Inferring Electric Vehicle Charging Patterns from Smart Meter Data for Impact Studies

Feng Li*‡, Élodie Campeau*, Ilhan Kocar†, and Antoine Lesage-Landry*
*Department of Electrical Engineering, Polytechnique Montreal, GERAD & Mila, Montreal, Quebec, Canada H3T 1J4
Email: {feng.li, elodie.campeau, antoine.lesage-landry}@polymtl.ca
†Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China
Email: ilhan.kocar@polyu.edu.hk
‡Eaton's CYME International T&D, Brossard, Quebec, Canada J4Z 0N5

***Abstract*—In this work, we propose a non-intrusive and training free method to detect behind-the-meter (BTM) electric vehicle (EV) charging events from the data measured by advanced metering infrastructure (AMI) such as smart meters. By leveraging the contextual information of EV charging, we formulate a mixed-integer convex quadratic program (MICQP) to detect EV charging events from customers' daily meter data. No labelled training data or hyperparameter tuning are required, and the MICQP can be efficiently solved. By collecting information about the start time, the charging duration, and the power level of each detected charging event, we infer customers' charging patterns in terms of probabilities of charging profiles through a data-driven approach using one year's meter data. In a numerical case study, we use the proposed approach to extract EV charging events from a test dataset of customers' meter data, and we demonstrate that similar detection accuracy is achieved as that of other learning-based approaches which use high-solution meter data. Finally, impacts of EV charging on the IEEE-8500 test feeder are presented in the case study by using the inferred charging patterns.**

***Index Terms*—Convex programming, data-driven, electric vehicle, smart meter, behind-the-meter, distribution networks.**

I. Introduction

With the increasing penetration levels of EVs on the power distribution networks, utility planners need to evaluate the impacts of EV charging on their networks to maintain system reliability and power quality. Due to the randomness associated with the usage of EVs, stochastic methods are usually adopted to analyze the impact by characterizing EV owners' charging habits in terms of probabilities of charging profiles [1]. While such probabilities can be constructed based on historical EV charging events, as EV chargers are often installed behind-the-meter (BTM), the charging events are invisible to utilities. It is hence challenging to directly observe customers' charging behaviours. In this work, we propose a method to detect customers' charging events from the smart meter data, which are readily available to the utilities. For each charging event detected, the start time, the duration, and the power level of the charger is extracted. We then infer customers' charging patterns in terms of probability distributions based on the extracted information out of all charging events detected.

In the literature, there exist mainly two groups of methods to detect EV charging events from the smart meter data: supervised learning-based and training-free methods. Supervised machine learning methods, e.g., convolutional neural networks and recurrent neural networks, have been trained to detect EV charging events with the best average accuracy of 71% [2], [3], [4], [5]. These methods primarily focus on identifying time periods associated with EV charging activities. Supervised classification methods, e.g., random forest algorithm, k-nearest neighbour, and classification-regression trees, are also deployed for qualitative purposes, such as distinguishing EV owners from other consumers [6], [7]. The downside of these methods is that their performance is highly dependent on the quality and quantity of data available for training, i.e., the training samples.

In addition to supervised learning-based methods, training-free methods have also been utilized in EV detection. These approaches take advantage of contextual information (e.g., standardized power level of EV chargers), and for this reason, they are widely applicable to customers' data on different distribution networks and/or regions. Models based on cross-correlation and pattern recognition are proposed in [4], [8]. These models use sliding windows and pattern search techniques to identify EV's charging events. For improved accuracy, filtering techniques are proposed in [9], [10] to specifically remove data segments that do not meet typical values of power levels and of duration of an EV charging event. Signal decomposition is also utilized and takes advantage of trends in measured meter data to target EV signatures [11]. These methods are easily interpretable, but are limited by the assumptions made about EV charging behaviours. More refined models further leverage statistical and probabilistic methods. For example, [12], [13] use independent component analysis (ICA) to detect EV charging events. Probabilistic models such as hidden Markov models (HMM) allow to account for the uncertainty around EV charging profiles in terms of the start time, the initial state-of-charge, duration, or the power level. In [14], [15], the authors used HMM to separate individual appliances from an aggregate load without requiring complete knowledge of the types of appliances in the households. Even though no training is required, hyperparameter tuning is still necessary for optimal performance.

In this work, we develop a method that can accurately detect the EV charging events from smart meter data through a mixed-integer convex program, which can be solved in a computationally efficient way. Our method leverages the contextual information of EV charging, and does not rely on any supervised or unsupervised learning models hence no labelled training samples or hyperparameter tuning are necessary, making it readily implementable for utilities. Through a data-driven approach, we construct the charging patterns of all customers in terms of probability distributions from the detection results. Using the inferred charging patterns, impacts of these customers' EV charging behaviours to the power distribution network can then be analyzed [1].

The rest of the paper is organized as follows: in Section II we state the main assumptions about the contextual information of EV charging. We then formulate a mixed-integer convex quadratic program to detect a customer's EV charging events during a day. In Section III, we present our methodology to infer EV charging patterns for a set of customers on the distribution network, based on the detected EV charging events. The charging patterns are represented by distributions of charging start time, duration, and power levels. We present a case study to illustrate the EV charging patterns extracted by our proposed approach. We then showcase the quality and accuracy of the inferred patterns by using them in a stochastic analysis of EV impacts to a test distribution network in Section IV. Finally, we conclude in Section V and point out some future work directions.

