How Many More Charging Stations Do We Need? A Data-Driven Approach Considering Charging Station Overflow Dynamics.
Weekx, Simon; Tal, Gil; Vanhaverbeke, Lieselot

Published in:
Transportation Research Record: Journal of the Transportation Research Board

Publication date:
2024

Document Version:
Accepted author manuscript

Link to publication

Citation for published version (APA):
Weekx, S., Tal, G., & Vanhaverbeke, L. (Accepted/In press). How Many More Charging Stations Do We Need? A Data-Driven Approach Considering Charging Station Overflow Dynamics. Transportation Research Record: Journal of the Transportation Research Board.

Copyright
No part of this publication may be reproduced or transmitted in any form, without the prior written permission of the author(s) or other rights holders to whom publication rights have been transferred, unless permitted by a license attached to the publication (a Creative Commons license or other), or unless exceptions to copyright law apply.

Take down policy
If you believe that this document infringes your copyright or other rights, please contact openaccess@vub.be, with details of the nature of the infringement. We will investigate the claim and if justified, we will take the appropriate steps.

Download date: 10. May. 2024
How Many More Public Charging Stations Do We Need? A Data-Driven Approach Considering Charging Station Overflow Dynamics.

Simon Weekx
MOBI Electromobility Research Centre
Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
Email: simon.weekx@vub.be
https://orcid.org/0000-0001-8876-0837

Gil Tal
University of California, Davis, Institute of Transportation Studies
1590 Tilia St, Davis, California, United States, 95616
Email: gtal@ucdavis.edu
https://orcid.org/0000-0001-7843-3664

Lieselot Vanhaverbeke
MOBI Electromobility Research Centre
Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
Email: lieselot.vanhaverbeke@vub.be
https://orcid.org/0000-0001-6107-0254

Keywords: Electric vehicles, Charging behavior, Infrastructure planning, Simulation, Charging infrastructure
ABSTRACT

The development of public charging infrastructure is crucial to support mass EV adoption. Although many cities worldwide already installed an initial network of public chargers, it is often unclear whether the current supply of infrastructure is in line with demand, and how many more charging stations are required to cope with future EV growth. In this sense, transactional charging data on the existing network can help answer these questions. We present a novel method that uses historical charging data as input to determine (1) how many more chargers are required to meet future demand, and (2) where to install these new chargers. By mining the individual charging behavior of EV drivers, we show that overflow dynamics can be found between charging stations. That is, when a preferred charging station is fully occupied, EV drivers are found to divert to other nearby charging stations. Identifying these dynamics allows us to simulate the impact of a demand increase on the charging infrastructure network more accurately. We found the required number of new chargers to be significantly lower when considering overflow dynamics. Our simulations indicate that for a doubling of the demand, 30-50% fewer charging points are needed compared to when overflow dynamics are neglected while maintaining the same failure rate (i.e., percentage of failed charging sessions in the network). Determining the exact number of chargers will depend on the failure rate policymakers are willing to accept, reflecting the trade-off between charging convenience and utilization.
1 INTRODUCTION

The electrification of the transport sector will require large investments in charging infrastructure worldwide. Besides having chargers available at home and the workplace, the public domain constitutes an important part of the total charger supply [1, 2]. This is especially the case in large cities and urbanized areas, where many residents and visitors don’t have access to off-street parking and thus rely on public infrastructure for their charging needs. Although many cities already have an initial network of public chargers up and running, many more additional chargers are required to keep supporting mass EV adoption [3]. Considering the scarcity of available space in densely populated areas and the cost of installing new chargers, key questions that urban planners and Charging Point Operators (CPOs) should address are how many more chargers are required to meet all future charging demand, and where to install these new chargers. Incorrect decisions on either of these dimensions may result in under-, or overutilized charging stations, causing a waste of public resources (in terms of space and money), and possibly hampering future EV adoption.

