Transportation Research Part C 115 (2020) 102637

Contents lists available at ScienceDirect

Transportation Research Part C

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

A data driven typology of electric vehicle user types and charging sessions

Jurjen R. Helmus[a,b,*], Michael H. Lees[b], Robert van den Hoed[b]

[a] *Computational Science Lab, University of Amsterdam, Netherlands*
[b] *Urban Technology, Amsterdam University of Applied Sciences, Netherlands*

ARTICLE INFO

Keywords:
Charging behavior
Electric vehicles
EV user types
Charging infrastructure
Typology
Simulation

ABSTRACT

The understanding of charging behavior has been recognized as a crucial element in optimizing roll out of charging infrastructure. While current literature provides charging choices and categorizations of charging behavior, these seem oversimplified and limitedly based on charging data.

In this research we provide a typology of charging behavior and electric vehicle user types based on 4.9 million charging transactions from January 2017 until March 2019 and 27,000 users on 7079 Charging Points the public level 2 charging infrastructure of 4 largest cities and metropolitan areas of the Netherlands.

We overcome predefined stereotypical expectations of user behavior by using a bottom-up data driven two-step clustering approach that first clusters charging sessions and thereafter portfolios of charging sessions per user. From the first clustering (Gaussian Mixture) 13 distinct charging session types were found; 7 types of daytime charging sessions (4 short, 3 medium duration) and 6 types of overnight charging sessions. The second clustering (Partition Around Medoids) clustering result in 9 user types based on their distinct portfolio of charging session types. We found (i) 3 daytime office hours charging user types (ii) 3 overnight user types and (iii) 3 non-typical user types (mixed day and overnight chargers, visitors and car sharing). Three user types show significant peaks at larger battery sizes which affects the time between sessions. Results show that none of the user types display solely stereotypical behavior as the range of behaviors is more varied and more subtle. Analysis of population composition over time revealed that large battery users increase over time in the population. From this we expect that shifts charging portfolios will be observed in future, while the types of charging remain stable.

1. Introduction

In recent years, electric mobility has gained a great deal of attention, leading to significant growth of the Electric Vehicle (EV) market and complementary development of necessary charging infrastructure. A large number of cities have developed extensive public charging infrastructures to facilitate the uptake for electric mobility. Local policy makers are faced with the challenge to provide sufficient accessibility to public chargers for prospective EV users, while balancing the investments required for charging infrastructure and considering public space, scarce parking resources and impact when developing exclusive charging spots. With the expected surge of electric mobility in the coming years, planning and rolling out a cost-effective charging infrastructure that suits the

* Corresponding author at: Amsterdam University of Applied Sciences, Urban Technology, Weesperzijde 190, 1097DZ Amsterdam, Netherlands.
E-mail address: j.r.helmus@hva.nl (J.R. Helmus).

https://doi.org/10.1016/j.trc.2020.102637
Received 6 November 2019; Received in revised form 27 March 2020; Accepted 28 March 2020
Available online 25 April 2020
0968-090X/

Nomenclature

Glassary table

Acronym	**Description**
AC	Alternating Current
AIC	Akaike Information Criterion
ARI	Adjusted Rand Index
BEV	(full) Battery Eclectic Vehicle
BIC	Bayesian Information Criterion
CP	Charging Point
CPO	Charging Point Operator
DBS	Distance Between Sessions
DC	Direct Current
EM	Expectation Maximization
EV	Electric Vehicle
EVV	Equal
GMM	Gaussian Mixture Model
HBS	Hours Between Sessions
PAM	Partition Around Medoid clustering
PHEV	Plugin Hybrid Electric Vehicle
POI	Point of Interest
SOCVVV	State Of Charging

charging needs for EV users is of crucial importance.

A crucial element in optimizing roll out of charging infrastructure is charging behavior of EV users. According to Morrissey et al (2016) an understanding of charging behavior of EV users at a variety of charge point types is essential in order to ensure that the charging needs of a growing EV population are catered for, and in order to choose the optimal charging infrastructure rollout plan in terms of economic and practical effectiveness. Therefore, gaining insight in users types allows researchers to simulate the rollout of charging infrastructure to a predicted local demand of EV users more accurately based on their activities (Momtazpour et al., 2012).

