Application Independent Flexibility Assessment and Forecasting for Controlled EV Charging

Marcus Voß, Mathias Wilhelm and Sahin Albayrak
DAI-Labor, Technische Universtität Berlin, Ernst Reuter Platz 7, 10587, Berlin, Germany

Keywords: EV Flexibility Modeling, EV Flexibility Forecasting.

Abstract: Electric vehicles (EVs) have been proposed to provide flexibility to the energy grid in various ways. With EV exhibiting very diverse usage patterns on the one hand, and many demand response (DR) schemes and their respective requirements on the other, aggregators of flexibility, as well as operators of controlled charging infrastructure, need models and methods to assess the suitability of specific EVs for specific schemes. In this paper, we provide an application independent flexibility model that allows quantifying the potential amount of flexibility based on a historical dataset. Further, we provide a process to assess the predictability of flexibility through modeling it as a short-term load forecasting problem suitable also for smaller aggregations. Our key findings using real-world data of over 200 charging points are that up- and downwards flexibility per interval have a similar magnitude, but it is unexpectedly low for the high number of charging points. Further, we find that forecast errors are quite high, although we can improve upon naïve benchmarks by almost 20% in mean absolute errors with learning models.

1 INTRODUCTION

In the transition to more intelligent decentralized energy systems with increasing shares of fluctuating and uncertain renewable energy supplies, aggregations of electric vehicles (EV) have been proposed to provide flexibility to the energy system. In this work, we aim at both market DR and physical DR (cf. (Palensky and Dietrich, 2011)). In market DR schemes, EVs are typically integrated through aggregators, and different energy markets have been proposed, such as reserve markets (Goebel and Jacobsen, 2016), day-ahead markets (Vayá and Andersson, 2015) or intraday/imbalance markets (Sortomme and El-Sharkawi, 2012). Here, the EVs may even be physically distributed, but typically within a virtual balancing group or virtual power plant. Physical DR may aim at for instance grid congestion control (Rivera et al., 2015), local load management, power-balance across districts (Yu et al., 2016), or local renewable energy share maximization (Tushar et al., 2016; Hrabia et al., 2015). In this case, the EVs are typically physically close in either a local area network or a part of the distribution grid. Neupane et al. (Neupane et al., 2015) evaluate the value of flexibility, by examining energy regulation markets. In their simulation study they show that even small loads can be aggregated efficiently, and find that if in the Nordic power markets 3.87% of total demand would be flexible, regulation costs can be reduced by 49%.

So with these schemes, on the one hand, there are EVs that exhibit very diverse usage patterns on the other. While some EVs used by private owners may exhibit commuter patterns with quite regular behavior, and long parked times at both work and home, others exhibit quite irregular behavior (Goebel and Voß, 2012). Within research projects, we encountered several different EV use cases, such as stationary and free floating car sharing, and different company cars in *Micro Smart Grid EUREF*, or parcel delivery and nursing service staff fleets in *Smart E-User* (see (Lützenberger et al., 2015) for project descriptions). These and ongoing projects cover private, public and semi-public controlled charging infrastructure, each with a central charging control. Further, EV may be fully electric or plug-in hybrids, which may exhibit different charging patter (Martinez et al., 2017).

So for a specific EV use case, aggregators need methods and tools to assess the potential flexibility of specific EVs and charging stations, to choose a suitable DR scheme. Or with a specific DR scheme in mind, the aggregator could improve an EV portfolio towards specific requirements, e.g. towards high up- or downwards flexibility at certain times. To analyze

Voß, M., Wilhelm, M. and Albayrak, S.
Application Independent Flexibility Assessment and Forecasting for Controlled EV Charging.
DOI: 10.5220/0006795601080119
In *Proceedings of the 7th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS 2018)*, pages 108-119
ISBN: 978-989-758-292-9

this, two dimensions are important: how much flexibility can a fleet provide within specific time intervals and how well can a specific fleet's flexibility be predicted. In this work, we provide an application independent simple flexibility model that allows to quantify the potential amount of flexibility for single intervals retrospectively, based on historical data, as well as a method to assess the predictability through modeling it as a short-term load forecasting problem. The model is suitable for smaller aggregations of EV (e.g., a few tens to hundreds of charging stations/EV), which typically have high uncertainties. Due to the high uncertainties for few stations, as well as typically unavailability of relevant state data (such as the state of charge of battery), our model allows only for control within a specific control band and makes assumptions independent from specific smart control systems.

First, we survey related work in the next section. Then, we present our flexibility model in section 3.1 and the method to assess the predictability in section 3.2. We evaluate the models and methods on a private real-world dataset of 214 charging stations. The dataset used is introduced in section 4.1 and we demonstrate the flexibility model in section 4.2 and evaluate an instantiation of the forecasting process with specific machine learning and benchmark models in section 4.3. Finally, we summarize our contributions and findings in section 5.

2 RELATED WORK

Work on EV flexibility aims at quantifying flexibility typically as a result of scheduling optimization with a specific goal, such as in (Goebel and Jacobsen, 2016), where EV flexibility is analyzed for utilization in pay-as-bid reserve markets and results of a mixed-integer linear program. They evaluate their approach on a simulation study of 10,000 vehicles. However, currently there are no systems in place where aggregators control such large fleets centrally. Hence, we focus on smaller fleets and not on specific scheduling solutions. Goebel et al. treat uncertainties through a schedule repair mechanism, but only make assumptions regarding the departure time distribution (them being normally distributed), but they don't discuss how to predict them. Similarly, Sortomme and El-Sharkawi (Sortomme and El-Sharkawi, 2012) don't base driving behavior on real data, but simple assumptions of EV behavior (i.e. 10% of unexpected leaving during the day and 20% in evening hours). Sundström and Binding (Sundstrom and Binding, 2012) make forecasts of EV usage a central part of their framework, but assume perfect forecasts. In contrast, Bessa et al. (Bessa et al., 2012) more specifically compare how their aggregation framework behaves with forecast errors by comparing a perfect forecast of EV load profiles to a naïve forecast – the same as used in our work as lower bound benchmark – and find that forecast errors increase costs by up to 17 %. Vayá and Andersson (Vayá and Andersson, 2015) demonstrate simulation results to analyze bidding strategies for EV fleets. They find that while for larger fleets (e.g. < 1,000 EV) the error due to uncertainty in driving patterns is neglectable, but for smaller aggregations uncertainty of EV usage is a large issue. Mathieu et al. (Mathieu et al., 2013) show more specifically in one of their two case studies (the other is for air conditioners) how the uncertainty of EV load profile prediction decreases drastically when moving from 1,000 to 100 vehicles. However, their findings are also based on simulated data.

