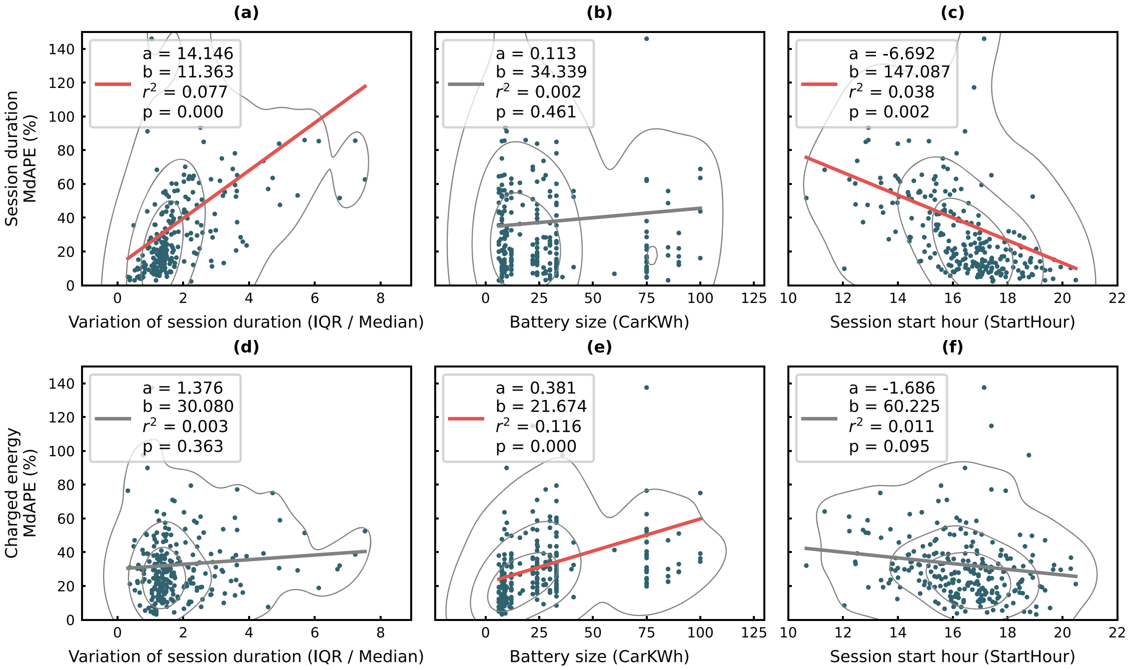


Fig. 7. Dependence of per-vehicle prediction performance on vehicle characteristics. The first row shows results for the duration predictions, the second row for energy predictions.

by splitting the dataset into two parts. One part only includes sessions that end on the same day as they start ("same-day") and the other part includes all the other sessions, i.e., those which always start on one day and end on another ("overnight"). Next, we retrain and test the previously best performing model (HGBR) separately on the day and overnight sessions, following the same procedure as described in Section 3. Fig. 8 reports the corresponding AE and APE values in comparison to the previous model trained on all sessions. Here, a clear difference in performance between the models is visible. For example, the session duration related MAE of same-day sessions is 1.35 h compared to 4.05 h of overnight sessions and 5.53 h of all sessions. This is expected because same-day sessions have a maximum duration of 24 h, which limits the range of possible observations and thus reduces the absolute prediction error. More importantly, there is a substantial improvement in performance for overnight sessions with a MAPE score of only 15% instead of 202% for all sessions. The same tendencies in the prediction of the duration can also be observed for the charged energy, but to a lesser extent. Here, the MAPE is again especially influenced by extreme outliers. We conclude that predicting the duration of overnight sessions separately significantly improves performance. In practice, this would require knowledge about whether a vehicle is needed on the same day or whether it will be plugged in until the next day. We consider this type of information as something a driver could provide when plugging in the EV and discuss possible use cases further in the next sections.

5. Discussion

We have identified three main insights regarding EV charging predictions. Firstly, there are large differences in the prediction performance for different vehicles that we have quantified. Secondly, the