Extracting rules of behavior for EV user charging can also be used to model and simulate future scenarios, for instance in agent based simulations (Sweda and Klabjan, 2011; Xydas et al., 2016a, 2016b). Understanding how these behavioral rules differ between different EV users allows rich forms of scenario testing, for example the effect of changing population compositions on charging infrastructure utilization. From these insights policy makers can be supported in anticipating charging infrastructure requirements and optimize investments.

While the importance of charging behavior has been acknowledged in research (Azadfar et al., 2015; Xydas et al., 2016a, 2016b; Daina et al., 2017a; Weldon et al., 2018), little attention has been paid in trying to define different types of EV users types based on their charging behavior. Current literature lacks a clear set of rules to describe when, where and how long different types of EV users will charge. Studies that have addressed typologies of charging behavior contain simplified categories based on simplistic stereotypes such as visitors, residents and office chargers. While in reality, there are likely to be different subtypes of visitors, residents and commuters with distinctly different charging behavior. Also, EV users may display both charging behavior of visitors (e.g. short sessions during daytime) as well as residents (overnight charging). Furthermore, in most studies of EV users, the categories are based on assumptions of these stereotypes, and very rarely are these stereotypes tested against real historical charging data.

The aim of this paper is to develop a typology of EV users by examining real-world data that captures charging behavior. A user typology characterized through behavior provides opportunities to build an agent-based model (Bonabeau, 2002) in which the rules describing the user behavior can be defined per user type. With a good understanding of charging behavior and different typical EV users, academics may develop models that in the end help policy makers to improve understanding of charging infrastructure in their cities. To overcome the aforementioned simplifications used in previous work, we suggest that EV user types should be conceptualized as a portfolio of underlying behavioral outcomes (charging sessions).

This paper starts by reviewing state of the art research regarding EV user charging behavior. Next we develop a typology of user types based on a two-step clustering: (i) clustering to define a typology of charging sessions, (ii) clustering of portfolios of session types per user to define user types. The distinct features of each typology are discussed to gain a better understanding of the behavior. Subsequently, both typologies are related to development over time in order to find relations between behavior and charging infrastructure maturity. The paper works towards a set of detailed EV user types and provides a structured approach how to achieve these types. As such, it contributes to scientific knowledge on charging behavior that allows to be built upon in future work.

2. State of literature on charging behavior

Charging behavior research in recent years has developed from potential limitations of using EVs in the context of limited charging infrastructures (Guo et al., 2012; Neubauer and Wood, 2014; Yang et al., 2016) towards insights in how EV users make charging decisions in a mature charging infrastructure environment with joint charging modes (Xu et al., 2017; Wolbertus and Van den Hoed, 2019a). Researchers have investigated charging behavior using various approaches. On one hand researchers focused on choice behavior, psychological and socio-demographic factors that influence charging behavior (Daina et al., 2017a, 2017b). Others have analyzed actual charging sessions using statistical methods (Haidar and Muttaqi, 2015; Wolbertus et al., 2018a). Due to adoption of EVs over time, recent research tends to use actual charging behavior (revealed preferences) of EV users (Khoo et al., 2014; Morrissey, Weldon and O'Mahony, 2016), rather than experiments with stated preferences (Daina et al., 2017b). A common factor is that research is focused on better modelling of charging choices, while acknowledging that user behavior may vary amongst EV users. This research focuses on identifying models for behavioral choices for different types of EV users by using EV users actual charging data.

For the ease of reading we focus on the categorization of user types in our literature overview. For completeness we included an overview of literature on charging behavior choice modelling and data driven research on charging behavior in Appendix C.