There is not much existing work on real-world EV usage forecasting. In a late broad study on data mining for energy-related time series no EV applications were presented (Martínez-Álvarez et al., 2015). Some work on predicting EV usage was done by deriving conclusions from combustion engine cars, such as data from a GPS-based study with 76 vehicles over a year in Winnipeg. It was used for instance in (Ashtari et al., 2012) to generate charging profiles. Panahi et al. (Panahi et al., 2015) forecast profiles based on generated data which again is based on a study where data was collected manually through driver's logs. Such approaches predict typical driver profiles, which could be and have been used to assess flexibility for market integration. However, they do not conduct actual short-term forecasts. Depending on the application, day-ahead forecasts or forecasts of up to a few hours may be of interest. Hence, for assessing the predictability we need to evaluate how well the short-term behavior can be predicted. For that Goebel and Voß analyzed how well the next departure time for residential drivers (first daily departure time) can be predicted (Goebel and Voß, 2012). There analyses was based on data of the Traffic Choices Study (TCS) collected by the Puget Sound Regional Council in 2005. They found that for a subset of "well-behaved" commuters the behavior is very predictable and offered a method to first cluster the EV based on that property and compared for different subsets how well the departure time can be predicted based on calendar-based features such as the hour of day, day of week and if the day is a workday or not. Similarly, Kirk et al. (Kirk and Dianov, 2015) offer an approach to the same problem, however more from a fleet then from an individual driver's perspective, which is based on a Gaussian modeling approach. They evaluated their approach on data of the NREL's Secure Transportation Data Project, and found that the number of vehicles that will depart in a 15-minute time interval can be predicted with high confidence (around 95%). For commercial infrastructure with small amounts of charging points (<20) it was shown through the example of a campus that for smaller groups of charging stations the amount of plugged-in vehicles is hard to predict (cf. (Hrabia et al., 2015)).

In summary, current work on EV flexibility focuses mainly on very large fleets (a few thousand EV), where uncertainty through usage is not such a big issue. For smaller fleets of a few tens or hundreds of EV however, uncertainty is a very important issue that needs to be considered and it has been shown that costs can increase significantly due to errors. Hence, there is a motivation to improve forecasting (e.g., compared to the presented results of a naive forecast). However, there is not much work yet on EV forecasting. This may be due to a lack of available datasets on EV fleet usage. All the presented studies are based on data for combustion engines, or on simulated data, hence not suitable for actual forecasting evaluation. Further, the studies focus on residential EV, but arguably commercial EV fleets will much earlier be used in aggregator applications, as they are easier to integrate into central control schemes and businesses are more prone to evaluating cost saving potentials through controlled charging.

3 CONCEPTUAL BACKGROUND

3.1 Flexibility Modeling

In general flexibility is the possibility to adapt an electricity consumption profile, e.g. through shifting or simply changing loads. Neupane et al. (Neupane et al., 2015) provide a general taxonomy of flexibility as *time flexibility* and *amount flexibility*. Time flexibility is the possibility of shifting the amount of energy in time, while amount flexibility is the range between the minimum and maximum energy demand at a particular point in time. We believe that for a generic approach the notion of *amount flexibility* matters more, as it is mostly of interest how much load can be changed upwards or downwards for a certain time interval and is therefore the flexibility notion we focus on in this work. Similarly to De Coninck et al. (De Coninck and Helsen, 2013) in the building energy management domain, we want to view flexibility as the possibility to deviate from a *reference scenario* within a certain time interval independently from the possible changes in other time intervals. Inter-dependencies between intervals of course exist in reality, for instance, several intervals may be affected in an optimal schedule. Therefore, viewing only one interval at a time can be seen as a theoretical lower and upper bound, which will, for the purpose of analyzing the potential of a fleet, still give a useful indication. For the EV charging domain, the reference scenario is the *naïve charging* case, where the vehicle charges with the highest possible power for each point of time, as determined by the infrastructure involved. Similar to (De Coninck and Helsen, 2013) and (Neupane et al., 2015), we refer to *positive flexibility* as increasing load for the considered interval and refer to *negative flexibility* as decreasing load, compared to the reference scenario.[1]

Table 1: Flexibility model parameters.

Parameter	Description
l	The length of the equally sized intervals $i \in I$. In the context of energy metering and markets, the intervals are typically fixed at $l = 15$ or $l = 60$ minutes.
S	Set of controllable charging stations for a fleet with parameter p_{min}, which denotes the configured minimum charging power.
C	A historic set of charging events of the form $(t_{arr}, t_{dep}, e_C, p_{max}) : T \times T \times E \times P$.
P_M	An ordered set of tuples $(i, p) : I \times P$ which represents a metered time series of historic average power values.

We calculate the flexibility with the two functions $CalculatePosFlexibility : I \rightarrow P$ and $CalculateNeg Flexibility : I \rightarrow P$. Here, I denotes a set of equally sized time intervals with length l. P models the domain of power values. A variable of this domain models the average power for a given interval. The result of the function is the amount flexibility, that is the maximum possible *change of power* within the considered interval. Table 1 gives an overview over parameters relevant to calculate the flexibility. This flexibility is calculated for a fleet, or more specifically a set of controllable charging stations S of a fleet. To calculate the flexibility, a historic set of charging events C has to be provided as well as a metered time

[1] Note, that an inverted notion is sometimes used for flexibility, mainly when generators are considered.

series P_M in a regularly sampled interval (e.g., 15 minutes or one hour).

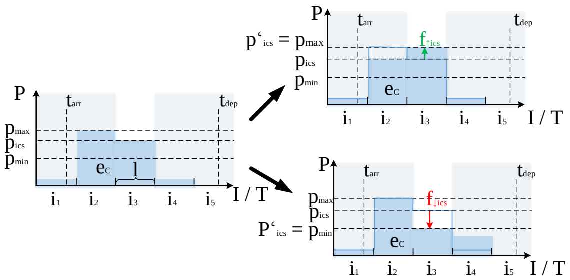


Figure 1: Overview of flexibility concepts for example charging event c.