To this end, policymakers can use the observed charging behavior from the initial charging network to make more informed decisions on how to upgrade it. Despite the high value of charging data as a source of information, research that shows how this data can be translated to policy decisions is still lacking. To bridge this gap in the literature, we present a novel model that can extract from existing charging data how many more chargers are required for different demand growth scenarios, and what the optimal locations of these new chargers are. The model itself consists of three parts: (1) grouping charging stations that are used interchangeably by groups of EV drivers into zones; (2) identifying the overflow dynamics between charging stations of the same zone (i.e., which chargers are used as an alternative to each other); and (3) simulating the effect of an increase in charging demand, considering the overflow.

2 LITERATURE REVIEW

Considerable research has been focused on the problem of how to design a robust network of charging stations for EVs. Already before EVs came to the market, initial research has used proxy variables to estimate which locations would be most suitable for locating EV chargers. Cai, et al. [4] provide an overview, indicating that traffic data, gas station locations, and vehicle ownership rates are often used as proxies. With the number of EV adopters starting to rise, another line of research has focused on characterizing early adopters in terms of their socio-demographic profile. Studies based on EV adoption data from the U.K. [5], Ireland [6], and Pennsylvania [7] indicate that early adopters are more likely to be found in sub-urban regions with a high income, higher median age, more house owners, smaller household sizes, and access to more than one car. These attributes have been used in case studies conducted in Milan [8] and the Tyne and Wear County (U.K.) [9] to locate charging stations based on where early adopters are most likely to live.

Seeking more data-driven approaches, others have modeled charging infrastructure from large-scale trajectory data, household travel surveys, or surveying the EV population. Cai, et al. [4] mine vehicle trajectory data stemming from an Internal Combustion Engine (ICE) taxi fleet in Beijing to study charging station planning, environmental-, and grid impacts. Mandev, et al. [10] collect fuel consumption data on Plug-in Hybrid EVs (PHEVs) from the U.S. and Canada to empirically study their charging behavior. They find that PHEV owners charge about once per day, mostly during the night. Li and Jenn [11] use individual activity-based travel diary data from California to determine optimal charging locations, the optimal charging strategy, and the required number of chargers. Lee, et al. [12] survey the EV population in California and find that the use of public, workplace, and home chargers is interconnected. They find that although the majority of EV drivers solely rely on charging at home, about 30-40% of the EV population display a mixed use of infrastructure depending on socio-demographic and EV characteristics.

While these findings provide useful inputs for urban planners to develop an initial network of chargers, the question remains how this network should be upgraded over time to keep up with future EV charging demand. In this sense, observed data on the usage of the existing charging network allows us to
understand the complex charging behavior of EV drivers [13], and hence constitutes an important source of information. However, fewer studies are found that show how charging data can translate to policy decisions on how many more chargers are needed and where they should be located. Wagner, et al. [14] and Pevec, et al. [15] use a two-step approach where they first predict the utilization of potential new chargers based on historical charging data from Amsterdam and the Netherlands respectively. In a second step, they use this prediction as input in an optimization model to find the optimal locations for new chargers. However, the authors do not consider to what extent the existing charging supply is sufficiently covering the charging demand, and how many more charging stations are required to cope with different EV growth scenarios. Helmus, et al. [16], Wolbertus, et al. [17], Hoekstra and Hogeveen [18] propose Agent-Based Simulation Models (ABMs) based on charging data from the Netherlands to answer those questions. However, they also have some limitations. First, ABMs are computationally expensive, and their models are trained on multi-year charging datasets, requiring a high level of data maturity. Second, while Helmus, et al. [16] model the charging behavior through the concept of EV driver interaction, they do so indirectly, through external decision rules.