2.1. Categorization of charging behavior and user types

In recent years work has been done in classifying particular types of charging behavior. In order to characterize charging demand Xydas (Xydas et al., 2016, 2016b) performed a k-means cluster analysis on connection profiles based on charging data from 21,918 charging events and 255 chargers in the UK. In total 14 clusters for power demand over 3 different regions were found, indicating that different locations may lead to different patterns of behavior. A further exploration on charging behavior of user type was not performed as the paper aimed to analyze risks of EV adoption for the electricity grid.

De Gennaro et al (2015) defined potential charging strategies as types of charging sessions based on parking activities. Although their research was based on GPS data from conventional vehicles, it does provide insight in different types of charging behavior. Examples are short stay charging at public parking spaces near shopping locations and long stay overnight charging in residential areas. Moreover, rather than defining charging behavior of a single user, they regard charging behavior at specific locations as a portfolio of charging sessions types. As the paper focuses on the perspective of the charging points rather than the users themselves the work does not elaborate on defining user types.

Other studies on charging behavior reveal a classification of charging behavior in fixed and flexible charging behavior (Bae and Kwasinski, 2012; Yang et al., 2016). The former relates to charging at fixed times and locations, such as home and work charging, the latter relates to charging during trips. The concept of regular versus irregular behavior is also found in the work of Kim et al. (Kim et al., 2017) whose model specifically focuses on inter charging times at public charging infrastructure. Based on charging data from Dutch municipalities they developed two different models for inter charging times; one for regular users and one for random users.

A drawback of several studies is that charging behavior is described with parametric statistics, whereas user behavior is much more complex than simple mean behavior of a user in a single dimension. Other studies have included heterogeneity in both user types and charging decisions using mixed logit models (Jabeen et al., 2013; Yu and MacKenzie, 2016). These models enable quantification of effects on charging decision at a user level (e.g. company/private cars) or contexts (day of week, time). This emphasizes that charging behavior should be considered in more sophisticated ways, for example considering a full distribution of arrival times, rather than parametric estimation, at least in the time direction.

In conclusion, although there are some categorizations of charging behavior, they seem oversimplified. They tend to focus on (i) limited number of properties of charging sessions, (ii) distinct users based on their mean behavior rather a distribution of behavior, and (iii) the allocation of DC charging on highways versus AC charging within cities.

2.2. Contributions of current work

Recent qualitative and data related research has increased our understanding of charging behavior has increased a great deal in recent years. It has led to valuable conceptualizations and models of charging behavior and its determining factors. Nevertheless, a classification of EV user types is lacking and current research largely uses simplified models of EV users as proxy of behavior. Particularly in simulation literature, EV users are often characterized as having stereotypical behavior, i.e. a single behavioral rule that governs their overall behavior (e.g., overnight charger, office charger) (Xydas et al., 2016a, 2016b; Pagani et al., 2019). Yet, we expect that user types may display different types of charging sessions, rather than being reduced to solely stereotypical behavior.

In this paper we aim to contribute to the understanding of charging behavior in two ways. First, methodologically, we aim to capture the complexity of charging behavior of individual EV users by using a new clustering-based method for identifying portfolios of behavior. By doing so, we assert that EV users differ in sets of behavioral rules and probabilities for these rules to apply of which each rule leads to a specific type of charging activity. The sets of rules and their probabilities are regarded as a portfolio that differs amongst user types. The behavioral rules are applied with different probabilities depending on the context of the situation, which may even be the inter-dependence between the rules, i.e., rule A is far more likely after applying rule B.

Second, we identified and made online available rules for charging behavior types and EV user types by applying our method to a large geographically diverse dataset. The dataset contains data over several years in a wide geographic area that varies in maturity and charging network density. We therefore expect a variety of user types, potentially with different sets of behavior related to both external factors (charging infrastructure) as well as car properties.

3. Methodology

This research works towards a typology of EV user types. An EV user type is defined by its portfolio of distinct session types. Sessions types in turn are defined as distinct charging transactions that can be distinguished based on typical charging related variables such as time, location and frequency. Our methodology involves a two-step procedure where first a typology of charging sessions is developed based on charging session features. Thereafter, the activity of an EV user is characterized by a particular fraction of different session types, or a particular portfolio of session types. The second step is to develop a typology of EV users based on their charging portfolio, clustering users with similar fractions of session types together. Changes over time in user type population and charging behavior composition are analyzed and related to contextual factors of EV adoption and public charging infrastructure rollout.