Figure 1 provides an overview of the nomenclature throughout a typical charging event c. A charging event is bound by the time of arrival t_{arr} (which we assume to be the plug-in time, as well as start of charging) and the departure time t_{dep} (which we assume to be the plug-out time). The curve presents an uncontrolled charging curve of the typical shape, where the load descents towards the end as the inner battery resistance increases. Domain T denotes points in time and E denotes amount of energy. $e_C \in E$ shows how much energy is charged within a charging event. Further, the figure demonstrates how we calculate flexibility from the naïve charging case (represented on the left) for a single charging event in a specific interval. On the right it shows the positive and negative flexibility $f_{\uparrow_i}$ and $f_{\downarrow_i}$ for interval i_3 which is the maximum possible increase (top) or decrease (bottom) in average load within one considered interval i (going from naïve charging power p_{ics} to controlled charging power p'_{ics}, for each charging station s and event c). This value is constrained by two constants: p_{max} and p_{min}, where p_{max} denotes the aforementioned maximum charging power of the naïve case and therefore is defined per charging event and for each interval, as it is determined by the specific car, station and cable used. The charging power must therefore remain below p_{max} for each interval of the charging event. Secondly, we introduce the minimum charging power p_{min} as a constraint, which shall not be undershot within each interval where the vehicle is connected and the battery could still charge. This minimum power comes on the one hand from practical issues, which arise either from the charging infrastructure, for instance EV may go into a standby mode if electric current is below a certain threshold. On the other hand side, it ensures that even if t_{dep} occurs much earlier than expected that at least some energy was charged. With range anxiety being one of the main inhibitors of EV spreading, we therefore desist from allowing the full degrees of freedoms within the control band, which is also the most common type of controlled charging currently in practice. From our prior experience with different e-mobility fleets and their use cases, such as car sharing, home care, delivery or company vehicles, as well as from the current state in the industry, we find that further neither the current state-of-charge (SoC) of the battery, nor the disposition data of the vehicles should be assumed to be known to a central optimization. The reasons range form lack of standard protocols, lack of processes and privacy concerns.

Positive Flexibility. We want to present more in detail how we calculate positive flexibility, the maximum possible increase in average load within one considered interval i which can be calculated independently per charging event c and station s:

$$f_{\uparrow ics} = p'_{ics} - p_{ics}, \quad p'_{ics} > p_{ics} \tag{1}$$

The charging power p'_{ics} is as stated constrained by p_{max} for each considered interval. Therefore, the following constraint must hold:

$$f_{\uparrow ics} <= p_{max} - p_{ics} \tag{2}$$

Further, we defined that the total energy e_c charged within a charging event should remain unaltered. This can be ensured if the amount in the relevant interval i is changed by as much, as the sum of all changes in all other intervals:

$$0 = f_{\uparrow ics} - \sum_{j \in I \setminus \{i\}} p_{jcs} - p'_{jcs} \tag{3}$$

The downward change in all the other intervals is then constrained as stated by p_{min}. Therefore, with equation 3, the change upwards within one interval is constrained as follows:

$$f_{\uparrow ics} <= \sum_{j \in I \setminus \{i\}} p_{jcs} - p_{min} \tag{4}$$

So in summary, the maximum change from p_{ics} to p'_{ics} is constrained by either maximum charging power (equation 2) or the amount that can be decreased within the other intervals (equation 4). So we can define the maximum positive flexibility for an interval i within a charging event c to be the minimum of the two:

$$f_{\uparrow ics} = \min\{p_{max} - p_{ics}; \sum_{j \in I \setminus \{i\}} p_{jc} - p_{min}\} \tag{5}$$

Finally, to calculate the total positive flexibility of all the charging stations:

$$CalculatePosFlexibility(i) = \sum_{s \in S} \sum_{c \in C} f_{\uparrow ics} \tag{6}$$

Negative Flexibility. Similarly, *negative flexibility* $f_{\downarrow_i}$ describes the maximum possible decrease in average load within one considered interval i:

$$f_{\downarrow ics} = p_{ics} - p'_{ics}, \quad p'_{ics} < p_{ics} \tag{7}$$

It is constrained by the minimum charging power:

$$f_{\downarrow ics} \quad <= \quad p_{ics} - p_{min} \tag{8}$$

As for positive flexibility, E_c should remain constant, therefore:

$$0 = f_{\downarrow ics} - \sum_{j \in I \setminus \{i\}} p'_{jcs} - p_{jcs} \tag{9}$$

Also, the charging power must for all other intervals of the charging event remain below p_{max}:

$$f_{\downarrow ics} \quad <= \quad \sum_{j \in I \setminus \{i\}} p_{max} - p_{jcs} \tag{10}$$

Therefore, we can define the maximum negative flexibility for an interval i within a charging event c to be the minimum of the constraints (equations 8 and 10):

$$f_{\downarrow ics} = \min\{p_{ics} - p_{min}; \sum_{j \in I \setminus \{i\}} p_{max} - p_{jc}\} \tag{11}$$

And finally, to calculate the total negative flexibility:

$$CalculateNegFlexibility(i) = \sum_{s \in S} \sum_{c \in C} f_{\downarrow ics} \tag{12}$$

3.2 Predictability Modeling

In addition to the amount of flexibility, we are interested in assessing in a second dimension: how well the amount can be predicted. For most market-targeted and other energy management use cases the relevant horizon for forecasting EV load is from a few hours up to day-ahead which in the load forecasting literature is mostly referred to as *short-term*. As for electric load in general, EV load is mostly depended on calendar-based independent variables (features). Weather-based features may also play a role, but with a lack of real-world data, this is not satisfactorily studied and will most likely vary for different fleets.

We calculate predictability of positive and negative flexibility with the function $CalculateResiduals : (I \to P) \to P$, where the input is the flexibility as described in section 3.1, and the result corresponds to the residuals based on the evaluation of a forecasting model on a test set. The calculation of predictability has the parameters in table 2, which are described in more detail in the following paragraphs. So more specifically, given a time series $F = \{(i_1, p_1), (i_2, p_2), \ldots, (i_n, p_n)\}$, where the values of p are the values of calculated flexibility as described above (either positive or negative), we split the available data at threshold $split$ into a training set of intervals F_T and an evaluation set F_E. Then the features are engineered according to Cov. The forecasting algorithm $\mathcal{A}(\mathcal{H})$ is trained and iteratively applied to produce forecasts of horizon h at forecast time t_{fore}. The forecast $\widehat{F} = \{(i_1, \widehat{p}_1), (i_2, \widehat{p}_2), \ldots, (i_n, \widehat{p}_n)\}$ is then compared against the actual series to calculate residuals $R = \{(i_1, p_1 - \widehat{p}_1), (i_2, p_2 - \widehat{p}_2), \ldots, (i_n, p_n - \widehat{p}_n)\}$.

Table 2: Predictability model parameters.

Parameter	Description
$\mathcal{A}(\mathcal{H})$	Forecasting algorithm to use with hyper-parameters $\mathcal{H}$.
h	Forecast horizon to evaluate forecast on.
Cov	Covariate structure to use.
t_{fore}	Forecast time.
$n_{fold} \in \mathbb{N}$	Number of folds for evaluation scheme.
$split \in [0,1]$	Split of dataset into training and evaluation parts.