II. Detection of EV charging events

Let $\mathbf{P} = [\mathrm{P}_i]_{i=1}^N \in \mathbb{R}^{N\times N_{\mathrm{D}}T}$ be the real power data measured in kW by the smart meters for a set of N customers and for N_{D} days, where T is the time horizon or the number of data samples during a day. Here, we assume that $\mathrm{P}_i \in \mathbb{R}^{N_{\mathrm{D}}T}$ is measured at a 15-minute interval, hence $T = 96$. Let $\mathbf{P}^{\mathrm{EV}}, \mathbf{P}^{\mathrm{BL}} \in \mathbb{R}^{N\times N_{\mathrm{D}}T}$ refer to, respectively, the power demands to charge customers' EVs and those of the baseload, i.e., all other appliances and devices in the households. Hence, we write $\mathbf{P} = \mathbf{P}^{\mathrm{EV}} + \mathbf{P}^{\mathrm{BL}}$. Our goal is to disaggregate $\mathbf{P}^{\mathrm{EV}}$ from $\mathbf{P}$ through a detection procedure for EV charging events.

A. Contextual information

In this section, we state several assumptions based on characteristics of EV charging events and results of statistical studies. In this work, we consider only the residential customers, as they account for up to 80% of charging events with diversified and stochastic charging patterns [16].

- The charging power levels depend on the vehicle models and the type of chargers. For residential customers, Level 1 or Level 2 chargers are usually used. As the time required for a full-charge by Level 2 chargers is shorter and the associated costs to install them keep decreasing over time, they are becoming more popular among residential customers [17]. Hence, the set of possible charging power levels depending on vehicle models is assumed to be $\mathcal{P} = \{3.6, 6.6, 7.2, 9.6, 11.5, 15\}$, in kW.
- Based on a statistical study on Pecan Street data [18], the duration of charging events is usually more than 30 minutes, and the number of charging events during a day is usually less than 3 [11]. In another statistical study on HVAC systems, the average duty cycle of an HVAC system is about 30 minutes [19]. In order to differentiate EV charging events from HVAC duty cycles, in this work we ignore charging events that last less than 1 hour.
- We assume that the profile of a charging event follows a rectangular waveform, i.e., the power consumption is constant during the entire charging period [20], and is voltage independent. The variations in power consumption are ignored when the charging starts and completes, as well as due to voltage fluctuations during charging. As shown later in the case study section, the assumption is valid, e.g., when comparing the waveform of the detected EV charging event with the true waveform.
- According to a survey on multi-vehicle households [21], 90% of EV owners have more than 1 vehicle but only 1% of them own another EV. Hence, we assume that only 1 EV is charged each day, even for households with multiple EVs. Hence, all charging events in a day should have the same power level.

We remark that in some cases where utilities only have data of lower temporal resolution, e.g., hourly measurements, our EV detection approach can still be applied but reduced accuracy may be expected.

B. Mixed-integer convex quadratic program

Let $\mathrm{P}_{i,d} \in \mathbb{R}^T$ be customer i's power demand during day d, and $\mathrm{P}^{\mathrm{EV}}_{i,d}, \mathrm{P}^{\mathrm{BL}}_{i,d} \in \mathbb{R}^T$ be the demand for EV charging and baseload, respectively. The mixed-integer convex quadratic program (MICQP) to determine $\mathrm{P}^{\mathrm{EV}}_{i,d}$ is formulated as:

$$\min_{\mathrm{P}^{\mathrm{EV}}_{i,d},\mathbf{x},\mathbf{y},\mathbf{z},\boldsymbol{\delta}} \quad \mathrm{Var}(\mathrm{P}^{\mathrm{BL}}_{i,d}) = \frac{1}{T}\sum_{t=1}^{T}\left(\mathrm{P}^{\mathrm{BL}}_{i,d}(t) - \overline{\mathrm{P}^{\mathrm{BL}}_{i,d}}\right)^2 \tag{1}$$

$$\begin{aligned}
\text{subject to} \quad & \overline{\mathrm{P}^{\mathrm{BL}}_{i,d}} = \frac{1}{T}\textstyle\sum_{t=1}^{T}\mathrm{P}^{\mathrm{BL}}_{i,d}(t), && (2)\\
& \mathrm{P}_{i,d} = \mathrm{P}^{\mathrm{EV}}_{i,d} + \mathrm{P}^{\mathrm{BL}}_{i,d}, && (3)\\
& \mathbf{x},\mathbf{y},\mathbf{z} \in \{0,1\}^T, && (4)\\
& \boldsymbol{\delta} \in \{0,1\}^{|\mathcal{P}|\times T}, && (5)\\
& \boldsymbol{\delta}^h \in \mathbb{R}^{|\mathcal{P}|}, && (6)\\
& \mathbf{1}_T^\top \mathbf{y} \le 2, && (7)\\
& \textstyle\sum_{n=0}^{\min\{T-t,3\}} \mathbf{y}(t+n) + \mathbf{z}(t+n) \le 1, \forall t, && (8)\\
& \boldsymbol{\delta}^\top \mathbf{1}_{|\mathcal{P}|} \le \mathbf{x}, && (9)\\
& \boldsymbol{\delta}^\top(\mathcal{P} - \varepsilon\mathbf{1}_{|\mathcal{P}|}) \le \mathrm{P}^{\mathrm{EV}}_{i,d} \le \boldsymbol{\delta}^\top(\mathcal{P} + \varepsilon\mathbf{1}_{|\mathcal{P}|}), && (10)\\
& \boldsymbol{\delta}^h = \frac{1}{T}\boldsymbol{\delta}\mathbf{1}_T, && (11)\\
& \boldsymbol{\delta}(p,t) \le 1 - \textstyle\sum_{n=1}^{|\mathcal{P}|-p}\boldsymbol{\delta}^h(p+n),\ \forall p,t, && (12)\\
& \mathrm{P}^{\mathrm{EV}}_{i,d}(t) - \mathrm{P}^{\mathrm{EV}}_{i,d}(t-1) \le \mathbf{y}(t)M,\ \forall t, && (13)\\
& \mathrm{P}^{\mathrm{EV}}_{i,d}(t-1) - \mathrm{P}^{\mathrm{EV}}_{i,d}(t) \le \mathbf{z}(t)M,\ \forall t, && (14)\\
& \mathrm{P}^{\mathrm{EV}}_{i,d}(t) \le \mathbf{x}(t)M,\ \forall t, && (15)\\
& \mathrm{P}^{\mathrm{EV}}_{i,d}(t), \mathrm{P}^{\mathrm{BL}}_{i,d}(t) \ge 0,\ \forall t, && (16)\\
& \mathbf{x}(t) - \mathbf{x}(t-1) = \mathbf{y}(t) - \mathbf{z}(t),\ \forall t. && (17)
\end{aligned}$$