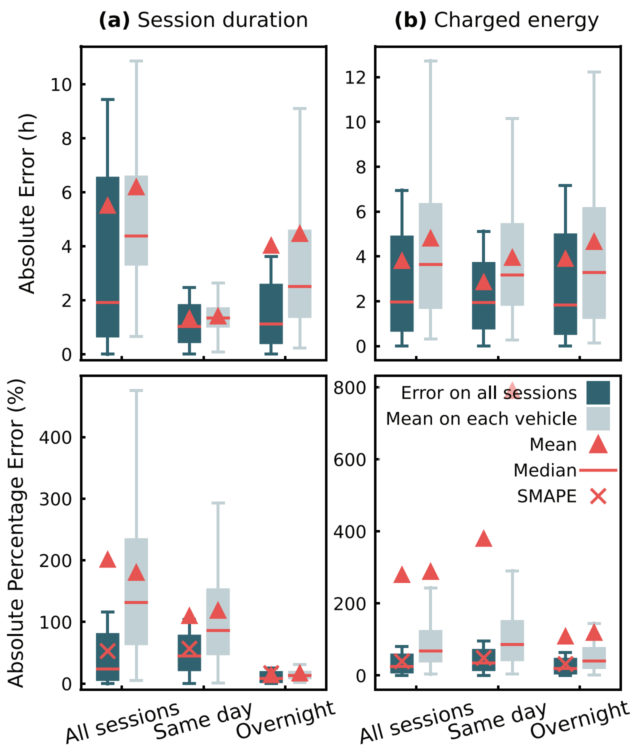


Fig. 8. Comparison of prediction performance for all sessions, those that end on the same day they start, and those that end on another day. The figure only shows data for the HGBR, the best-performing model.

plugged-in duration can be predicted better for vehicles that typically start charging in the afternoon. Thirdly, knowledge that a vehicle charges over night substantially increases prediction performance. In the following, we discuss the real-world implications of our findings and address limitations and potential for future work.

5.1. Contributions and real-world implications

While confirming that predicting the charging duration and charged energy of EVs is overall difficult, we show that performance differs substantially between vehicles. On average, these prediction differences are stable over time. Thus, the value of predictions for smart charging needs to be assessed on an individual level. Furthermore, we uncover that (i) vehicles which start charging later in the day have higher performance for plugged-in duration predictions and (ii) knowing if vehicles will be unplugged on the same day they are plugged in substantially improves the predictions of overnight sessions to a level that makes them usable in real-world applications.

Together, these results make it necessary to rethink the application of charging demand predictions in real-world settings as discussed in the following for the concrete use case of a decision support systems for residential smart charging. Here, predictions could either be used to automatically configure the smart charging parameters (departure time and SoC), or serve as recommendations for drivers. In both cases, grid flexibility is optimized by basing the parameters on the actual demand for a vehicle's availability rather than relying on the perceived need of drivers for the vehicle's availability. In fact, multiple studies have reported that drivers tend to wrongly estimate the SoC that is necessary for their EV to meet driving demand [68–70].

At this point, we recall that overnight sessions play a more important role than same-day sessions for smart charging. As explained in Section 4, (i) 79% of vehicles charge more often overnight than within the same day, (ii) overnight sessions exhibit far greater flexibility potential, and (iii) these sessions predominantly coincide with or cause evening peak loads. Thus, overnight charging is more important to be managed effectively but also offers more load shifting potential. For this reason, we argue that smart charging applications should especially focus on optimizing overnight charging. Based on our findings, we suggest that the default configuration of a decision support system for residential smart charging should be to always initially assume that a vehicle will be charged overnight when it is plugged in. Users should then overwrite this setting either per session or in general if this type of behavior is not desired. Alternatively, in a recommendation setting, users could specify if their departure is on the same or the following day and predictions could only be used to recommend times for the next day. Such a standard configuration would be advantageous for several reasons: Firstly, charging is controlled in such a way that the electricity grid is impacted as little as possible. This can significantly reduce demand peaks and the general load on the grid, and lower users' costs for charging if dynamic tariffs reflect grid congestion. Secondly, as shown previously, under the assumption of overnight charging, the algorithms perform substantially better in predicting charged energy and session duration, better reflecting real demand. This in return leads to better parameters provided to drivers on how to configure the charging system or when charging will be finished. Thus, it may lead to higher acceptance of smart charging decision support systems. We note again that we are in no way advocating taking away users' ability to control charging. Instead, users should always be able to control the charging time themselves, change the configuration or inform their smart charging systems of the desired duration of the session. We simply argue that the assumption that overnight charging as standard is beneficial for the reasons outlined above.

To assess the validity of the proposed approach, two situations in which this assumption would be erroneous must be discussed. The first case concerns users who typically charge overnight but want to charge within a single day for a specific session. Such behavior can be considered an outlier due to special circumstances and thus we assume that users know in advance when the vehicle is needed on the same day, and that they can explicitly inform the system at the moment of plugin. Furthermore, our findings on flexibility (Fig. 3) show that in such cases the departure time often also coincides with the time the vehicle finishes charging. In fact, a user might even decide to plug in the vehicle at a certain point of time precisely because of knowing when the vehicle should finish charging. In any case, in this scenario, predicting charging parameters would be of little relevance to users. The second case affects users who typically unplug their vehicle on the same day they plug it in, which means that our predictions would be wrong most of the times. However, these users are also not good candidates for smart charging in general, because their charging offers very little flexibility. Hence, also in this case a wrong prediction would not substantially change the users expectations. Consequently, the cases in which prediction performance is relevant are those where the assumption does hold. To conclude, it is worthwhile to quantify the prediction performance on only the overnight sessions, while keeping in mind, that cases exist in which smart charging is not viable at all.