In this study, we propose a novel method that uses transactional charging data to simulate how many more charging stations are needed for different future EV growth scenarios, as well as what the optimal locations of those charging stations are. The contributions of our study can be summarized as follows. First, the majority of existing studies use simulated data or variables related to EV ownership to model charging infrastructure. We contribute by demonstrating how observed charging data at the level of the individual EV driver can be used as input for infrastructure planning, allowing for more data-driven planning. Second, to be able to determine how many more charging stations are required, it is necessary to know whether the current supply of infrastructure is in line with demand. Through the concept failure rate (i.e., percentage of failed charging sessions in the network) we show how demand and supply can be matched and are able to formulate exactly how many more chargers are required. Third, our model introduces the novel concepts of “charging zones” (see Definition 1) and “overflow sessions” (see Definition 2) and demonstrates how they can be empirically retrieved from charging transactions. This enhances the deeper understanding of charging infrastructure data, providing new tools for planners to aid the decision on where to locate new charging stations.

3 DATA
The dataset includes one full year of charging transactions (also referred to as sessions) from Dec. 2021 to Nov. 2022, recorded at 197 on-street public level 2 charging stations in the Brussels Capital Region. Each charging station has a unique location and is equipped with two Charging Points (i.e., CPs, also referred to as chargers). Only stations that were available for the full period of observation (installed before Dec. 1st, 2021) are included. Each charging transaction is characterized by a unique ID of the charging station (indicating where the session happened) and EV driver (indicating who performed the session). The whole set of charging stations in Brussels is referred to as the charging network. Descriptive statistics on both are described below.

3.1 Charging stations
OpenStreetMap is used to calculate the walking time between each pair of stations, which is used later as input in the clustering model (section 4.1). Figure 1 shows a map of all charging stations in the dataset and Figure 2 shows the distribution of the walking time from each station to the nearest other station. The mean walking time to the closest neighbor is 5.03 minutes, and for 99% of all the charging stations, the walking time is less than 15 minutes.
3.2 EV Drivers

Each EV driver in the dataset is identified with a unique ID based on the charging card that was used for the transaction. This allows us to analyze charging behavior at the micro level of the individual EV driver. Table 1 below describes the EV driver based on where they charge (multiple or single locations) and when they charge (day-only, night-only, or day-and-night). Most of the EV drivers (57%) only charge at one single location. Most of this group is formed by daytime-only users, most likely visitors who infrequently use the public charging stations. Of the EV drivers that charge at multiple locations (43%), the majority are day-only or day-and-night users. These are most likely residents and visitors who use the charging infrastructure more frequently than single-location EV drivers.

<table>
<thead>
<tr>
<th>Where do EV drivers charge?</th>
<th>Day-and-night</th>
<th>Day-only</th>
<th>Night-only</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple</td>
<td>17%</td>
<td>26%</td>
<td>&lt;1%</td>
<td>43%</td>
</tr>
<tr>
<td>Single</td>
<td>5%</td>
<td>48%</td>
<td>4%</td>
<td>57%</td>
</tr>
<tr>
<td>Total</td>
<td>22%</td>
<td>74%</td>
<td>4%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1 EV Driver typology.

4 METHODS

The novelty of our model is that we combine both the geographical information on the charging stations (section 3.1) and the individual charging data on the EV driver (section 3.2) to identify overflow dynamics. The analysis includes three steps. First, charging stations are clustered together into charging zones. Second, the overflow dynamics within the zones are mined from the charging data. Third, the simulation model assesses the impact of an increase in charging demand. An overview of the nomenclature that is used can be found in Table 2.
4.1 Step 1: Clustering

The goal of the clustering is to group together charging stations that are used interchangeably by multiple EV drivers. The underlying idea is that when a preferred charging station is not available for an EV user to charge, it is likely that the user will charge their vehicle at the next available charging station, given that it is still within a reasonable walking distance of the driver’s destination. Identifying these groups, defined here as charging zones, allows us to model how one station within a zone will be impacted by the activity at the other stations in that zone.

Although clustering charging stations is already studied in the EV literature [15, 19-21], no research to this date has considered the underlying group of EV drivers that visit the station as a clustering criterion. Traditional clustering methods (e.g., K-means [21, 22], Gaussian Mixture Models [13, 23], DBSCAN [24]) are less suitable because of the combinatorial explosion of all possible “station-EV driver” combinations [25]. Hence, we utilize association rule mining to find groups of charging stations that are all visited interchangeably by multiple EV drivers.