The objective (1) is to minimize the variance of the baseload $\mathrm{P}_{i,d}^{\mathrm{BL}}$ after $\mathrm{P}_{i,d}^{\mathrm{EV}}$ is subtracted from $\mathrm{P}_{i,d}$, where $\overline{\mathrm{P}_{i,d}^{\mathrm{BL}}}$ is the mean value of the resulting baseload as computed in (2). The binary variables $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \{0,1\}^T$ indicate, respectively, whether an EV is being charged ($\mathbf{x}(t)=1$), the start time ($\mathbf{y}(t)=1$), and the end time ($\mathbf{z}(t)=1$) of a charging event for each time step t during the horizon T. Here, as we assume that EVs are not charged more than twice during a day, we limit the detection of charging events to 2 by (7), i.e., $\mathbf{y}(t)=1$ at no more than 2 time steps. As we do not consider charging events that last less than 1 hour, we use (8) to enforce that we cannot have an active $\mathbf{y}$ and an active $\mathbf{z}$ during any time window of 4 time steps. With (9) and (10), the power level at each time step t takes a single value from $\mathcal{P}$ (within some tolerance $\varepsilon > 0$), and the activated power level is indicated by the binary variable $\boldsymbol{\delta} \in \{0,1\}^{|\mathcal{P}| \times T}$. Here, $|\mathcal{P}|$ is the cardinality of $\mathcal{P}$, and $\mathbf{1}_{|\mathcal{P}|}$ is the column vector of ones of dimension $|\mathcal{P}|$. It is possible that different power levels may occur during a charging event, i.e., step changes, hence we need to ensure that the same power level is adopted during all charging events of the day. An intermediate variable $\boldsymbol{\delta}^h \in \mathbb{R}^{|\mathcal{P}|}$ is defined in (11) to calculate the fraction of power levels activated during the day, where $\mathbf{1}_T$ is the column vector of ones of dimension T. By (12) we force that the higher power level with a corresponding non-zero $\boldsymbol{\delta}^h$ value is always used during this day. By (13)-(17), the profile of a charging event must follow a rectangular waveform at the selected power level, where M is a large constant, i.e., $M \gg \max(\mathcal{P})$. Finally, the resulting $\mathrm{P}_{i,d}^{\mathrm{EV}}$ and $\mathrm{P}_{i,d}^{\mathrm{BL}}(t)$ should be non-negative at all times.

Note that for the simplicity in notation, the subscripts i and d on all binary and intermediate variables, namely $\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\delta}$, and $\boldsymbol{\delta}^h$, are omitted in the formulation of the MICQP.

III. Customers' EV charging patterns

Let T^{s} and T^{d} be random variables on sample space $\mathcal{T}$ and P^{c} be a random variable on sample space $\mathcal{P}$, which represent, respectively, the start time, the duration, and the charger power level of EV charging events for all customers on a given network. Here, $\mathcal{T} = [0, \Delta t, 2\Delta t, \cdots, (T-1)\Delta t]$ is the discretized time horizon with a time step $\Delta t = 15$ minutes. The charging patterns can be characterized through the inferred probability mass functions (PMFs) of these random variables, namely, $\hat{f}_{T^{\mathrm{s}}}(t)$, $\hat{f}_{T^{\mathrm{d}}}(t)$, and $\hat{f}_{P^{\mathrm{c}}}(p)$.

Let $t_{i,d}^{\mathrm{d}}, t_{i,d}^{\mathrm{s}}, t_{i,d}^{\mathrm{e}}$, and $P_{i,d}^{\mathrm{c}}$ be, respectively, the duration, the start time, the end time, and the power level of the EV charging events detected for customer i on day d by solving the MICQP proposed in the previous section. From the solution obtained, we compute $t_{i,d}^{\mathrm{s}} = \{t \in \mathcal{T} \mid \mathbf{y}(t) = 1\}$, $t_{i,s}^{\mathrm{e}} = \{t \in \mathcal{T} \mid \mathbf{z}(t) = 1\}$, $t_{i,d}^{\mathrm{d}} = t_{i,s}^{\mathrm{e}} - t_{i,s}^{\mathrm{s}}$, and $P_{i,d}^{\mathrm{c}} = \{\max_t(\mathrm{P}_{i,d}^{\mathrm{EV}})\}$. We note that as we detect up to two charging events in each day, two values may be collected in each of them. By repeatedly solving the MICQP for all customers and for all days, i.e., to extract all charging events in $\mathbf{P}$, we obtain the following multisets of detected results:

$$
\begin{aligned}
\mathcal{T}^{\mathrm{s}} &= \{t_{i,d}^{\mathrm{s}},\ d = 1,2,\cdots,N_{\mathrm{D}},\ i = 1,2,\cdots,N\} \\
\mathcal{T}^{\mathrm{d}} &= \{t_{i,d}^{\mathrm{d}},\ d = 1,2,\cdots,N_{\mathrm{D}},\ i = 1,2,\cdots,N\} \\
\mathcal{P}^{\mathrm{c}} &= \{P_{i,d}^{\mathrm{c}},\ d = 1,2,\cdots,N_{\mathrm{D}},\ i = 1,2,\cdots,N\},
\end{aligned}
$$