5.2. Limitations

While our research is a step towards using predictions of EV charging sessions in real-world applications, it comes with a few limitations. Firstly, the dataset stems from a single EV project in the United Kingdom with participants spread across the British Midlands, South West, and South Wales [59]. Thus charging behavior may be slightly different for other geographic regions. Secondly, the project included periods of smart charging, which could have influenced the observed charging behavior. Especially the third phase of the project was set up to incentivize users to leave their vehicle plugged in as long as possible. Therefore, we followed the study authors suggestions [59] and excluded all sessions from that phase. Nonetheless, some participants may have altered their charging behavior even without specific incentives just due to participating in the project. With the rapidly increasing share of EVs and the increased prevalence of smart charging, more representative data will become available in the future. Thirdly, to limit the scope of analyses, we only employ simple and easy to understand machine learning techniques. While other studies show comparable results when using more complex neural networks, we focus on a detailed investigation of the performance heterogeneity of simpler models. Similarly, we focus on point predictions and do not use probabilistic models. Finally, a deeper exploration of feature selection methods is omitted as this study follows related work and only derives simple features from the vehicle charging data alone. Although advanced features such as GPS data or drivers' calendar schedules would likely improve prediction performance, we intentionally focus on readily available data that is realistic for use in real-world applications. Despite these limitations, our study demonstrates the value of in-depth analysis of predictive models and demonstrates the conditions under which EV charging predictions are feasible for applications.

5.3. Future work

Our study can be extended in multiple ways. Firstly, the analysis should be extended to new datasets. Especially using data from real-world offerings can help to overcome the opt-in selection bias common in voluntary study projects. Additionally, data recorded outside of any smart charging project would show completely uninfluenced charging demand and thus address the limitations discussed in the previous section. Secondly, while our analysis focuses on intuitive metrics to compare the prediction performance between individual vehicles, future work could conduct similar analyses for probabilistic models, which have the potential to capture some of the uncertainty inherent in any prediction. In addition, the potential of training individual models for each vehicle could be expanded. In preliminary studies with a separate model per vehicle, we did not find significant differences in performance compared to our approach of using a single model for all vehicles (results are not shown as they are beyond the scope). However, the differences between individual vehicles and their characteristics could be further explored in this way. Finally, any use of predictions needs to be evaluated in a real application. Therefore, they should be used as configuration parameters or recommendations for departure time and SoC in future smart charging projects. This way the feasibility and relevance of such predictions in the real world can be investigated.

6. Conclusion

This study aims to increase the usability of charging parameter predictions for real-world smart charging applications. We train simple machine learning models to predict the plugged-in duration and charged energy for a large and heterogeneous set of 267 electric vehicles with 59,520 sessions measured over a period of 22 months. While the algorithms achieve similar performance as reported in previous studies, our main contribution is a detailed analysis of the performance differences across a range of vehicles. We find that the strongly skewed distribution of per-vehicle average prediction performance is highly influenced by extreme outliers. For predicting charging session duration, the MAPE of the best model across all sessions is 202% compared to a median of per-vehicle MAPE below 37.6%. Similarly, for charged energy predictions, we find a MAPE of 281% for all sessions versus a median per-vehicle MAPE of 32.6%. These results suggest that studies in the context of predicting EV charging behavior should report a more diverse set of metrics to correctly evaluate and analyze the performance of their predictions, in particular with respect to per-vehicle performance. In fact, the differences in performance between vehicles are much larger than the performance differences between models. Therefore, we analyze the characteristics of vehicles with good and bad predictability, and find that the plugged-in duration of vehicles that typically start to charge in the afternoon can be predicted significantly better than others. These sessions predominantly correspond to overnight charging sessions. A comparison of models trained separately on overnight sessions and same-day sessions, which end on the same day they start, confirms this: When knowing that a session lasts over night, the MAPE is reduced from over 200% to 15%, i.e., more than 13 times. This finding is important when using predictions of plugged-in duration and charged energy to determine a vehicle's charging parameters in a smart charging setting. Importantly, in a smart charging setting, overnight session are both more relevant because they coincide with residential peak demand times in the evening, and offer higher flexibility, as vehicles are typically plugged in longer than they require to charge. Consequently, we propose to always assume overnight charging when using predictions in a smart charging setting and letting customers overwrite the system if this behavior is not desired. As our results are limited to analyses of a single dataset, future work could analyze different datasets and further investigate the impact of larger models and performance differences when using probabilistic predictions.