Forecasting Algorithms. We treat the problem of forecasting flexibility of small EV fleets similarly to a load forecasting problem for buildings. As shown in the aforementioned survey (Martínez-Álvarez et al., 2015), the most popular algorithms applied are artificial neural networks (ANN), support vector regression (SVR) and multiple linear regression (MLR), and each may perform better on different datasets (e.g., different level of aggregation). We will hence include these as well as other non-linear regression algorithms and compare all against two simple benchmarks which should present the lower bounds. Each of the algorithms has hyper-parameters which influence the performance of the algorithms. We fine-tune the selection of these parameters using grid search and the cross validation scheme as described below. The algorithms $\mathcal{A}$ and their respective most important hyper-parameters $\mathcal{H}$ to tune are the following:

- *Naïve Last Day Type* ($Na\ddot{\imath}ve(\tau)$): A simple benchmark model that predicts the value based on the value of the same time of day from the previous day of the same day type as determined by τ. It is expected to be the lower bound of forecasting accuracy. The parameter τ could be $workday/\neg workday$, $weekday/weekend$ or $day-of-week$.
- *Middle 4–of–6 Day Type* ($Middle4of6(\tau)$): An-

other simple benchmark that predicts the value for an hour of the day based on the mean value of the same hour of the past 6 days of the same day type as determined by τ, ignoring the minimum and the maximum value. The parameter τ could be $workday/\ \neg workday$, $weekday/weekend$ or $day-of-week$.

- *Ridge Regression* ($MLR(r)$): Multiple linear regression with l2 regularization and ridge parameter r.
- *K-nearest neighbors* ($kNN(k)$): A regression scheme that predicts based on the nearest k neighbors. If $k>1$ the mean of the cluster is used.
- *Gaussian Processes* ($GP(l,K)$): A probabilistic regression scheme based on fitting multivariate Gaussian distributions. With a noise level l and a Kernel K with hyper-parameters. Here, the radial basis function (RBF) kernel with parameter γ and the polynomial kernel with maximum exponent e are applied.
- *Artificial Neural Network* ($ANN(\alpha,m,n,(h_i)_{i=1}^n)$): A regression scheme that uses backpropagation to train the weights of a network of nodes that are activated by a sigmoid activation function. The backpropagation algorithm is parameterized with a learning rate α and a momentum m. Further, the amounts of hidden nodes n have to be specified (h_i being the number of hidden neuron on hidden layer i).
- *Support Vector Regression* ($SVR(c,K)$): Support Vector Machine for regression with complexity constant c and a kernel K analogous to the Gaussian Process.

As most of the algorithms are only capable of learning one dependent variable (the class value) at a time, we use an *iterated* (also *recurrent* or *closed-loop*) prediction scheme to produce multiple-step ahead forecasts. That means in order to forecast steps beyond one-step ahead up to horizon h, $p_{t+2},p_{t+3},\dots,p_{t+h}$, the previous forecasted values $p_{t+1},p_{t+2},\dots,p_{t+h-1}$ are used as input.

Feature Modeling. As stated above, assuming no other external available data such as booking times, we can base our prediction mainly on the historical consumption and the dependency on its historic values (*autocorrelation*) and calendar-based features, as well as weather information. Table 3 gives an overview of the proposed features. These should be chosen according to the specific requirements of the application and the specific fleet data as shown in section 4.3 applying cross-validation as described in the following paragraph. Therefore, we add lagged values of consumption to Cov. For very short-term forecasting the lagged consumption of the recent time intervals up to lag_{short} are most important. Electric load forecasting data exhibit typically daily and weekly seasonality, so that any lagged values lag_{med} of the same time of the day within a sliding window of one week may be relevant features. Additionally, some calendar-based variables are included as categorical labels. At minimum the hour of the day hr and the day of the week d should be included. If the data spans several months the month of the year m should be included. If holiday data is available a variable may be used to model workdays and non-workdays w. If available, we propose to add the lagged temperature $Temp_{i-1}$, if the fleet under consideration is in the same region and data is available. For the machine learning algorithms these categorical variables are one-hot encoded (dummy encoded) and all numeric features are normalized (scaled to the interval of $[0,1]$). Depending on the residual characteristics, transformations of the load data may be considered.

Table 3: Proposed input features.

Feature	Definition
Short-term lags	$p_{t-1},p_{t-2},\dots,p_{t-(60/l)\cdot lag_{short}}$
Medium-term lags	$p_{t-(60/l)\cdot 24},p_{t-(60/l)\cdot 48},\dots,p_{t-(60/l)\cdot lag_{med}}$
Hour of day	$hr\in\{0,1,\dots,23\}$
Day of week	$d\in\{1,2,\dots 7\}$
Month of year	$m\in\{1,2,\dots 12\}$
Weekday or not	$w\in\{weekday,\neg weekday\}$
Temperature	$Temp_{i-1}$

Evaluation Scheme. A classic approach to time series forecast evaluation is out of sample (OSS) approach, where a model is fitted (trained) on a first part of the data and then iterativly evaluated on a second part. Per iteration a forecast is produced and error metrics are calculated. When using highly non-linear predictive models n-fold cross-validation (CV) is a common good practice to evaluate the generalization of the model to an independent dataset. However, the general scheme of splitting up the dataset randomly into training and evaluation parts has been shown to be problematic for time series data, as the folds may no longer be i.d.d. (cf. (Bergmeir and Benítez, 2012)). However, a "blocked" CV scheme, where large parts of the data remain together, may still improve upon the OOS scheme and could be used. We propose to utilize a time series specific blocked cross-validation scheme (TSCV), where at any moment only blocks of the past are used to train the model, which basically simulates forecasting operation, were future blocks would not be available. The size of each block could be any length, but should best be chosen according to typical calendar intervals, such as days, weeks or months, depending on the size of the dataset, the forecast horizon h and the highest lagged value lag_{med}. More specifically each block should be considerably larger than the sum of the highest lag and the horizon.