We remark that $\mathcal{T}^{\mathrm{s}}, \mathcal{T}^{\mathrm{d}}$, and $\mathcal{P}^{\mathrm{c}}$ contain the same number of elements. The probability mass functions (PMFs) for the random variables $T^{\mathrm{s}}, T^{\mathrm{d}}$, and P^{c} are then approximated by the empirical distributions:

$$
\begin{aligned}
\hat{f}_{T^{\mathrm{s}}}(t) &= \frac{1}{|\mathcal{T}^{\mathrm{s}}|} \sum_{t^{\mathrm{s}} \in \mathcal{T}^{\mathrm{s}}} \mathbb{1}_{\{t^{\mathrm{s}}\}}(t),\ t \in \mathcal{T} \\
\hat{f}_{T^{\mathrm{d}}}(t) &= \frac{1}{|\mathcal{T}^{\mathrm{d}}|} \sum_{t^{\mathrm{d}} \in \mathcal{T}^{\mathrm{d}}} \mathbb{1}_{\{t^{\mathrm{d}}\}}(t),\ t \in \mathcal{T} \\
\hat{f}_{P^{\mathrm{c}}}(p) &= \frac{1}{|\mathcal{P}^{\mathrm{c}}|} \sum_{p^{\mathrm{c}} \in \mathcal{P}^{\mathrm{c}}} \mathbb{1}_{\{p^{\mathrm{c}}\}}(p),\ p \in \mathcal{P},
\end{aligned}
\tag{18}
$$

where $\mathbb{1}$ is an indicator function, i.e., $\mathbb{1}_{\{p^{\mathrm{c}}\}}(p) = 1$ if $p^{\mathrm{c}} = p$ and $\mathbb{1}_{\{p^{\mathrm{c}}\}}(p) = 0$ otherwise.

Based on $\mathcal{T}^{\mathrm{s}}$ and $\mathcal{T}^{\mathrm{d}}$, we then construct a set of all possible charging profiles in per-unit (p.u.) values:

$$
\mathcal{L} = \{l_{t^{\mathrm{s}},t^{\mathrm{d}}}(t) \mid t^{\mathrm{s}} \in \underline{\mathcal{T}}^{\mathrm{s}}, t^{\mathrm{d}} \in \underline{\mathcal{T}}^{\mathrm{d}}\}, \tag{19}
$$

where $\underline{\mathcal{T}}^{\mathrm{s}}$ and $\underline{\mathcal{T}}^{\mathrm{d}}$ are the support of $\mathcal{T}^{\mathrm{s}}$ and $\mathcal{T}^{\mathrm{d}}$, respectively, and

$$
l_{t^{\mathrm{s}},t^{\mathrm{d}}}(t) = \begin{cases} 1, & \text{if } t^{\mathrm{s}} \le t < t^{\mathrm{s}} + t^{\mathrm{d}}, \\ 0, & \text{otherwise.} \end{cases} \tag{20}
$$

A profile $l_{t^{\mathrm{s}},t^{\mathrm{d}}}(t)$ reconstructs the per-unit rectangular waveform of a charging event that has been detected from customers' smart meter data. To conduct an impact analysis of EV charging, we also need to calculate the probability that each $l_{t^{\mathrm{s}},t^{\mathrm{d}}}(t)$ is adopted, which is given by:

$$
\begin{aligned}
\Pr[l_{t^{\mathrm{s}},t^{\mathrm{d}}}] &= \Pr[T^{\mathrm{s}} = t^{\mathrm{s}}, T^{\mathrm{d}} = t^{\mathrm{d}}] \\
&= \Pr[T^{\mathrm{s}} = t^{\mathrm{s}} \mid T^{\mathrm{d}} = t^{\mathrm{d}}] \hat{f}_{T^{\mathrm{d}}}(t^{\mathrm{d}}),
\end{aligned}
\tag{21}
$$

where $\Pr[T^{\mathrm{s}} = t^{\mathrm{s}} \mid T^{\mathrm{d}} = t^{\mathrm{d}}]$ is approximated by m/n, where n is the total number of detected charging events lasting for t^{d} and m is the number of charging events starting at t^{s} out of the n events. We remark that as $t^{\mathrm{d}} \in \underline{\mathcal{T}}^{\mathrm{d}}$, we have $n \ge 1$.

IV. Case study

We use the Pecan Street data [18] to evaluate the performance of our method to extract customers' EV charging habits. The dataset contains smart meter data measured for 76 customers with EVs for the full year of 2021. Daily power demand of EV charging is also recorded for each customer, which serves as the "ground truth" to verify the accuracy of the detection results using our approach.