CRediT authorship contribution statement

Markus Kreft: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft, Writing – review & editing. **Tobias Brudermueller:** Visualization, Writing – original draft, Writing – review & editing. **Elgar Fleisch:** Supervision, Writing – review & editing, Project administration. **Thorsten Staake:** Supervision, Methodology, Writing – review & editing, Project administration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Markus Kreft reports financial support was provided by the Swiss Federal Office of Energy. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

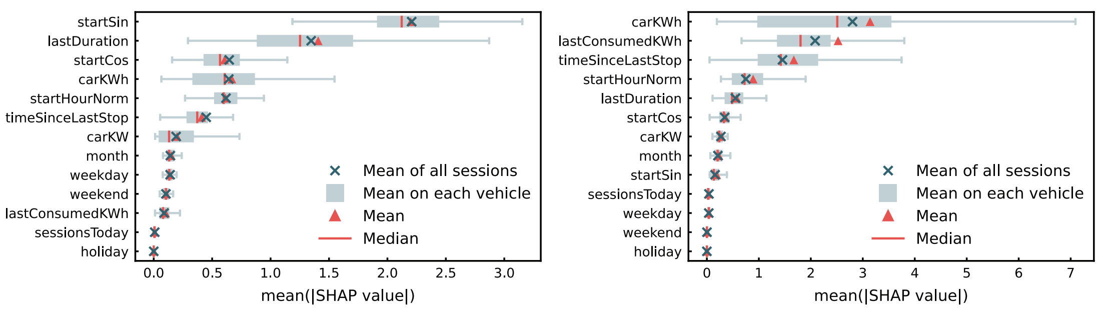
Code available at: https://github.com/markus-kreft/ev-charging-prediction. The openly available, 3rd party dataset comes from the Electric Nation Project [59] and can be downloaded from National Grid Electricity Distribution at: https://www.nationalgrid.co.uk/electric-nation-data.

Acknowledgments

This research is part of a project funded by the Swiss Federal Office of Energy, Switzerland under the grant number SI/502271. We thank David Schaurecker for helpful discussions on data analysis and interpretation.

Appendix

See Figs. 9 and 10.



(a) Session duration **(b)** Charged energy

Fig. 9. Feature importance for the HGBR model. The graphs show mean absolute SHAP values across all sessions as well as the distributions of the means across the sessions of each vehicle.