**Definition 1** (Charging zone). Given a set of charging stations $S$, the accompanying set of charging transactions $T$, a maximum walking time $t_{\text{max}}$ and a minimum support level $e_{\text{min}}$, a charging zone $z \in Z$ is defined as the largest possible subset of stations $z \subseteq S$, for which the following two conditions hold:

1. $\text{time}(s_i, s_j) \leq t_{\text{max}}, \forall s_i, s_j \in z$: the walking time between any two stations in the zone is less than or equal to the maximum walking time.

2. $\text{supp}(z) \geq e_{\text{min}}$: the number of EV drivers that visited all stations in the zone is greater than or equal to the minimum support level.
Based on Definition 1, all charging zones can be found in two steps. First, all largest groups of stations are identified for which the walking time between any pair of them is less than $t_{\text{max}}$. Walking time is preferred as distance metric over driving time because the decision of the EV driver to charge at an alternative station (e.g., ‘B’) when a preferred station (e.g., ‘A’) is not available will most likely depend on the walking time from B to the EV driver’s final destination, rather than on the driving time from A to B. Within graph theory, finding these largest groups of nodes that are all connected is a well-known problem referred to as The Maximal Clique Problem, and algorithms exist to solve it [26]. These largest groups of stations are referred to as candidate zones. Second, for each candidate zone, the support (i.e., the number of EV drivers that have visited all stations in the zone) is calculated from the charging transactions. If the support level is greater than or equal to $ev_{\text{min}}$, a charging zone is found, and the algorithm continues to the next candidate zone. If not, the support level of each sub-zone (if any) is calculated until all sub-zones have been searched. This search procedure is known as the problem of finding all Maximal Frequent Itemset in Association Rule Mining, and algorithms to solve it can be found in [27].

### 4.2 Step 2: Identifying overflow dynamics.

Once charging zones have been identified, we can mine for the presence of overflow dynamics within the zones. Overflow dynamics show to which (if any) charging stations EV drivers divert when a preferred station is not available. For this, we employ the concept of an overflow session, which is defined in Definition 2.

**Definition 2 (Overflow session).** Given a set of charging stations $S$, a set of charging zones $Z$, and the accompanying set of charging transactions $T$, an overflow session $t_s \in T$ of station $s \in S$ is defined as a charging session that occurred at a different station $u \in Z_s$ ($u \neq s$), because station $s$ was fully occupied (i.e., both plugs were in use).

Definition 2 implies causality, meaning that an overflow session only happens at a certain station because another station was occupied. To identify overflow sessions, we statistically test for the presence of causality as is shown in the process in Figure 3. This is an iterative process that is repeated for each charging station $s \in S$ in the dataset. First, all zones that contain the station $s$ are identified ($= Z_s$). Next, we loop over each EV driver that has charged at least once in any of these zones ($= EV_s$). For each driver, we find its most used station (i.e., favorite station) within the zones and then look at all instances where the driver does not charge at the favorite station ($= T_{s,Z_s}$), but at any other station in any of the zones $Z_s$. Next, we calculate the proportion of these sessions, for which 2 minutes before the start time of the sessions, the favorite station $s$ was unavailable ($= \hat{p}_{s,Z_s}$). The 2-minute time difference represents the moment where the EV driver finds out its preferred station is unavailable and decides to drive to the other nearby station. The higher this proportion, the more likely that the EV driver is charging elsewhere because the favorite station is unavailable.