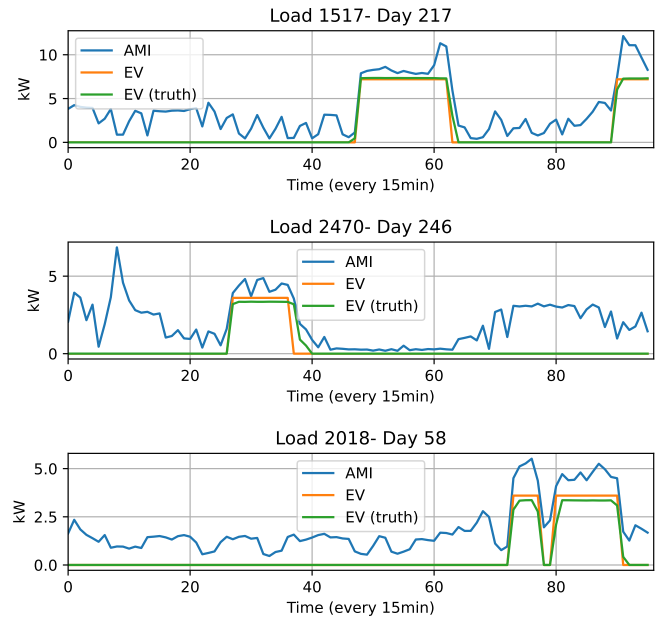

Fig. 1: Examples of EV charging events detected from smart meter data measured throughout a day at every 15 minutes

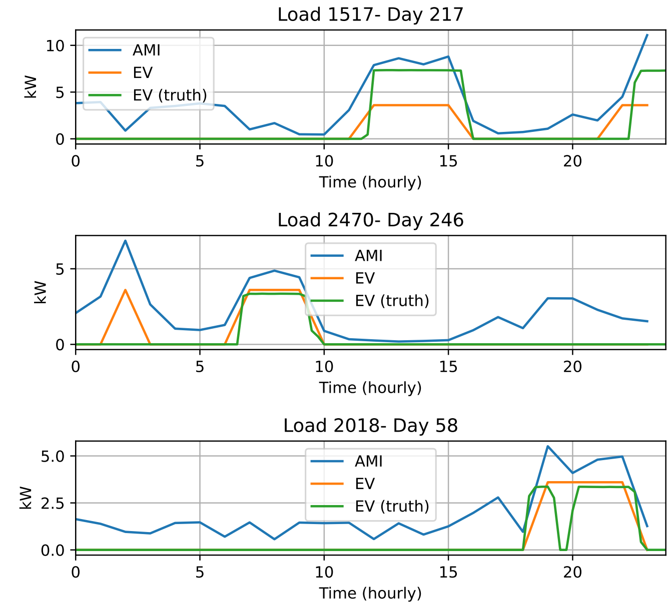

Fig. 2: EV charging events detected for the same customer/day as in Figure 1 but from smart meter data measured at every hour

A. *Distributions of charging patterns*

The first step is to detect daily charging events of each customer from the smart meter data by solving the MICQP. We remark that some customers have photovoltaic (PV) systems, so the measured consumption has already been offset by the PV generation. In order to compare the detected events with the ground truth, the smart meter data for these customers are pre-processed by subtracting the PV generation. Figures 1 and 2 show examples of the charging events detected during a day of selected customers. Two experiments are performed: (1) using smart meter data measured at a 15-minute interval and (2) using hourly measurements.

We observe in Figure 1 that EV charging events are accurately detected from the meter data measured at every 15 minutes in terms of start time, duration, and the power level. We demonstrate that our method can accurately detect the EV charging events in the following cases: multiple events occur during the day (e.g., Load 1517 and Load 2018), periods with constant higher power consumption are not identified as EV charging (e.g., Load 2470), and the pause between two charging events is correctly recognized (e.g., Load 2018). We also test our approach on the same customers/days but using hourly meter measurements, and the results are shown in Figure 2. We observe that while EV charging events can still be detected, the level of accuracy is reduced mainly due to the lack of data. For example, the power level is determined to be 3.6kW for Load 1517 because the baseload consumption during the day would become negative if the power level were set at 7.2kW. For Load 2470, a peak demand at 2AM is mistakenly identified as a charging event for 1 hour. Lastly, the brief pause between the two charging events for Load 2018 is not picked up, hence the extracted information on start time and duration of the charging event is not exact.

To benchmark the performance of our proposed approach to detect the EV charging events, we compare it with two other algorithms: one based on signal decomposition [11] and the other using a trained CNN model [5]. We adopt the two metrics used in [11] to evaluate the detection performance on customer i's data:

- *F1 Score* F1 that measures the accuracy of detection results. The score is defined in terms of sample counts in True Positive (TP), False Positive (FP), and False Negative (FN) conditions:
$$\mathrm{F1}_i = \frac{2TP_i}{2TP_i + FP_i + FN_i}.$$
- *Explained Variance Score* $\mathrm{E}_{\mathrm{var}}$ that measures the dispersion or discrepancy between the detected and true EV consumption data. The $\mathrm{E}_{\mathrm{var}}$ score is defined as:
$$\mathrm{E}_{\mathrm{var},i} = 1 - \frac{\mathrm{Var}(\mathrm{P}_i^{\mathrm{EV}} - \mathrm{P}_{i,\mathrm{true}}^{\mathrm{EV}})}{\mathrm{Var}(\mathrm{P}_{i,\mathrm{true}}^{\mathrm{EV}})},$$
where $\mathrm{P}_i^{\mathrm{EV}}, \mathrm{P}_{i,\mathrm{true}}^{\mathrm{EV}} \in \mathbb{R}^{N_{\mathrm{D}}T}$ are the detected and true EV charging consumption for all days, respectively.

We remark that the closer the two metric values are to 1, the better the performance of the algorithm is.