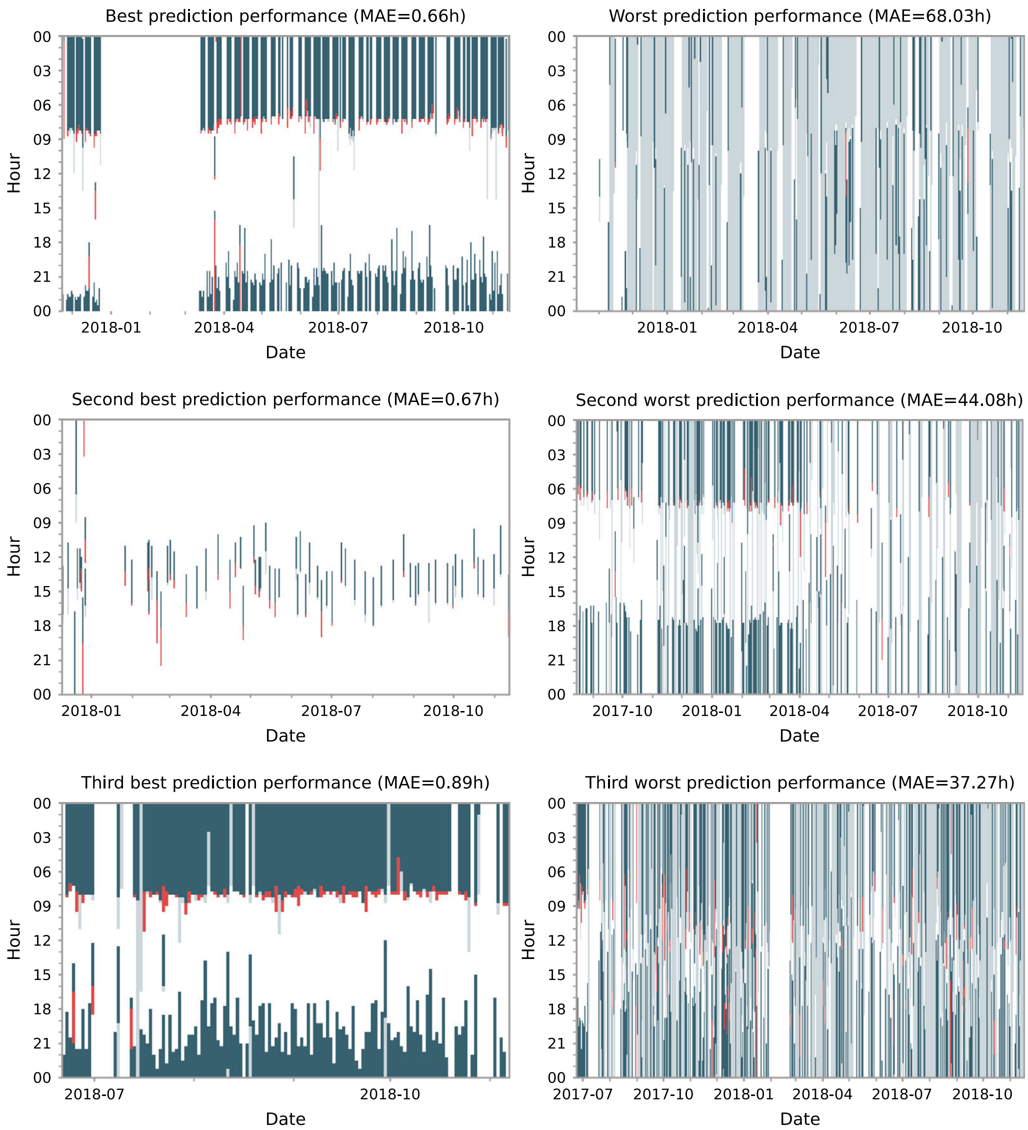


Fig. 10. Time series of plugged-in sessions for the three vehicles with best (left column) and worst (right column) duration prediction performance. The heat map shows the date and time of day for the whole duration of data availability of the given vehicle, including training and testing data. The color encodes if a vehicle was plugged in or predicted to be plugged in (if the point in time represented by the position on the heat map falls into a plugged-in session or not). **White:** true negative (correct prediction that not plugged in). **Dark Blue:** true positive (correct prediction that plugged in). **Light Blue:** false positive (prediction that plugged in, but it was not). **Red:** false negative (prediction that not plugged in, but it was).

References

[1] IEA. Global EV outlook 2023. 2023, https://www.iea.org/reports/global-ev-outlook-2023. [Accessed 31 October 2023].

[2] IEA. Tracking clean energy progress 2023. 2023, https://www.iea.org/reports/tracking-clean-energy-progress-2023. [Accessed 6 November 2023].

[3] Muratori M. Impact of uncoordinated plug-in electric vehicle charging on residential power demand. Nat Energy 2018;3(3):193–201. http://dx.doi.org/10.1038/s41560-017-0074-z.

[4] IEA. Electricity grids and secure energy transitions. 2023, https://www.iea.org/reports/electricity-grids-and-secure-energy-transitions. [Accessed 6 November 2023].

[5] Sadeghian O, Oshnoei A, Mohammadi-ivatloo B, Vahidinasab V, Anvari-Moghaddam A. A comprehensive review on electric vehicles smart charging: solutions, strategies, technologies, and challenges. J Energy Storage 2022;54:105241. http://dx.doi.org/10.1016/j.est.2022.105241.

[6] Lauvergne R, Perez Y, Françon M, Tejeda De La Cruz A. Integration of electric vehicles into transmission grids: A case study on generation adequacy in europe in 2040. Appl Energy 2022;326:120030. http://dx.doi.org/10.1016/j.apenergy.2022.120030.

[7] Xu C, Behrens P, Gasper P, Smith K, Hu M, Tukker A, Steubing B. Electric vehicle batteries alone could satisfy short-term grid storage demand by as early as 2030. Nature Commun 2023;14(1):119. http://dx.doi.org/10.1038/s41467-022-35393-0.

[8] García-Villalobos J, Zamora I, San Martín JI, Asensio FJ, Aperribay V. Plug-in electric vehicles in electric distribution networks: A review of smart charging approaches. Renew Sustain Energy Rev 2014;38:717–31. http://dx.doi.org/10.1016/j.rser.2014.07.040.

[9] Wang Q, Liu X, Du J, Kong F. Smart charging for electric vehicles: a survey from the algorithmic perspective. IEEE Commun Surv Tutor 2016;18(2):1500–17. http://dx.doi.org/10.1109/COMST.2016.2518628.

[10] Mohanty S, Panda S, Parida SM, Rout PK, Sahu BK, Bajaj M, Zawbaa HM, Kumar NM, Kamel S. Demand side management of electric vehicles in smart grids: A survey on strategies, challenges, modeling, and optimization. Energy Rep 2022;8:12466–90. http://dx.doi.org/10.1016/j.egyr.2022.09.023.

[11] Yao Z, Wang Z, Ran L. Smart charging and discharging of electric vehicles based on multi-objective robust optimization in smart cities. Appl Energy 2023;343:121185. http://dx.doi.org/10.1016/j.apenergy.2023.121185.