Evaluation Metrics. To assess the predictability visually, mostly the actual residuals are used, as this absolute measure allows for better comparison with the flexibility results, and is most intuitively interpretable from the application side. Therefore, we propose to report the mean absolute error (MAE) if scores are needed. However, to compare performance of the algorithms on one specific dataset we propose as is common, to use root mean square error (RMSE). To compare the performance across datasets scale independent measures should be reported. Due to the intermittency of the flexibility time series and the many actual values of zero, the relative measure mean absolute percentage error (MAPE) is not defined. For the symmetric MAPE (sMAPE), also the denominator tends to be too close to zero, making the calculation unstable (cf. (Hyndman, 2006) for a discussion of forecasting error metrics). We propose to use the scale independent metric coefficient of variation of the RMSE, the CV(RMSE), which is defined as follows, with $\bar{F}$ the mean of the flexibility values:

$$CV(RMSE) = \frac{RMSE}{\bar{F}} \tag{13}$$

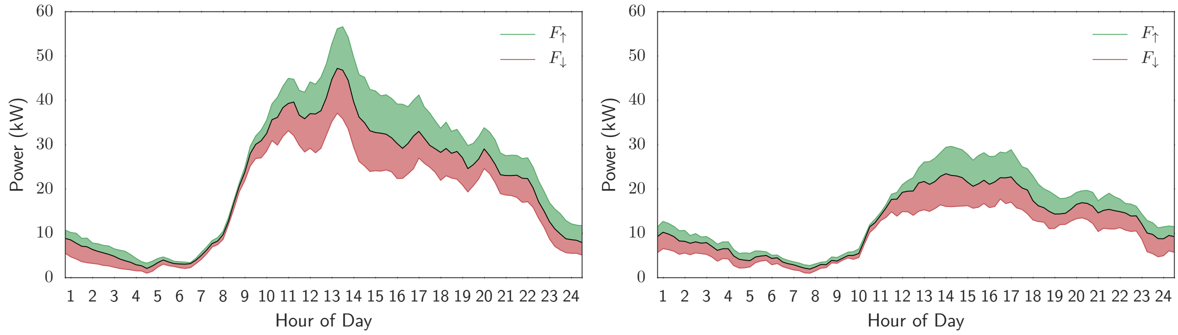
4 EXAMPLE CASE STUDY

4.1 Dataset

We want to give an example assessment of flexibility and its predictability using the introduced models within a case study. We have no *complete* suitable real-world dataset at hand, but we merged two private real-world datasets *Charging Events* and *Charging Metering* and add some additional static information based on suitable assumptions (*Charging Station Information*). The first contains data of 214 different charging points operated by a German distribution grid operator over the course of a year (2013) in two different German cities. It contains event-based information for in total 25,462 charging events (corresponding to C), most interestingly for us plug-in and plug-out times, as well as the amount of energy charged per charging event (t_{arr}, t_{dep}, e_C). The second dataset consists of 497 metered charging curves which we collected within different research projects of the *International Showcase of Electric Mobility (Berlin-Brandenburg)*, and are hence of a variety of different types of mostly unknown vehicles. These charging curves represent P_m and are associated to the charging events so that it best fits to the charged amount e_c of the charging event. In absence of any recorded information regarding p_{max} (which could be communicated via charging protocols), here the maximum metered value of an event is assumed to be the maximum. Third we needed to add some static information to the charging stations, most importantly the minimum charging power allowed p_{min}. Here we used different thresholds $p_{min} = 1.2kW$, $p_{min} = 2.4kW$ and $p_{min} = 4.2kW$. 1.2 kW and 4.2 kW correspond to about 6 Ampere in 230 V/1-phase and 400 V/3-phase systems. These are to typical technical threshold in mode 2 systems. We also compiled a mixed set with 1.2 kW and 2.4 kW, where the thresholds where assigned randomly as it best fits the data. This mixed set is used below, when no other information is given.

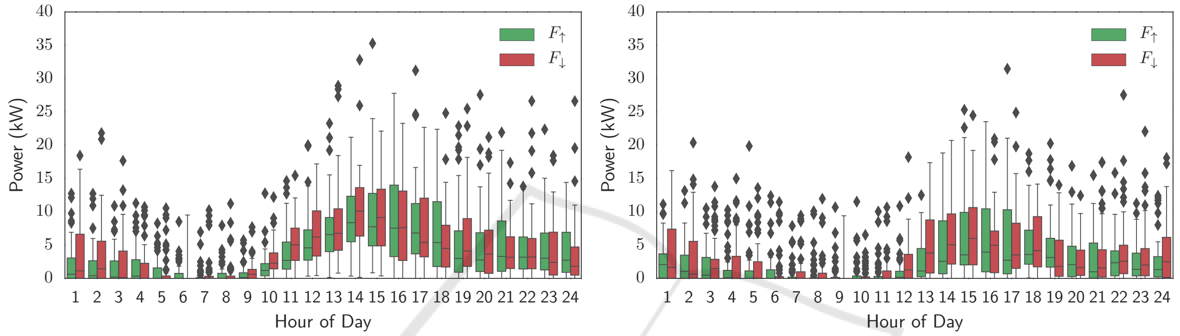
4.2 Flexibility Case Study

Figures 2a and 2b demonstrate how positive and negative flexibility compare to the whole available load. They are the mean positive and negative flexibility for all Mondays (fig. 2a) and all Sundays (fig. 2b) in the dataset, so over a year and the mixed variation for p_{min}. It becomes obvious that for our dataset most flexibility is available from around noon, where also average loads reach a maximum of around 50kW, and into the evening where loads are decreasing. Here it seems that the data does not come from private households, where one would expect larger peaks around 6-8 p.m. On Sundays the load is a lot lower peaking a little later than on Mondays at 20kW. Loads and flexibility are quite low in the middle of the night, around 3-6 a.m., which is the time when the connected vehicles are all charged to their maximum capacity. Figure 2 shows the distribution of positive and negative flexibility per interval for all Mondays (figure 2c) and all Sundays (figure 2d) in the dataset, so over a year and the mixed variation for p_{min}. It becomes obvious that for our dataset most flexibility is available from around noon, where also average total loads reach a maximum of around 50 kW (not shown here due to space issues), and into the evening where loads are decreasing. Here it seems that the data does not come from private household charging stations, where one would expect larger peaks around 6-8 p.m. On Sun-



(a) Mean pos. and neg. flexibility in comparison to total load for all Mondays.

(b) Mean pos. and neg. flexibility in comparison to total load for all Sundays.



(c) Distribution of pos. and neg. flexibility per interval for Mondays ($l = 60min$).

(d) Distribution of pos. and neg. flexibility per interval for Sundays ($l = 60min$).

Figure 2: Flexibility case study results.

days the load is a lot lower peaking a little later than on Mondays at 20 kW (also not shown). Loads and flexibility are quite low in the middle of the night, around 3-6 a.m., which is the time when the connected vehicles are all charged to their maximum capacity. The highest flexibility is also around the early afternoon and it lags the highest value of power by few hours. Also, positive flexibility is just slightly higher than negative flexibility, at around 7-8 kW possible increase or decrease and is lowest in the hours between 3-5 a.m. with mean flexibility close to 0 kW. It is further interesting to see: while the mean is quite similar for Mondays between 1 p.m. and 5 p.m., the variance is increasing in these intervals. When compared with the loads, it becomes obvious that for our data the mean flexibility for this fleet is only slightly higher on Mondays as compared to Sundays, while the difference in total load is much more significant.