However, this is not yet evidence for causality because it is possible that the EV driver wanted to charge elsewhere, and that the favorite station was fully occupied at that time by random chance. Therefore, we also calculate the proportion $p_{s,Z_s}$ one would expect due to random chance (based on the utilization rate of $s$) and compare it with the actual observed proportion $\hat{p}_{s,Z_s}$. This makes it possible to verify the hypothesis that $\hat{p}_{s,Z_s} = p_{s,Z_s}$, and the hypothesis is rejected if $P(\hat{p}_{s,Z_s} | p_{s,Z_s})$ (= $p$-value) is less than 5%. In this case, all charging sessions during which the favorite station was unavailable are labeled as ‘overflow’. Once this process is complete, all charging dynamics have been mapped and can be summarized in an origin-destination ‘overflow matrix’, which shows for each station how many overflow sessions were found at any other station.
4.3 Step 3: Simulation

Lastly, the effect of an increase in demand on the charging network is simulated. This is done in two steps as shown in Figure 4. In the first step, new charging sessions are randomly generated. Given the demand increase \( \Delta D \), the number of charging sessions that need to be generated per station is calculated based on the share of sessions that the station was responsible for in the entire dataset. Next, fictive charging sessions are generated by randomly sampling a \((\text{Start time}, \text{Duration})\) value, and a \text{DayOfYear} value from the station’s empirical distribution. Figure 5 and Figure 6 show these distributions for one station in the dataset. The pattern in Figure 5 highlights the importance of simultaneously drawing start time and duration values, as both are highly correlated to each other (e.g., sessions that start in the evening are more likely to last until the next morning, and thus have a longer duration).
In the second step, all charging sessions (both the existing and newly generated) are sorted on their start time and iteratively passed through to verify their status. In the baseline model, only two statuses are possible: either the charging station where the session is planned to happen is available ("OK" status), or the station is fully occupied, and the session cannot be executed ("FAILED" status). However, in the overflow model, we consider that EV drivers will divert to other nearby charging stations when a preferred station is unavailable (see shaded area in Figure 4). For each charging station, all other stations that can capture overflow (and their order of preference) can be retrieved from the overflow matrix (see section 4.2). After running the simulation, investigating the failed sessions per station is of particular interest, as this allows us to determine how many more CPs would be required to resolve them.

5 RESULTS

5.1 Charging zones

The charging zones are visualized for two areas in Brussels in Figure 7 and Figure 8 below, using a threshold walking time $t_{\text{max}} = 15$ minutes and a minimum support $ev_{\text{min}} = 4$ EV drivers. For Brussels, in total 413 zones are found of sizes between 1 to 5 stations. The minimum support level largely depends on the size and maturity of the dataset. Given that only one year of charging data is available, we choose 4 EV drivers as the threshold, however, sensitivity analysis on this parameter could shed more light on its optimal value.

Displaying the charging network in terms of the different zones provides new insights for planners when deciding on the location of new stations. Each zone indicates that all the charging stations within that zone are being visited by a shared group of EV drivers. As such, these stations may serve as an
alternative for each other when one is fully occupied. Although Definition 1 ensures that one zone can never be a full subset of another zone, many are found to be partially overlapping. This reflects the fact that a single charging station can serve multiple groups of EV drivers operating in its near surroundings. In areas where the density of charging stations is high (Figure 8), more partially overlapping zones are found.

5.2 Overflow dynamics

The overflow dynamics are visually represented in overflow plots, as shown in Figure 9. The nodes in the graph represent the charging stations (respective to their geographical location), and the arcs between them indicate the number of overflow sessions that have been identified between the stations. The size of the nodes is proportional to the charging activity measured at that station and the width of the arcs with the number of overflow sessions. The arcs always originate at the node with the same color as the arc. For instance, 134 overflow sessions have been identified that took place at station D, only because station C was occupied. Some overflow originating from station C is even captured by the more remote stations B and A. Figure 10 shows the same overflow plot for one of the charging zones in the high-density station area from Figure 8. Given the high density of the charging stations in this area, many overflow relations are found.

The overflow plots also show that nearby charging stations tend to capture more overflow sessions than remote stations, reflecting the preference of EV drivers to charge their vehicles as close as possible to their initially preferred charging station. The distribution of distances between the preferred station and the eventually chosen overflow station is presented in Figure 11. The mean distance is 402m (median 362m) and 95% of all overflow sessions occurred within 788m of the preferred station. Hence, charging stations located more than roughly 800m away from each other are unlikely to be used as alternatives, and will not have an overlapping catchment area anymore. These plots also give insight into the cannibalization potential (i.e., one charging station attracting charging sessions at the expense of another), which can be a useful tool for urban planners when evaluating candidate sites for installing new stations.