[12] Dudek E. The flexibility of domestic electric vehicle charging: the electric nation project. IEEE Power Energy Mag 2021;19(4):16–27. http://dx.doi.org/10.1109/MPE.2021.3072714.

[13] Octopus Energy. Learning from intelligent octopus: At-home EV charging behaviours & automation. 2022, https://www.centrefornetzero.org/res/learning-from-intelligent-octopus-at-home-ev-charging-behaviours-automation/. [Accessed 27 June 2022].

[14] CKW. CKW smart charging. 2023, https://www.ckw.ch/gebaeudetechnik/e-mobilitaet/smartcharging. [Accessed 25 July 2023].
[15] Alexeenko P, Bitar E. Achieving reliable coordination of residential plug-in electric vehicle charging: A pilot study. Transp Res D 2023;118:103658. http://dx.doi.org/10.1016/j.trd.2023.103658.
[16] Vayá MG, Andersson G. Smart charging of plug-in electric vehicles under driving behavior uncertainty. In: Karki R, Billinton R, Verma AK, editors. Reliability modeling and analysis of smart power systems. Reliable and sustainable electric power and energy systems management, New Delhi: Springer India; 2014, p. 85–99. http://dx.doi.org/10.1007/978-81-322-1798-5_6.
[17] Vandael S, Claessens B, Ernst D, Holvoet T, Deconinck G. Reinforcement learning of heuristic EV fleet charging in a day-ahead electricity market. IEEE Trans Smart Grid 2015;6(4):1795–805. http://dx.doi.org/10.1109/TSG.2015.2393059.
[18] Xu Z, Hu Z, Song Y, Zhao W, Zhang Y. Coordination of PEVs charging across multiple aggregators. Appl Energy 2014;136:582–9. http://dx.doi.org/10.1016/j.apenergy.2014.08.116.
[19] 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. http://dx.doi.org/10.1038/s41597-021-00932-9.
[20] Strobel L, Pruckner M. OMOD: an open-source tool for creating disaggregated mobility demand based on OpenStreetMap. Comput Environ Urban Syst 2023;106:102029. http://dx.doi.org/10.1016/j.compenvurbsys.2023.102029.
[21] 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. http://dx.doi.org/10.1016/j.etran.2021.100131.
[22] Libertson F. Requesting control and flexibility: exploring Swedish user perspectives of electric vehicle smart charging. Energy Res Soc Sci 2022;92:102774. http://dx.doi.org/10.1016/j.erss.2022.102774.
[23] Haack J, Akyol B, Tenney N, Carpenter B, Pratt R, Carroll T. VOLTTRON™: an agent platform for integrating electric vehicles and smart grid. In: 2013 international conference on connected vehicles and expo. ICCVE, 2013, p. 81–6. http://dx.doi.org/10.1109/ICCVE.2013.6799774.
[24] Valogianni K, Ketter W, Collins J. Enabling sustainable smart homes: an intelligent agent approach. In: Thirty fifth international conference on information systems. 2014, p. 20.
[25] Cheng X, Zhang R, Yang L. Consumer-centered energy system for electric vehicles and the smart grid. IEEE Intell Syst 2016;31(3):97–101. http://dx.doi.org/10.1109/MIS.2016.52.
[26] Kraschewski T, Brauner T, Eckhoff S, Breitner MH. Transformation to sustainable building energy systems: a decision support system. ICIS; 2020, p. 18.
[27] Schwenk K. A smart charging assistant for electric vehicles considering battery degradation, power grid and user constraints. 2022, http://dx.doi.org/10.5445/IR/1000148141.
[28] Majidpour M, Qiu C, Chu P, Pota HR, Gadh R. Forecasting the EV charging load based on customer profile or station measurement? Appl Energy 2016;163:134–41. http://dx.doi.org/10.1016/j.apenergy.2015.10.184.
[29] Bikcora C, Refa N, Verheijen L, Weiland S. Prediction of availability and charging rate at charging stations for electric vehicles. In: 2016 international conference on probabilistic methods applied to power systems. PMAPS, 2016, p. 1–6. http://dx.doi.org/10.1109/PMAPS.2016.7764216.
[30] Graham J, Teng F. Vehicle-to-grid plug-in forecasting for participation in ancillary services markets. In: 2023 IEEE Belgrade powertech. 2023, p. 01–6. http://dx.doi.org/10.1109/PowerTech55446.2023.10202778.
[31] Saputra YM, Hoang DT, Nguyen DN, Dutkiewicz E, Mueck MD, Srikanteswara S. Energy demand prediction with federated learning for electric vehicle networks. In: 2019 IEEE global communications conference. GLOBECoM, 2019, p. 1–6. http://dx.doi.org/10.1109/GLOBECOM38437.2019.9013587.
[32] Buzna L, De Falco P, Ferruzzi G, Khormali S, Proto D, Refa N, Straka M, van der Poel G. An ensemble methodology for hierarchical probabilistic electric vehicle load forecasting at regular charging stations. Appl Energy 2021;283:116337. http://dx.doi.org/10.1016/j.apenergy.2020.116337.
[33] Ashtari A, Bibeau E, Shahidinejad S, Molinski T. PEV charging profile prediction and analysis based on vehicle usage data. IEEE Trans Smart Grid 2012;3(1):341–50. http://dx.doi.org/10.1109/TSG.2011.2162009.
[34] Orzechowski A, Lugosch L, Shu H, Yang R, Li W, Meyer BH. A data-driven framework for medium-term electric vehicle charging demand forecasting. Energy AI 2023;14:100267. http://dx.doi.org/10.1016/j.egyai.2023.100267.
[35] Jahangir H, Tayarani H, Ahmadian A, Golkar MA, Miret J, Tayarani M, Gao HO. Charging demand of plug-in electric vehicles: forecasting travel behavior based on a novel rough artificial neural network approach. J Clean Prod 2019;229:1029–44. http://dx.doi.org/10.1016/j.jclepro.2019.04.345.
[36] Huber J, Dann D, Weinhardt C. Probabilistic forecasts of time and energy flexibility in battery electric vehicle charging. Appl Energy 2020;262:114525. http://dx.doi.org/10.1016/j.apenergy.2020.114525.