Although the benchmark algorithms also use Pecan Street datasets to test the performance, it is not clear what customers and time periods are included in the datasets. Therefore, it is not possible to directly compare the metric values of a specific customer. Instead, we report the minimum, mean, median, and maximum of F1 and $\mathrm{E}_{\mathrm{var}}$ values obtained for all customers in

	Proposed approach		Decomposition [11]		CNN [5]	
	F1	E_{var}	F1	E_{var}	F1	E_{var}
Min	0.696	0.512	0.771	0.611	0.765	0.857
Mean	0.876	0.805	0.906	0.831	0.892	0.890
Median	0.889	0.819	0.909	0.852	0.911	0.891
Max	0.967	0.972	0.982	0.974	0.925	0.924

TABLE I: Performance comparison with other algorithms

the dataset used in this work, and compare them with those reported in [5], [11] using their datasets. The values are shown in Table I. We note that the performance of our approach is comparable to that of the benchmark algorithms. Although the minimum, mean, and median values from our approach are slightly lower, we should emphasize that the meter data used in both benchmark algorithms have 1-minute intervals, while data with 15-minute intervals are used in our work. As pointed out in [11], as the data measurement interval increases, the accuracy of the algorithm decreases (e.g., the mean E_{var} drops to 0.69 if data with 15-minute intervals are used in [11]). Further, to obtain high accuracy, the CNN approach [5], being a supervised-learning method, needs a large number of labelled training samples which are difficult to obtain in practice. While the method in [11] does not require training, tuning of the hyperparamters in both stages of the decomposition is required to differentiate EV charging from the use of air conditioners. In our approach, by leveraging the readily available contextual information of EV charging events, we do not require high-resolution data, and we achieve comparable accuracy with that of the literature without any model training or hyperparameter tuning.

Next, we compute the empirical distributions for customers' charging habits as in (18) based on charging events detected for all customers during a full year. We use the meter data with 15-minute intervals due to the better accuracy of the detection results. We compare the estimated PMFs and the cumulative distribution functions (CDFs, except for the charger power levels) computed from the detection results with those from the ground truth in Figure 3. The following can be inferred from the results for the customers in this Pecan Street dataset:

- Customers tend to charge EVs during evenings until early mornings, with slightly higher probabilities starting in the evenings;
- Each charging event usually does not last more than 9 hours;
- Daily energy consumption for charging EVs rarely exceed 60kWh, and;
- More than half of the EVs are charged at 3.6kW.

We remark that in the distributions of the charging start time, high probabilities of starting the charging at midnight are observed. This is because some charging events start during the evening and continue until the next day. In this case, as we only detect events during one day's time window, the event is broken into two parts with the second part starting at midnight the next day. Also, as we only detect events lasting more

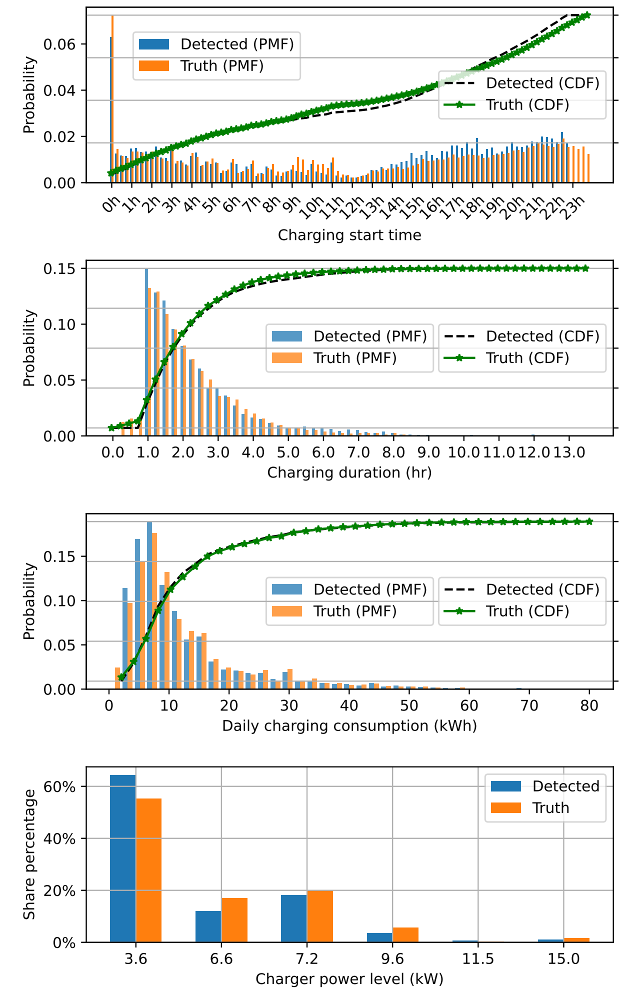

Fig. 3: Distributions of start time, duration, daily energy, and charger power levels based on detected charging events

than 1 hour, we obtain a zero probability for charging events starting on and after 11PM and for charging duration less than 1 hour, which are observable differences when comparing the distributions in Figure 3.

B. Impact analysis of EV charging

In this part of the case study, we use the inferred charging patterns to study the impact of EV charging to the distribution network using the stochastic method described in [1]. The IEEE-8500 test feeder [22] is selected with some modifications to demonstrate the results. We construct the set of all possible per-unit charging profiles based on the detected results using (19) and (20). The resulting set contains a total of 4224 profiles. We then use (21) to calculate the probability that each profile will be adopted by the customers on the network. We illustrate the probabilities of all profiles as a heat map in Figure 4, where each profile is represented by the tuple $(start\ time, duration)$. For example, (8PM, 2h30m) represents the profile for which charging starts at 8PM and lasts for 2 hours and 30 minutes. To get the EV charging