4.3 Predictability Case Study

As for flexibility, we will evaluate the predictability modeling on the real world dataset as described above in section 4.1. We want to compare all the different forecasting algorithms $\mathcal{A}(\mathcal{H})$. If not otherwise stated, the analyses are done on the full set of 214 charging stations with the mixed p_{min} and $l = 60min$. The forecasting is implemented in JAVA and is based on the WEKA[2] machine learning library to implement the machine learning models, as well as the benchmark models described in section 3.2.

Figure 3 shows the yearly profile of hourly values of positive and negative flexibility, as well as the weekly profile of a random week. It can be seen that there are as expected clear daily and weekly patterns. Most days have peaks around the early afternoon and in the evening. There is no obvious strong yearly pattern for our data. We analyzed autocorrelation and found that there is significant correlation in the lags preceding the current time, as well as in all lags that are multiples of 24 hours for both positive and negative flexibility. We therefore include the lagged consumption of the last 3 hours which is most valuable for short-term forecasts and then the multiples of 24 hours within a window of a week. Unfortunately we cannot easily associate the dataset with weather information, as the charging station locations are within at least two different Ger-

[2] http://www.cs.waikato.ac.nz/~ml/weka/

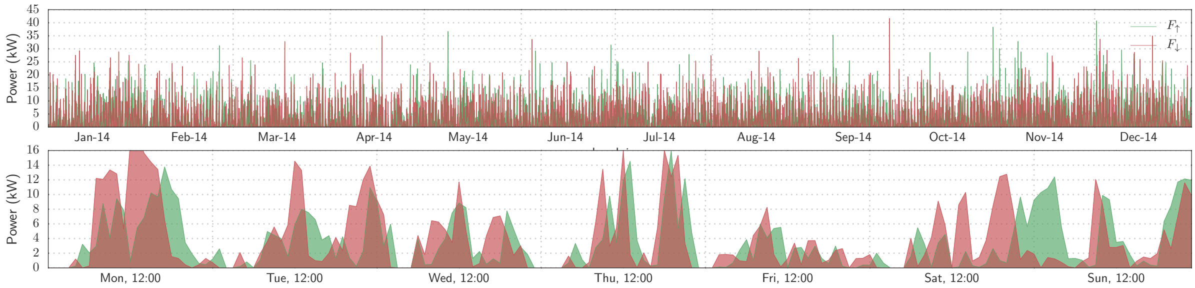


Figure 3: Yearly (top) and weekly (bottom) profile of positive and negative flexibility.

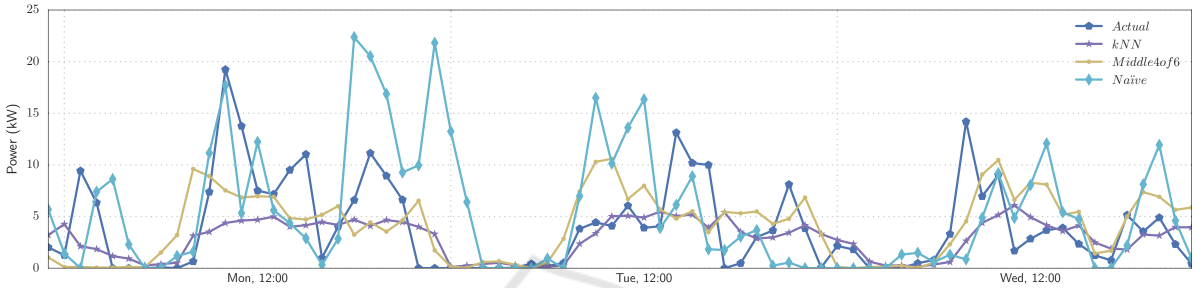


Figure 4: Comparison of forecasts.

man cities. For similar reasons we could not associate the data with local holiday information, which could also be done in future improvements. Further we add the hour of the day, the day of the week and the month of the year as additional covariates, so $Cov = \{p_{t-1}, p_{t-2}, p_{t-3}, p_{t-24}, p_{t-48}, p_{t-72}, p_{t-96}, p_{t-120}, p_{t-144}, p_{t-168}, hr, d, m\}$.

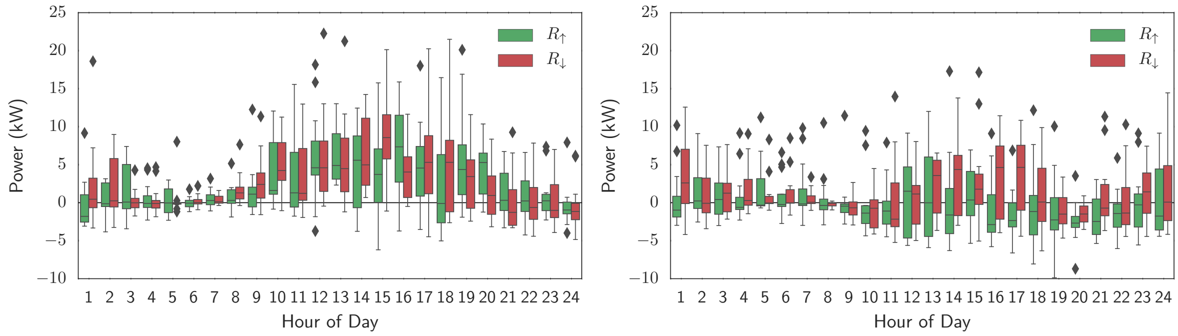
In a first run we compared the performance of the algorithms on the raw dataset, for day-ahead forecasting with $h = 24$, $t_{fore} = 00{:}00$. When analyzing the residuals, we found they exhibit a very skewed distribution with high heteroscedasticity and non-zero mean. Changing the complexity parameters of the learning models could weaken the high bias slightly, but usually at the cost of higher RMSE. Therefore, we decided to apply the models on transformed data. We explored therefore log transformations, the inverse hyperbolic sine transformation, as well as the Box-Cox (*box*) transformation, which we found improved residuals the most, while not considerably worsen forecast results. The Box-Cox transformation is defined as follows:

$$p' = \begin{cases} \frac{(p+C)^{\lambda}-1}{\lambda}, \text{if} \lambda \neq 0, \\ ln(p+C), \text{if} \lambda = 0 \end{cases} \tag{14}$$

Here, $C = 1$ is a shift parameter to avoid logarithm of zero, λ is a form parameter which is fitted to the data based on a likelihood function. We fitted it in one variation to the whole data and in another to the data excluding zero values, which resulted in $\lambda = 0.21$ for the first and $\lambda = 0.3$ for the second. The latter variation improved residual properties the most, so that the distribution of the residuals only exhibited a slight skew and slightly fat tails, but was other than that close to normal using histograms and QQ-plots (not shown for spatial reasons), which would make most importantly fitting prediction intervals in this transformed domain a lot easier. However, for both the Kolmogorov-Smirnov and Shapiro-Wilk test, the hypothesis of a normal distribution has to be rejected at p-value 0.05. In the following, all analyses are done on *box*-transformed data with $\lambda = 0.3$.