[37] Chen Y, Alamin KSS, Jahier Pagliari D, Vinco S, Macii E, Poncino M. Electric vehicles plug-in duration forecasting using machine learning for battery optimization. Energies 2020;13(16):4208. http://dx.doi.org/10.3390/en13164208.
[38] Phipps K, Schwenk K, Briegel B, Mikut R, Hagenmeyer V. Customized uncertainty quantification of parking duration predictions for EV smart charging. 2023, http://dx.doi.org/10.36227/techrxiv.22674997.v1.
[39] Powell S, Cezar GV, Min L, Azevedo IML, Rajagopal R. Charging infrastructure access and operation to reduce the grid impacts of deep electric vehicle adoption. Nat Energy 2022;7(10):932–45. http://dx.doi.org/10.1038/s41560-022-01105-7.
[40] 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. http://dx.doi.org/10.1109/ACCESS.2020.3023388.
[41] Aguilar-Dominguez D, Ejeh J, Dunbar ADF, Brown SF. Machine learning approach for electric vehicle availability forecast to provide vehicle-to-home services. Energy Rep 2021;7:71–80. http://dx.doi.org/10.1016/j.egyr.2021.02.053.
[42] Perry D, Wang N, Ho S-S. Energy demand prediction with optimized clustering-based federated learning. In: 2021 IEEE global communications conference. GLOBECoM, 2021, p. 1–6. http://dx.doi.org/10.1109/GLOBECOM46510.2021.9685647.
[43] El-Azab H-AI, Swief RA, El-Amary NH, Temraz HK. Seasonal electric vehicle forecasting model based on machine learning and deep learning techniques. Energy AI 2023;14:100285. http://dx.doi.org/10.1016/j.egyai.2023.100285.
[44] Ai S, Chakravorty A, Rong C. Household EV charging demand prediction using machine and ensemble learning. In: 2018 IEEE international conference on energy internet. ICEI, 2018, p. 163–8. http://dx.doi.org/10.1109/ICEI.2018.00037.
[45] Cai H, Xin Y, Martin H, Raubal M. Optimizing electric vehicle charging schedules based on probabilistic forecast of individual mobility. AGILE: GISci Ser 2022;3:1–13. http://dx.doi.org/10.5194/agile-giss-3-3-2022.
[46] Schwenk K, Phipps K, Briegel B, Hagenmeyer V, Mikut R. A benchmark for parking duration prediction of electric vehicles for smart charging applications. In: 2021 IEEE symposium series on computational intelligence. SSCI, 2021, p. 1–8. http://dx.doi.org/10.1109/SSCI50451.2021.9660063.
[47] Straka M, Buzna Ľ, Refa N, Mazuelas S. The role of asymmetric prediction losses in smart charging of electric vehicles. Int J Electr Power Energy Syst 2022;143:108486. http://dx.doi.org/10.1016/j.ijepes.2022.108486.
[48] Genov E, Cauwer CD, Kriekinge GV, Coosemans T, Messagie M. Forecasting flexibility of charging of electric vehicles: tree and cluster-based methods. Appl Energy 2024;353:121969. http://dx.doi.org/10.1016/j.apenergy.2023.121969.
[49] 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. Phoenix AZ USA: ACM; 2019, p. 139–49. http://dx.doi.org/10.1145/3307772.3328313.
[50] Frendo O, Gaertner N, Stuckenschmidt H. Improving smart charging prioritization by predicting electric vehicle departure time. IEEE Trans Intell Transp Syst 2021;22(10):6646–53. http://dx.doi.org/10.1109/TITS.2020.2988648.
[51] Xiong Y, Wang B, Chu C-c, Gadh R. Distributed optimal vehicle grid integration strategy with user behavior prediction. 2017, http://dx.doi.org/10.48550/arXiv.1703.04552, arXiv:1703.04552.
[52] Chung Y-W, Khaki B, Li T, Chu C, Gadh R. Ensemble machine learning-based algorithm for electric vehicle user behavior prediction. Appl Energy 2019;254:113732. http://dx.doi.org/10.1016/j.apenergy.2019.113732.
[53] Venticinque S, Nacchia S. Learning and prediction of e-car charging requirements for flexible loads shifting. In: Montella R, Ciaramella A, Fortino G, Guerrieri A, Liotta A, editors. Internet and distributed computing systems. Lecture notes in computer science, Cham: Springer International Publishing; 2019, p. 284–93. http://dx.doi.org/10.1007/978-3-030-34914-1_27.
[54] Office of Energy Efficiency & Renewable Energy. Electric vehicle charging etiquette. 2023, https://www.energy.gov/energysaver/electric-vehicle-charging-etiquette. [Accessed 9 November 2023].
[55] Tesla. Supercharger idle fee | tesla support. 2023, https://www.tesla.com/support/charging/supercharger/idle-fee. [Accessed 9 November 2023].
[56] Department for Transport. National travel survey: 2017. 2018, https://www.gov.uk/government/statistics/national-travel-survey-2017. [Accessed 23 November 2023].
[57] Federal Highway Administration Office of Policy Information. 2017 NHTS data user guide. 2018.
[58] Department for Transport. Electric chargepoint analysis 2017. 2018, https://www.gov.uk/government/statistics/electric-chargepoint-analysis-2017-domestics.
[59] National Grid Electricity Distribution. National grid - electric nation data. 2023, https://www.nationalgrid.co.uk/electric-nation-data. [Accessed 9 November 2023].
[60] Calearo L, Marinelli M, Ziras C. A review of data sources for electric vehicle integration studies. Renew Sustain Energy Rev 2021;151:111518. http://dx.doi.org/10.1016/j.rser.2021.111518.
[61] Brudermueller T, Breer F, Staake T. Disaggregation of heat pump load profiles from low-resolution smart meter data. In: Proceedings of the 10th ACM international conference on systems for energy-efficient buildings, cities, and transportation. buildSys '23, New York, NY, USA: Association for Computing Machinery; 2023, p. 228–31. http://dx.doi.org/10.1145/3600100.3623731.