In the remainder of this section we describe the dataset used in our approach, we then describe and justify the clustering methods used to define the session types and the clustering method used to define the user types.

3.1. Dataset of charging sessions

For this paper a dataset of more than 5.82 million charging transactions from January 2017 until March 2019 on public charging infrastructure from the Dutch metropolitan area was used (Helmus et al., 2018a, 2018b; Wolbertus et al., 2018b). It concerns sessions of 133 thousand EV users (approximately 60% of all EV users in NL) made on 7079 level 2 public Charging Stations (28% of all public level 2 CPs in the Netherlands, see Fig. A1 in the Appendix A). The dataset contains diverse population of infrastructure types; dense in large cities (55% of CPs in the dataset, limited private parking), diversely concentrated in cities (30% of CPs in the dataset), and sparse in rural environments (15% of CPs in the dataset) (Hoed et al., 2014; Kansen et al., 2018). The dataset contains CPs nearby

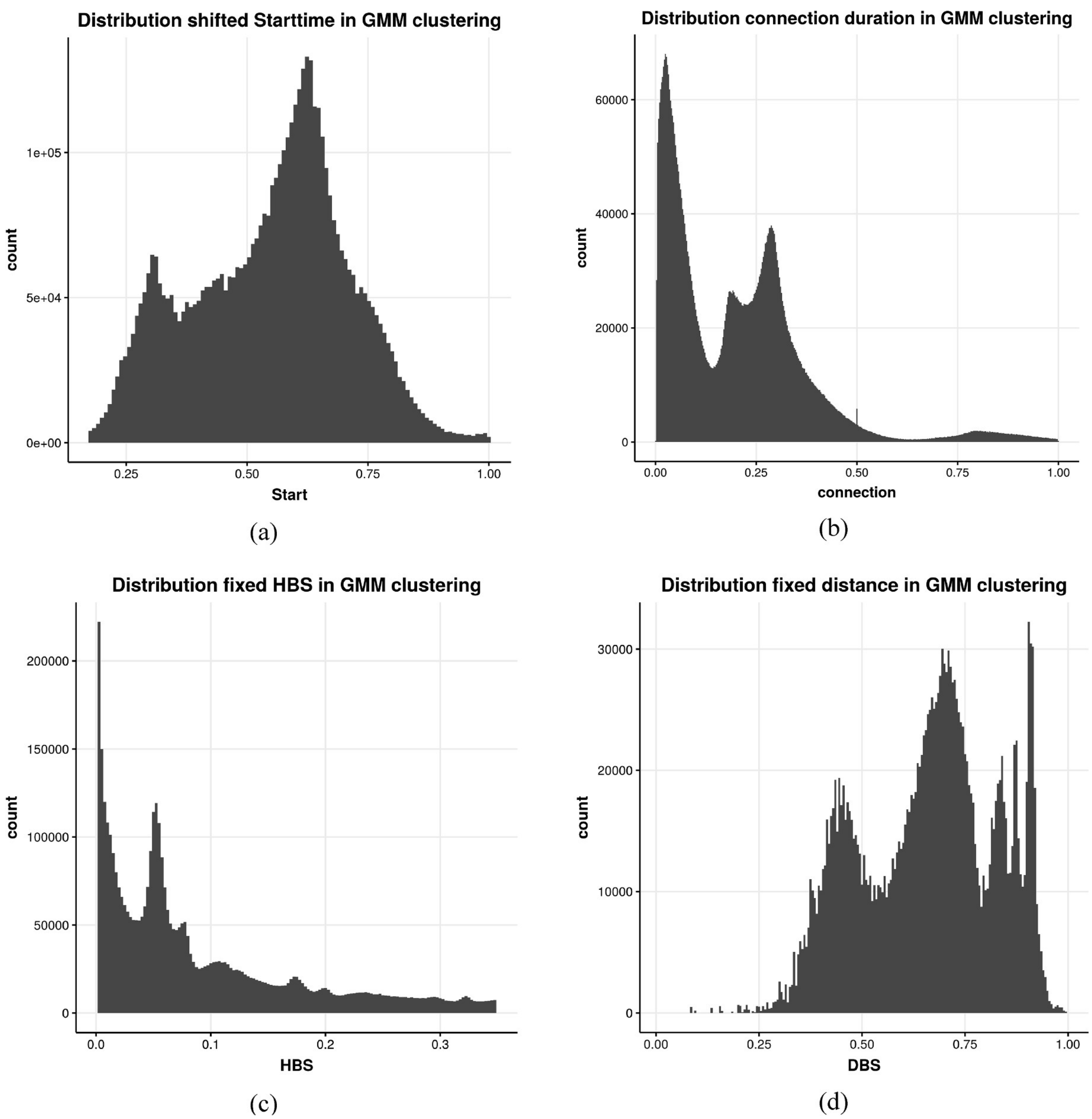
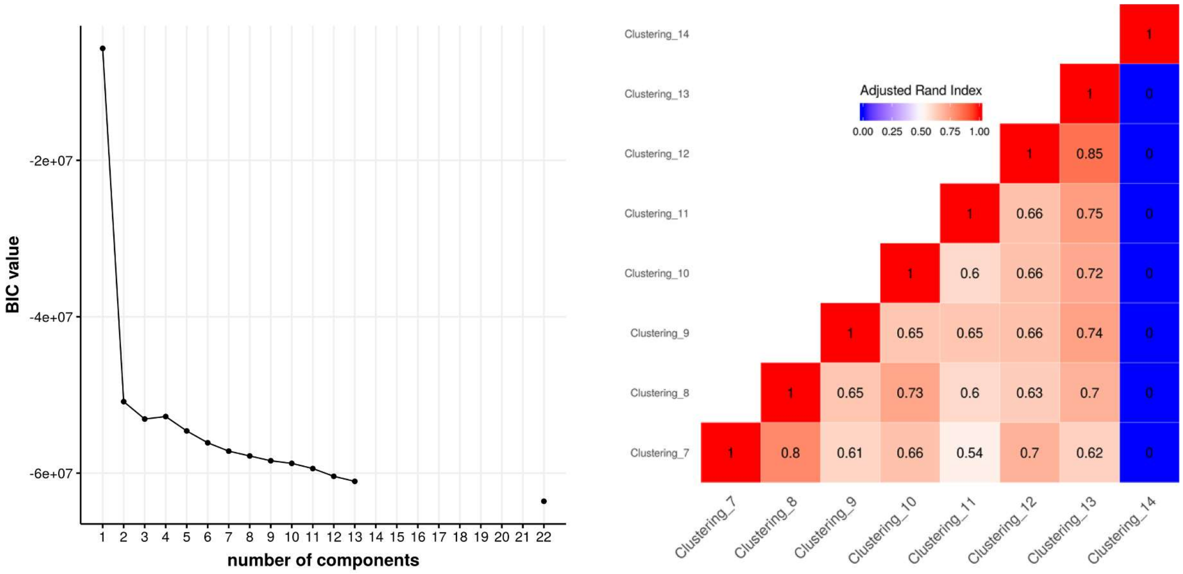


Fig. 1. Distribution of normalized variables start connection time (a), connection duration (b), hours between sessions (c), logdistance between sessions (limited from 0.001 to 1) (d).