Table 4 reports results for the comparison of all the algorithms on the transformed data for day-ahead forecasting, with $n_{fold} = 3$, $split = 0.75$ and as before $h = 24$, and $t_{fore} = 00{:}00$. Here, the learning approaches can as expected considerably improve upon the *Naïve* benchmark model, with the best model improving the MAE by 19.7% and RMSE by 19.3%. *kNN* and *MLR* were performing most robust and with best-achieved error values and were quite fast with a second or less for training and just milliseconds for forecasting on a laptop with 2.69GHz and 12GB RAM. *GP*, *SVR* and *ANN* which are more complex algorithms were very prone to overfitting and are hence often performing worse on the evaluation sets. This becomes for instance obvious when looking at the best performing kernel, the poly-kernel with low exponents of 2 and 3.

Figure 4 shows example traces of the actual values and the predictions on random three days. Here, *kNN* shows exemplarily how the learning approaches



(a) Distribution of pos. and neg. errors per interval for Mondays ($l = 60min$).

(b) Distribution of pos. and neg. errors per interval for Sundays ($l = 60min$).

Figure 5: Predictability case study results.

Table 4: Comparison of forecasting algorithms on *box* transformed data.

$\mathcal{A}$ $\mathcal{H}$		*kNN* $k = 40$	*MLR* $r = 10.0$	*GP* $K = Poly(e = 3)$, $l = 5.0$	*SVR* $K = Poly(e = 2)$, $c = 1.0$	*Middle4of6* *weekday*/ $\neg$*weekday*	*ANN* $\alpha = 0.05$, $m = 0.2$, $n = 1$, $h_1 = 32$	*Naïve* last same day
$P_\uparrow$	MAE	**3501.1**	3505.7	3544.7	3635.6	3690.1	3736.7	4362.1
	RMSE	**5249.7**	5460.5	5647.3	5339.2	5369.8	5965.1	6499.3
	$\bar{d}_{train}$	0.18	1.25	430.50	70.68	0.02	43.29	**0.01**
	$\bar{d}_{pred}$	0.038	0.024	0.06405	0.04019	0.003	0.016	**0.001**
$P_\downarrow$	MAE	**3486.5**	3511.9	3556.7	3635.6	3833.7	3556.6	4323.9
	RMSE	**5523.1**	5529.6	5646.6	5339.2	5687.9	5663.0	6368.0
	$\bar{d}_{train}$	0.20	1.10	487.48	95.56	0.01	64.49	**0.01**
	$\bar{d}_{pred}$	0.042	0.021	0.06037	0.05617	0.003	0.017	**0.001**

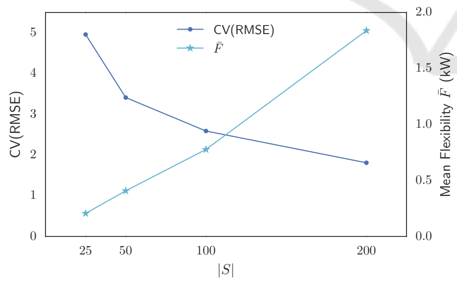


Figure 6: Comparison of different sizes of charging stations.

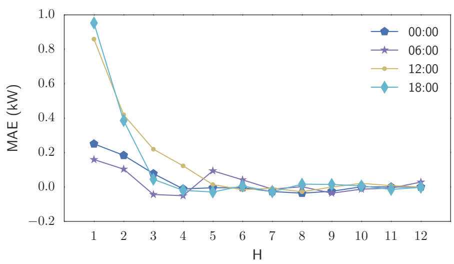


Figure 7: Comparison of errors for different horizons, presented as difference from the day-ahead errors.

tend to underestimate the values more often when compared to the benchmarks, but so avoiding higher penalties due predicting peaks at times, where there are none. One such example is presented on Monday, where the *Naïve* forecaster predicts very high peaks based on the same day one week ago. It also shows that the *Middle4of6* does not forecast the peak, as it averaged out this high peak. It generally predicts higher values than the learning approaches. Figure 5 allows a visual predictability assessment, by showing the errors per hour and for the whole dataset and determined by the *kNN* algorithm for Mondays and Sundays. Here, variability is slightly less for weekends and is typically higher for negative flexibility. Errors are also highest for the time were flexibility is typically highest, around noon and in the evening. The forecasts are less biased on the weekend when compared to the weekdays. Figure 6 shows how the forecast accuracy changes for different sizes of the dataset. Here, four different sizes of datasets were randomly generated (25, 50, 100 and 200). While mean flexibility scales as expected linearly with size, the forecasting accuracy, expressed by CV(RMSE), decreases exponentially, with relatively higher errors for smaller sets. Figure 7 shows how errors behave over different horizons for different times of the day when compared to a day-ahead forecast without any short-term lags. During the day, the forecast can be improved by as much as an average 1kW compared to the day-ahead forecast for the same time, confirming the importance of the short-term lags. This effect is significant for horizons of around 3-4 hours.