Figure 9 Overflow plot for the 4-station charging zone in Figure 7.
Figure 10 Overflow plot for one of the 4-station charging zones in Figure 8.

Figure 12 depicts the relationship between the number of overflow sessions and the utilization rate of a charging station. To preserve the confidentiality of the data, utilization is min-max normalized between 0 and 1 (1 representing the highest utilization found and 0 the lowest). Interestingly, the scatterplot shows an exponential relationship between both variables. Overflow sessions are rarely found for stations with a low utilization level, however as utilization increases the number of overflow sessions tends to increase more than proportional. High observed utilization does not guarantee more overflow as this will still depend on the number of nearby stations available to capture overflow. Highly remote charging stations will always have zero overflow sessions, as no nearby station is available to capture demand.

5.3 Future growth scenarios

Once the overflow dynamics between charging stations have been mapped, they can be used as input in the simulation model. Figure 13 below shows the percentage of all sessions labeled as failed for different demand increase scenarios while keeping the charging infrastructure supply at the same level. The percentage of failed sessions is consistently lower in the overflow model compared to the baseline.
model, as the former assumes EV drivers to use a nearby overflow station (if available) when their preferred station is occupied.

![Figure 13 Percentage of failed sessions per demand increase.](image1)

Given all failed sessions in a certain growth scenario, it is possible to determine the minimum number of new CPs that are required to reach $x\%$ failed sessions (assuming that the CPs that capture the most failed sessions will be installed first). As each additional CP that is installed will capture a certain share of failed sessions, Figure 14 shows the remaining percentage of failed sessions after a given number of CPs have been installed, for a given demand increase ($\Delta D$) of 20%. The overflow model illustrates that the same reduction in failed sessions can be achieved with fewer CPs. Figure 15 shows the required number of CPs in the overflow model for different demand growth scenarios. All curves show to have a long tail, skewed to the right. This means that to fully reduce the number of failed sessions to zero, a significant number of CPs would be required, most of which would capture only very few failed sessions.

![Figure 14 Additional CPs required for a 20% demand increase (overflow vs. baseline).](image2)
The exact number of new CPs that are required will depend on the failure rate that policymakers are willing to accept ($r_{\text{max}}$). This is summarized in Table 3, where for a given demand increase and failure rate, the required number of CPs is given both in the overflow and the baseline model. The table shows that overall, when considering overflow capacity at nearby charging stations, between 32% to even 100% less CPs are required to reach the same failure rate. The latter case (i.e., $\Delta D = 20\%$, $r_{\text{max}} = 5\%$) indicates that the existing charging infrastructure network has sufficient overflow capacity available to capture the 20% demand increase while keeping the percentage of failed sessions under 5%, and thus no additional infrastructure is required. When considering the case where demand doubles ($\Delta D = 100\%$), between 30 to 50% fewer charging points are needed compared to when overflow dynamics are neglected while maintaining the same failure rate.

The locations of where to install the new CPs can be retrieved from the model as well. Figure 16 below shows for $\Delta D = 100\%$ and $r_{\text{max}} = 2\%$ the locations of the 319 additional CPs according to the overflow model. The map shows how many more charging points should be installed at each existing charging station in the network to meet a doubling of demand while keeping the failure rate below 2%.