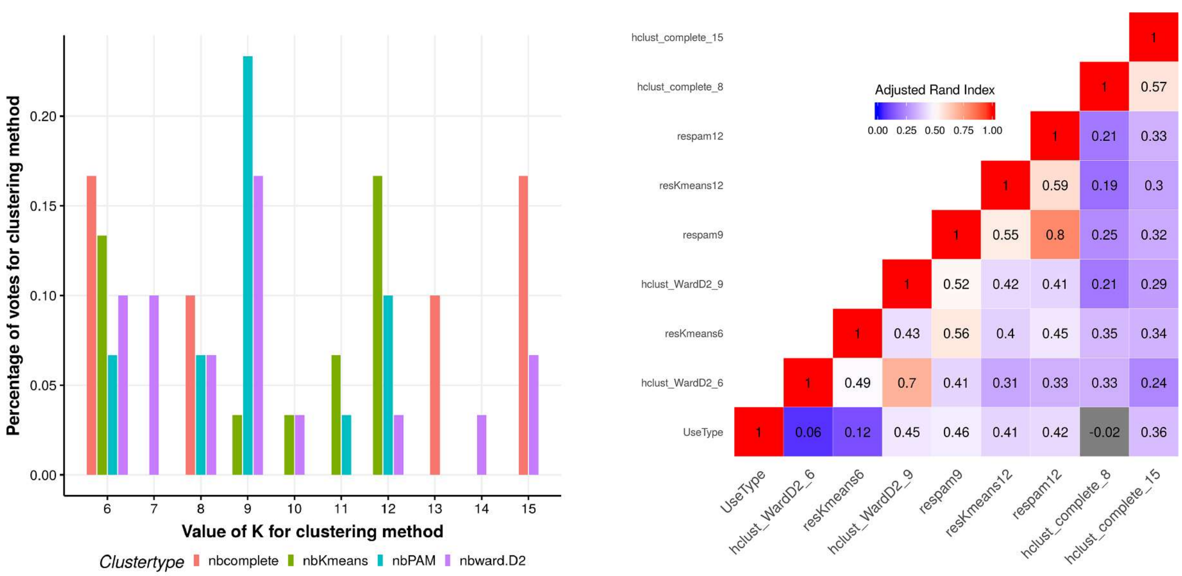


(a) (b)

Fig. 2. Result of BIC analysis for VVV model G = 1.0.22 with Bayesian Regularization(a), Adjusted Rand Index for VVV model G = 7 to 14 (b).

Table 1
Overview of user types, number of sessions and number of users.

UseType	Number of sessions	Percentage of total	Number of users	Percentage of total
Regular users	4047,080	94.8%	27,637	96.5%
Shared Fleet	116,764	2.7%	443	1.5%
Taxi	106,729	2.5%	552	2.0%
Total	**4,270,573**		**28,668**	



(a) (b)

Fig. 3. Number of high scores of quality indices for different clustering (a), ARI values of different clusterings and on the known UseTypes (b).

Table 2
Overview Session types, abbreviations and distribution properties.