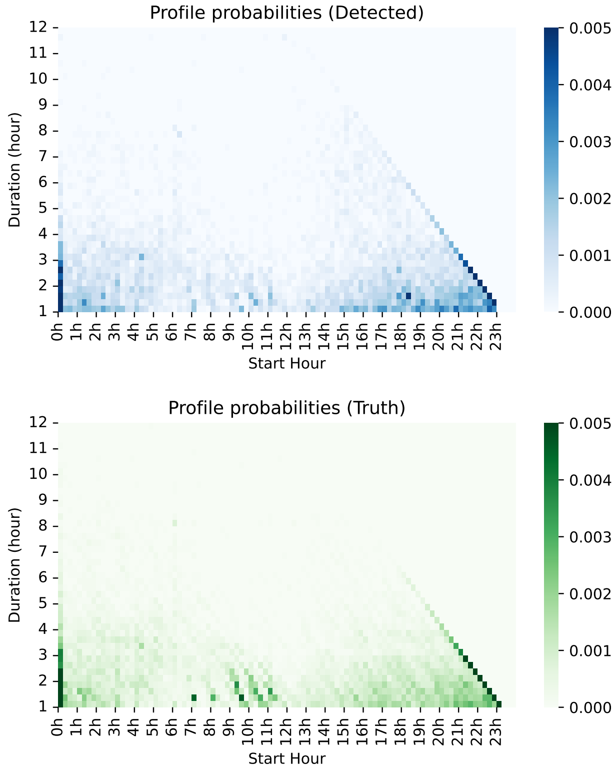

Fig. 4: Probabilities of profiles

profiles in kW, we multiply the per-unit profiles with the 6 power levels in $\mathcal{P}$. The distribution of the power levels is shown in the bottom right plot of Figure 3. For simplicity, we assume that the distribution of power levels and the probabilities that per-unit charging profiles are adopted by the customers are independent.

In Figure 5 we show the impacts of EV charging on the loading level of the substation transformer at penetration rates of 10%, 30%, 50%, and 80%. The penetration rate is defined as the ratio of the number of EVs connected to the network over the total number of customers. We use both the inferred probabilities and the true probabilities as in Figure 4 to perform the impact analyses. At each penetration rate, we show the loading curve over one day without any EV connected and compare the loading curves obtained from the two sets of probabilities. We observe that as the penetration grows, impacts to the transformer loading level increase, and the transformer becomes overloaded during the evening hours at 80% EV penetration. Due to differences in the profile distributions, errors between the two loading curves are more observable at higher penetration rates. We extract the per-unit errors at all penetrations and show them in Figure 6. Errors are mostly higher during the 9h-11h and evening hours, which are the periods where differences in the profile distributions are more observable in Figure 4. At 80% penetration rate, the worst error is around 0.06 p.u. or 6%.

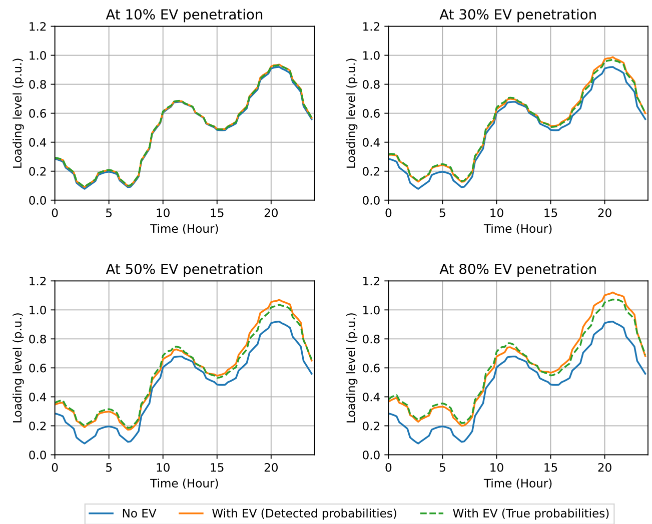

Fig. 5: Comparison of the substation transformer loading levels using detected and true profile probabilities at various EV penetration rates

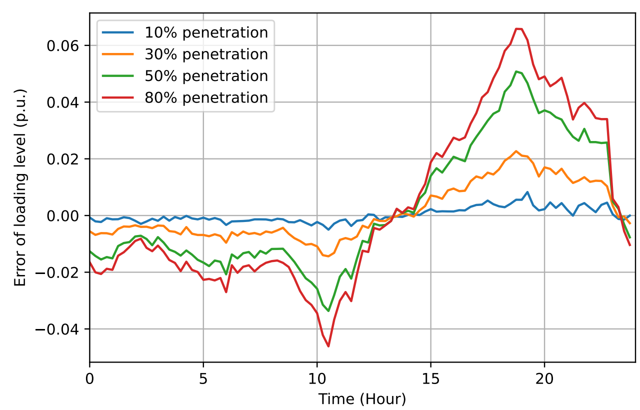

Fig. 6: Result errors of the substation transformer loading levels using detected and true profile probabilities at various EV penetration rates

V. Conclusion

In this work, a non-intrusive and training-free method is proposed to detect BTM EV charging events based on customers' smart meter data. Our approach does not require labelled training data nor hyperparameter tuning, and achieves a similar level of accuracy in extracting information of charging events as that of the literature by using meter data measured at every 15 minutes. Through a data-driven approach, we infer customers' charging patterns in terms of probability distributions of charging profiles from the detection results during an entire year. We compare the inferred probability distributions with those from the ground truth, and illustrate that even if there exist some minor differences between the two sets of distributions, no significant error occurs in the results of EV charging impact analyses. The inferred probability distributions from our approach allow utilities to not only evaluate impacts that

EV charging may bring to power distribution networks, but also design incentive programs to mitigate equipment overload for better planning and operation of their networks [23]. In practice, as EV owners' charging behaviours may shift over time, the variations can be captured by periodically applying the method on customers' smart meter data to update the inferred probability distributions.