<table>
<thead>
<tr>
<th>Demand increase ($\Delta D$)</th>
<th>Required # CPs</th>
<th>Failure rate ($r_{\text{max}}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>40%</td>
<td>322</td>
</tr>
<tr>
<td></td>
<td>60%</td>
<td>425</td>
</tr>
<tr>
<td></td>
<td>80%</td>
<td>527</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>620</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1%</th>
<th>2%</th>
<th>5%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BL OF $\Delta$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>133</td>
<td>28</td>
<td>79%</td>
</tr>
<tr>
<td></td>
<td>2%</td>
<td>245</td>
<td>95</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>47</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>135</td>
<td>21</td>
<td>84%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>208</td>
<td>66</td>
<td>68%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>280</td>
<td>122</td>
<td>56%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>347</td>
<td>177</td>
<td>49%</td>
</tr>
</tbody>
</table>

Table 3 Required number of CPs per demand increase and per failure rate (OF= overflow; BL=baseline).
1 6 DISCUSSION

By grouping charging stations into zones and mining the individual charging behavior of EV drivers, we find underlying overflow dynamics between charging stations. These dynamics reveal how EV drivers behave when a preferred station is unavailable and which stations capture overflow sessions from each other. When overflow is found at a charging station, this also indicates that the true demand for this station is higher than what was observed from the charging data. Hüttel, et al. [28] found that this censorship bias occurs up to 61% of the time in some areas in Copenhagen, and that it decreases the performance of demand estimation models. Our study can alleviate this bias as the overflow dynamics give a more accurate view of the actual demand for charging stations.

Interestingly, it was found that the number of overflow sessions measured at a station increases more than proportional to the utilization rate of that station. Similar results were found by Helmus, et al. [16] who reported a non-linear relationship between system inconvenience (i.e. failed connection attempts) on the one hand and the number of EV drivers in the charging network on the other hand. However, a fundamental difference between the work of Helmus, et al. [16] and the model presented in this paper, is that Helmus, et al. [16] measure inconvenience indirectly from running agent-based simulation models, while in this paper we retrieve overflow sessions directly from mining the observed charging behavior at the level of the individual EV driver.

Policymakers and CPOs can use the concept of overflow sessions to assess possible cannibalization when installing new charging stations nearby existing ones. We furthermore found that charging stations located more than roughly 800m away from each other are unlikely to be used as alternatives for each other. This threshold distance can be used as input for urban planners when creating a charging network that is based on coverage (such as in e.g. [29], [30] and [31]). CPOs can furthermore use this threshold to determine when a competing charging station is operating in the same catchment area as an existing one.

Our models show that considerably fewer CPs are required when incorporating overflow dynamics. For example, for a doubling of the demand, 30-50% fewer charging points are needed compared to when overflow dynamics are neglected (see Table 3). Ignoring the charging behavior of EV drivers overestimates the number of CPs required and results in overinvestments. At first sight, this might be nuanced by the fact that the overflow simulation model assumes that all EV drivers are willing to charge at any nearby available overflow station. As in reality, this may not hold true for all EV drivers, the true number of required CPs might be higher than reported in the overflow model. Nevertheless, the...
concept of overflow dynamics only considers the ‘instant’ substitution between charging stations of the same zone. Depending on the state-of-charge of the vehicle and the range anxiety of the EV driver, some drivers may decide to delay the charging session elsewhere in the city at a different time or to consolidate it with the next charging session. This ‘delayed’ and ‘consolidate’ type of substitution is not considered in the model and could further decrease the required number of CPs.

Determining the exact number of new CPs to be installed does not solely depend on the future growth of EVs but is also determined by the failure rate policymakers are willing to accept. In the extreme case of reaching 0% failed session, a considerable number of CPs will need to be installed as demonstrated by Figure 15. This is not advisable as large investments would be required, while the utilization would be extremely low because of the small share of failed sessions that each additional CP would capture. Although a too-high failure rate can result in congested charging stations and user inconvenience, having a minimum rate might be desirable. Previous research has indicated the mixed usage of charging stations for both parking and charging [21, 32], and thus not every transaction reflects an inherent demand for charging. Policymakers should set a failure rate that reflects their trade-off between user convenience (i.e., always being able to charge) and utilization (i.e., having charging stations that are sufficiently used).