Abbreviation	Time	Distance Between Sessions	Connection Duration	Hours Between Sessions
DT-SL-MD-NM	DayTime, Start norm dist μ 7:59, σ1.13 hrs End norm dist μ 16:47, σ1.17 hrs	Same Location, no distribution	Medium Duration, norm dist μ 8.80 hrs, σ1.03 hrs	Narrow Medium HBS, norm dist μ 15.3 hrs, σ 2.75 hrs
DT-DL-MD-WL	DayTime, Start norm dist μ 8:42, σ1.38 hrs End norm dist μ 16:35, σ2.27 hrs	Different Location, Multimodal, peaks a 1000 m, 21 m, 32 km 50 km	Medium Duration , Skewed left norm dist med 8.16, σ2.09	Wide Long, multiple modes, (i) 2 hrs, (ii) 16 hrs (iii) 40 hrs
DT-SL-MD-WL	DayTime, Start skewed right med 8:40, σ4.26 End norm dist μ 16:20, σ3.70 h	Same Location, no distribution	Medium Duration , Bimodal, peak at 2 hrs, peak at 8.10 hrs	Wide Long, multiple modes peaks 16 h + N-days*24 h
DT-DL-SD-NS	DayTime, Start norm dist μ 14:08, σ3.89 hrs End norm dist μ 15:35, σ40.49 hrs	Different Location, Multimodal, peaks a 1000 m, 21 m, 32 km 50 km	Short Duration, μ 15.3 hrs exp decay, full width half max 2.45 hrs	Narrow Short, Left skewed norm dist, μ 1.07 hrs, σ 0.5 hrs
DT-DL-SD-WS	DayTime, Start norm dist μ 14:30, σ4.02 hrs End norm dist μ 16:40, σ40.68 hrs	Different Location, Multimodal, peaks a 1000 m, 21 m, 32 km , 50 km	Short Duration, μ 15.3 hrs exp decay, full width half max 2.77 hrs	Wide Short, multiple modes peaks at 5,21,42 hrs
DT-DL-SD-WL	DayTime, Start norm dist μ 13:06, σ30.68 hrs End norm dist μ 16:10, σ30.92 hrs	Different Location, Multimodal, peaks a 1000 m, 21 m, 32 km, 50 km	Short Duration, μ 15.3 hrs exp decay, full width half max 3.17 hrs	Wide Long , multiple modes 3–7 * 24 h repeating for 4 weeks
DT-SL-SD-WS	DayTime, Start norm dist μ 13:47, σ3.63 hrs End norm dist μ 16:12, σ30.65 hrs	Same Location, no distribution	Short Duration, exp decay, full width half max 3.89 hrs	Wide Short, multiple modes three peaks at 17, 10, 21 hrs
ON-SL-MD-NM	OverNight, Start skewed right norm dist med 18:27, σ2.66 hrs, End skewed right norm dist med 7:32, σ1.17 hrs	Same Location, no distribution	Medium Duration, norm dist μ 12.75 hrs, σ1.99 hrs	Narrrow Medium, norm dist μ 12.75 hrs, σ1.99 hrs
ON-SL-LD-NS	OverNight, Start skewed right norm dist med 17:42, σ3.96 hrs End skewed right norm dist med 9:05, σ3.17 hrs	Same Location, no distribution	Long Duration, norm dist μ 40.11 hrs, σ3.75 hrs	Narrrow Short, multiple modes, Peaks at 2,12, 36, 50 hrs
ON-SL-MD-NL	OverNight, Start skewed right norm dist med 17:59, σ3.19 hrs End skewed right norm dist med 8:19, σ3.49 hrs	Same Location, no distribution	Medium Duration, norm dist μ 15.80 hrs, σ50.82 hrs	Multimodal pattern, skip N-days
ON-SL-MD-NS	OverNight, Start skewed right norm dist med 18:15, σ4.97 hrs End skewed right norm dist med 9:23, σ2.93 hrs	Same Location, no distribution	Medium Duration, norm dist μ 15.09 hrs, σ40.48 hrs	Narrow Short, Bimodal, short 2.5, 8 hrs
ON-DL-MD-NS	OverNight, Start skewed right norm dist med 18:47, σ5.43 hrs End skewed right norm dist med 8:16, σ2.12 hrs	Different Location, exp decay, full width half max 0.93 km	Medium Duration, norm dist μ 13.22 hrs, σ3.46 hrs	,Multimodal, 1 ,12, 24–36 hrs
ON-DL-MD-WL	OverNight, Start skewed right norm dist med 18:12, σ5.41 hrs End skewed right norm dist med 9:21, σ4.22 hrs	Different Location, exp decay, full width half max 0.88 km	Medium Duration, Bimodal, peak at 14 and 28 hrs	Wide Long Multimodal pattern, skip N days

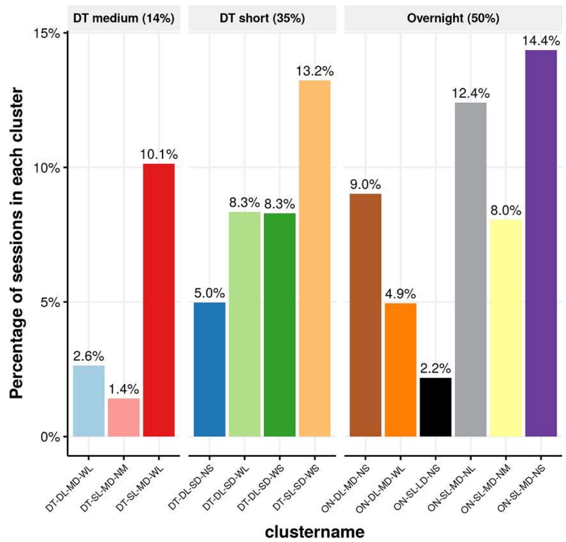


Fig. 4. Histogram of session types in percentage.