5 CONCLUSION

With the goal to implement a generic flexibility assessment framework for EV aggregators to analyze EV fleets and charging stations regarding their flexibility, we developed a flexibility model incorporating our project experience and reflecting current controlled charging strategies in the industry. It does not allow for complete rescheduling of charging events, but only for controlling load within a control band. We note that the model can only be used retrospectively, as e.g. for planning ahead, uncertain data such as arrival times would have to be forecasted (which is current work in project Mobility2Grid), or have to be provided by charging protocolls such as ISO 15118. But as the specific operation would differ anyhow for each use-case, we focused on an application-independent model that would still give a useful indication to a fleet operator about flexibility characteristics of its fleet. For instance, in the case study on real-world data, we found that up- and downwards flexibility per interval have a similar magnitude and lag the peaks in load by few hours. As expected the available amount differs for weekdays and weekends and is highest in the afternoon. We further find that it is unexpectedly low for the high number of charging points. This can be explained by the data being recorded in a year with not yet much electric traffic in the considered cities, and therefore a lot of charging stations with few charging events. Regarding uncertainty, we confirmed that for smaller amounts of EV, uncertainty is quite high. We further found that learning approaches can improve upon the simple benchmark by almost 20% (considering mean absolute errors), and that less complex models such as ridge regression and kNN regression perform more robust than more complex models such as ANN, SVR and Gaussian Processes for hourly data without temperature influences. Currently, there are unfortunately no publicly available datasets on EV fleet and charging station usage. Some work on predicting EV usage was done by deriving conclusions from combustion engine cars. More real-world data is needed to improve research in this area and also to evaluate our approach more thoroughly and for other fleets. To further improve forecasts it could help to train different models for different parts of the day (e.g. each hour, or day and night), the week (e.g., per day, per weekend) or even for different parts of the fleet, as forecasts tend to be very biased in certain situations. This could also be done automatically, by first clustering the data, similar to the approach by Wijaya et al. (Wijaya et al., 2015) for households, or mixture models could be utilized. More features such as other weather variables, or EVSE backend information (e.g. ID of vehicles) could be used. Here, privacy-preserving data-management strategies are an interesting research direction. The flexibility model could be extended for instance to allow calculation for multiple intervals and also to model vehicle-to-grid schemes. The concepts have been implemented as a web-based tool prototype within the research project *EVA service company (Electric Vehicle Aggregator)* to demonstrate how they could be applied by aggregators to optimize their portfolio or choose the right market scheme.

ACKNOWLEDGMENT

The work in this paper has been partially funded from research projects *EVA service company (Electric Vehicle Aggregator)* funded by EIT Digital, as well as project EUREF Forschungscampus Mobility2Grid (FKZ: 03SF0520A).

REFERENCES

Ashtari, A., Bibeau, E., Shahidinejad, S., and Molinski, T. (2012). PEV charging profile prediction and analysis based on vehicle usage data. *IEEE Transactions on Smart Grid*, 3(1):341–350.

Bergmeir, C. and Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213.

Bessa, R. J., Matos, M. A., Soares, F. J., and Lopes, J. A. P. (2012). Optimized bidding of a EV aggregation agent in the electricity market. *IEEE Transactions on Smart Grid*, 3(1):443–452.

De Coninck, R. and Helsen, L. (2013). Bottom-up quantification of the flexibility potential of buildings. *Building simulation, 13th international conference of the international building performance simulation association, IBPSA, Aix-les-Bains, France*.

Goebel, C. and Jacobsen, H.-A. (2016). Aggregator-controlled EV charging in pay-as-bid reserve markets with strict delivery constraints. *IEEE Transactions on Power Systems*, 31(6):4447 – 4461.

Goebel, C. and Voß, M. (2012). Forecasting driving behavior to enable efficient grid integration of plug-in electric vehicles. *Online Conference on Green Communications (GreenCom)*, pages 74–79.

Hrabia, C.-E., Küster, T., Voß, M., Pardo, F. D. P., and Albayrak, S. (2015). Adaptive multi-stage optimisation for EV charging integration into smart grid control. In *International Conference on Principles and Practice of Multi-Agent Systems*, pages 622–630. Springer.

Hyndman, R. J. (2006). Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, 4(4):43–46.

Kirk, N. H. and Dianov, I. (2015). Towards predicting first daily departure times: a gaussian modeling approach for load shift forecasting. *arXiv preprint arXiv:1507.04502*.

Lützenberger, M., Masuch, N., Küster, T., Freund, D., Voß, M., Hrabia, C.-E., Pozo, D., Fähndrich, J., Trollmann, F., Keiser, J., et al. (2015). A common approach to intelligent energy and mobility services in a smart city environment. *Journal of Ambient Intelligence and Humanized Computing*, 6(3):337–350.

Martinez, C. M., Hu, X., Cao, D., Velenis, E., Gao, B., and Wellers, M. (2017). Energy management in plug-in hybrid electric vehicles: Recent progress and a connected vehicles perspective. *IEEE Transactions on Vehicular Technology*, 66(6):4534–4549.

Martínez-Álvarez, F., Troncoso, A., Asencio-Cortés, G., and Riquelme, J. C. (2015). A survey on data mining techniques applied to electricity-related time series forecasting. *Energies*, 8(11):13162–13193.

Mathieu, J. L., Vayá, M. G., and Andersson, G. (2013). Uncertainty in the flexibility of aggregations of demand response resources. In *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, pages 8052–8057. IEEE.

Neupane, B., Pedersen, T. B., and Thiesson, B. (2015). Evaluating the value of flexibility in energy regulation markets. In *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*, pages 131–140. ACM.

Palensky, P. and Dietrich, D. (2011). Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE transactions on industrial informatics*, 7(3):381–388.

Panahi, D., Deilami, S., Masoum, M. A., and Islam, S. M. (2015). Forecasting plug-in electric vehicles load profile using artificial neural networks. In *Power Engineering Conference (AUPEC), 2015 Australasian Universities*, pages 1–6. IEEE.

Rivera, J., Goebel, C., and Jacobsen, H.-A. (2015). A distributed anytime algorithm for real-time EV charging congestion control. In *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*, pages 67–76. ACM.

Sortomme, E. and El-Sharkawi, M. A. (2012). Optimal scheduling of vehicle-to-grid energy and ancillary services. *IEEE Transactions on Smart Grid*, 3(1):351–359.

Sundstrom, O. and Binding, C. (2012). Flexible charging optimization for electric vehicles considering distribution grid constraints. *IEEE Transactions on Smart Grid*, 3(1):26–37.

Tushar, W., Yuen, C., Huang, S., Smith, D. B., and Poor, H. V. (2016). Cost minimization of charging stations with photovoltaics: An approach with ev classification. *IEEE Transactions on Intelligent Transportation Systems*, 17(1):156–169.

Vayá, M. G. and Andersson, G. (2015). Optimal bidding strategy of a plug-in electric vehicle aggregator in day-ahead electricity markets under uncertainty. *IEEE Transactions on Power Systems*, 30(5):2375–2385.

Wijaya, T. K., Vasirani, M., Humeau, S., and Aberer, K. (2015). Cluster-based aggregate forecasting for residential electricity demand using smart meter data. In *Big Data (Big Data), 2015 IEEE International Conference on*, pages 879–887. IEEE.

Yu, R., Zhong, W., Xie, S., Yuen, C., Gjessing, S., and Zhang, Y. (2016). Balancing power demand through ev mobility in vehicle-to-grid mobile energy networks. *IEEE Transactions on Industrial Informatics*, 12(1):79–90.