[62] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12(85):2825–30.

[63] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. LightGBM: a highly efficient gradient boosting decision tree. In: Advances in neural information processing systems. Vol. 30, Curran Associates, Inc.; 2017, p. 3149–57.

[64] Barthel V, Schlund J, Landes P, Brandmeier V, Pruckner M. Analyzing the charging flexibility potential of different electric vehicle fleets using real-world charging data. Energies 2021;14(16):4961. http://dx.doi.org/10.3390/en14164961.

[65] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Advances in neural information processing systems. Vol. 30, Curran Associates, Inc.; 2017, p. 4765–74.

[66] Brian Everitt. The Cambridge dictionary of statistics. Cambridge University Press; 1998.

[67] Bonett DG. Confidence interval for a coefficient of quartile variation. Comput Statist Data Anal 2006;50(11):2953–7. http://dx.doi.org/10.1016/j.csda.2005.05.007.

[68] Franke T, Neumann I, Bühler F, Cocron P, Krems JF. Experiencing range in an electric vehicle: understanding psychological barriers. Appl Psychol 2012;61(3):368–91. http://dx.doi.org/10.1111/j.1464-0597.2011.00474.x.

[69] Eider M, Sellner D, Berl A, Basmadjian R, De Meer H, Klingert S, Schulze T, Kutzner F, Kacperski C, Stolba M. Seamless electromobility. In: Proceedings of the eighth international conference on future energy systems. Shatin Hong Kong: ACM; 2017, p. 316–21. http://dx.doi.org/10.1145/3077839.3078461.

[70] Pevec D, Babic J, Carvalho A, Ghiassi-Farrokhfal Y, Ketter W, Podobnik V. Electric vehicle range anxiety: an obstacle for the personal transportation (R)evolution? In: 2019 4th international conference on smart and sustainable technologies. SpliTech, 2019, p. 1–8. http://dx.doi.org/10.23919/SpliTech.2019.8783178.