Our detection approach requires that the meter data contain only consumption; in other words, if customers have power-generating devices installed such as PV or battery systems, the generation must be subtracted from the meter data. As this information may not be available in practice, in the future we wish to extend our approach by adding an extra step to estimate BTM generation (e.g., [24]) and exclude them from the meter data. Our approach can be further improved by relaxing the assumption of rectangular charging waveform, i.e., by considering variations of charging power in time.

ACKNOWLEDGMENT

The authors would like to thank Eaton's CYME International T&D and The Natural Sciences and Engineering Research Council of Canada (NSERC) for their support of this work.

REFERENCES

[1] F. Li, I. Kocar, and A. Lesage-Landry, "A rapid method for impact analysis of grid-edge technologies on power distribution networks," *IEEE Transactions on Power Systems*, pp. 1–12, 2023.

[2] S. Wang, L. Du, J. Ye, and D. Zhao, "Robust identification of EV charging profiles," in *2018 IEEE Transportation Electrification Conference and Expo (ITEC)*, pp. 1–6, 2018.

[3] B. I. Fesche, V. Hoffmann, K. Ingebrigtsen, I. N. Christie, and M. Punnerud, "Signal discovery in the smart grid-finding: Electric vehicle charging patterns in power consumption data," *CIRED 2019 Conference*, no. 1531, 2019.

[4] V. Hoffmann, B. I. Fesche, K. Ingebrigtsen, I. N. Christie, and M. Punnerud, "Automated detection of electric vehicles in hourly smart meter data," *CIRED 2019 Conference*, no. 1531, 2019.

[5] D. Yang, X. Gao, L. Kong, Y. Pang, and B. Zhou, "An event-driven convolutional neural architecture for non-intrusive load monitoring of residential appliance," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 2, pp. 173–182, 2020.

[6] A. Verma, A. Asadi, K. Yang, and S. Tyagi, "A data-driven approach to identify households with plug-in electrical vehicles (PEVs)," *Applied Energy*, vol. 160, pp. 71–79, 2015.

[7] A. Verma, A. Asadi, K. Yang, A. Maitra, and H. Asgeirsson, "Analyzing household charging patterns of plug-in electric vehicles (PEVs): A data mining approach," *Computers & Industrial Engineering*, vol. 128, pp. 964–973, 2019.

[8] P. Zhang, C. Zhou, B. G. Stewart, D. M. Hepburn, W. Zhou, and J. Yu, "An improved non-intrusive load monitoring method for recognition of electric vehicle battery charging load," *Energy Procedia*, vol. 12, pp. 104–112, 2011. The Proceedings of International Conference on Smart Grid and Clean Energy Technologies (ICSGCE 2011).

[9] A. Shaw and B. P. Nayak, "Electric vehicle charging load filtering by power signature analysis," in *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*, pp. 71–75, 2017.

[10] Z. Zhang, J. H. Son, Y. Li, M. Trayer, Z. Pi, D. Y. Hwang, and J. K. Moon, "Training-free non-intrusive load monitoring of electric vehicle charging with low sampling rate," in *IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society*, pp. 5419–5425, 2014.

[11] Y. Xiang, Y. Wang, S. Xia, and F. Teng, "Charging load pattern extraction for residential electric vehicles: A training-free nonintrusive method," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 7028–7039, 2021.

[12] A. A. Munshi and Y. A.-R. I. Mohamed, "Unsupervised nonintrusive extraction of electrical vehicle charging load patterns," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 266–279, 2019.

[13] A. A. Munshi and Y. A.-R. I. Mohamed, "Extracting and defining flexibility of residential electrical vehicle charging loads," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 448–461, 2018.

[14] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 26, pp. 356–362, Sep. 2021.

[15] S. Wang, L. Du, J. Ye, and D. Zhao, "A deep generative model for non-intrusive identification of EV charging profiles," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 4916–4927, 2020.

[16] G. Tal, D. Chakraborty, A. Jenn, J. H. Lee, and D. Bunch, "Factors affecting demand for plug-in charging infrastructure: An analysis of plug-in electric vehicle commuters," *UC Office of the President: University of California Institute of Transportation Studies*, 2020.

[17] S. Habib, M. M. Khan, F. Abbas, and H. Tang, "Assessment of electric vehicles concerning impacts, charging infrastructure with unidirectional and bidirectional chargers, and power flow comparisons," *International Journal of Energy Research*, vol. 42, no. 11, pp. 3416–3441, 2018.

[18] Pecan Street Dataport. http://dataport.pecanstreet.org, Last accessed on 2022-04-25.

[19] K. S. Cetin and A. Novoselac, "Single and multi-family residential central all-air HVAC system operational characteristics in cooling-dominated climate," *Energy and Buildings*, vol. 96, pp. 210–220, 2015.

[20] E. Proedrou, "A comprehensive review of residential electricity load profile models," *IEEE Access*, vol. 9, pp. 12114–12133, 2021.

[21] L. W. Davis, "Electric vehicles in multi-vehicle households," *Applied Economics Letters*, vol. 30, no. 14, pp. 1909–1912, 2023.

[22] R. F. Arritt and R. C. Dugan, "The IEEE 8500-node test feeder," in *IEEE PES T&D 2010*, pp. 1–6, 2010.

[23] F. Li, I. Kocar, and A. Lesage-Landry, "Mitigating equipment overloads due to electric vehicle charging using customer incentives," in *2023 IEEE Power & Energy Society General Meeting (PESGM)*, (Orlando, USA), 2023.

[24] S. Lin and H. Zhu, "Enhancing the spatio-temporal observability of grid-edge resources in distribution grids," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5434–5443, 2021.