The simulation model decides where and how many charging stations should be installed based on reducing the number of failed charging sessions. As this will always be region-specific, the results as presented in Table 3 and Figure 16 should not directly be generalized to other cities. However, the model itself can easily be calibrated for any city, given that transactional charging data is available. Furthermore, our model can be extended to other performance metrics besides the failure rate. For CPOs, it might be more interesting to see how many more charging stations can be installed while keeping the expected utilization or consumed volume (kWh) of the new CPs sufficiently high. Instead of deciding on a maximum allowable failed rate, a minimum required utilization rate or volume can be set. The overflow relations between charging stations are important to consider as they reduce the costs of expanding the charging network while achieving the same level of performance.

7 CONCLUSIONS

While most research on modeling EV charging infrastructure has been focused on designing an initial charging network, this study already looks further to see how existing charging networks can be upgraded over time to keep up with future EV demand. We analyze one full year of charging transactions at 197 charging stations (394 CPs) and demonstrate how this can be used to determine (1) how many CPs are needed for future growth scenarios, and (2) where those CPs should be located.

The main findings of our model can be summarized as follows. First, we present a novel method to empirically extract overflow dynamics between charging stations from transactional charging data. This enhances our understanding of charging networks and shows that to understand the charging behavior at one station, it is necessary to also consider what happens at other nearby stations. Given the detailed spatial level of the overflow plots, they can furthermore be used as a tool to evaluate candidate sites when deciding to install new charging stations. Second, we demonstrate the importance of incorporating overflow dynamics between stations. As demand increases, the required number of charging stations is found to be considerably lower when considering that EV drivers will charge at other nearby stations when a preferred station is unavailable. This reflects the ability of the charging network to capture more demand due to the available overflow capacity. Third, we show how the exact number of CPs required depends on the allowable failed rate, which is a trade-off between convenience and utilization that needs to be set by policymakers.

Finally, our research has some limitations. The simulation models will decide on the optimal number of chargers and their locations based on the failed sessions registered at different stations. The recommended locations for installing new chargers are thus bound to the existing locations of the stations in the network. Hence, we recommend policymakers also install charging stations at new locations as a way to measure demand and subsequently use our method to expand the charging network in a data-driven manner. Besides, the simulation of fictive charging sessions in our study is straightforward and
relied on re-sampling existing sessions. More advanced methods (e.g., see [33]) are available, and we plan to incorporate them in our future work. Similarly, we assume that each EV driver has a favorite station, found as the most used station in the specific zone. This may simplify reality as EV drivers can have no (e.g., some visitors) or even multiple favorite stations within a specific zone. A possible solution would be to differentiate between regular and random EV drivers [34] and conduct a more extensive analysis of their charging station choice.

In future work, we plan to incorporate new evaluation metrics into the model (e.g., utilization rate and consumed volume), investigate whether an optimal failure rate can be found, and extend the model to the San Francisco region (which has a different charging location policy). Besides, there are several other promising lines for future research to investigate. First, it would be useful to analyze the trade-off between installing new stations to serve overflow or installing additional plugs directly at existing stations. Second, the concept of overflow dynamics could be of interest for map services to more accurately direct EV drivers to available charging stations. Third, the model can be extended to rural areas and DC fast chargers. Finally, it would be interesting to combine our model with traditional forecasting methods.

ACKNOWLEDGMENTS
We acknowledge TotalEnergies for their contribution to the data collection.

FUNDING
S.W. acknowledges fellowships from the Research Foundation Flanders (FWO-Vlaanderen) under Grant No. 11I5122N and Grant No. V411523N.

AUTHOR CONTRIBUTIONS
The authors confirm their contribution to the paper as follows: Study conception and design: SW, GT, LV; Data collection: SW, LV; Analysis and interpretation of results: SW, GT; Draft manuscript preparation: SW, GT. All authors reviewed the results and approved the final version of the manuscript.

CONFLICTS OF INTEREST
Declaration of interest: none.
REFERENCES


S. Weekx, C. De Cauwer, and L. Vanhaverbeke, "One charging station is not the other: spatial-temporal analysis of public charging stations," in *36th International Electric Vehicle Symposium and Exhibition (EVS36)*, Sacramento, California, USA, 2023, vol. 36.