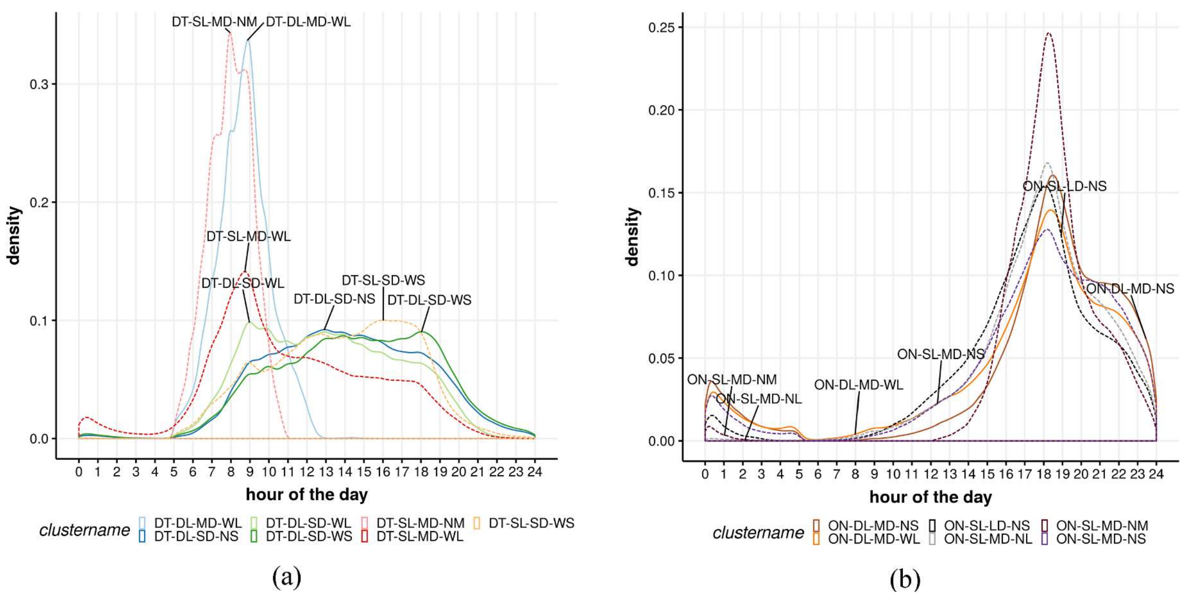


Fig. 5. Density plot of Start connection time for Daytime (a) and Overnight sessions (b).

Points of Interest (POIs) that may induce specific behavior. We do not expect this dataset to have any properties to be distinct from any other EV population.

This research is based on level 2 public charging and does not cover private and DC charging. This can be legitimized given that on average in the Netherlands 75% of households are dependent on public charging facilities (in metropolitan areas this share is likely higher) (Hoekstra and Refa, 2017; Kansen et al., 2018). Furthermore, research has shown that DC charging is limitedly used, and particularly to fill up EVs when necessary (Idaho National Laboratory, 2015; Wolbertus and Van den Hoed, 2019b). Also, DC is typically used for corridor charging and related to trip extension, while AC charging is related to activity patterns. In our research we focus on explaining behavior as related to typical activity patterns which relates to more complex behavioral patterns.

The available dataset does not contain a direct link between each EV user and its sociodemographic properties. Yet, it is known that 80–90% of the EVs in the Netherlands are company cars, either company owned or lease. As such, we expect that socio-demographic characteristics are biased towards higher income, larger mileage and higher educated persons (Nederland Elektrisch, no date; Hoekstra and Refa, 2017). We do expect a shift towards lower age and middle income as many less expensive EVs (e.g. Tesla Model3) have entered the market since 2018 (RVO, 2018).

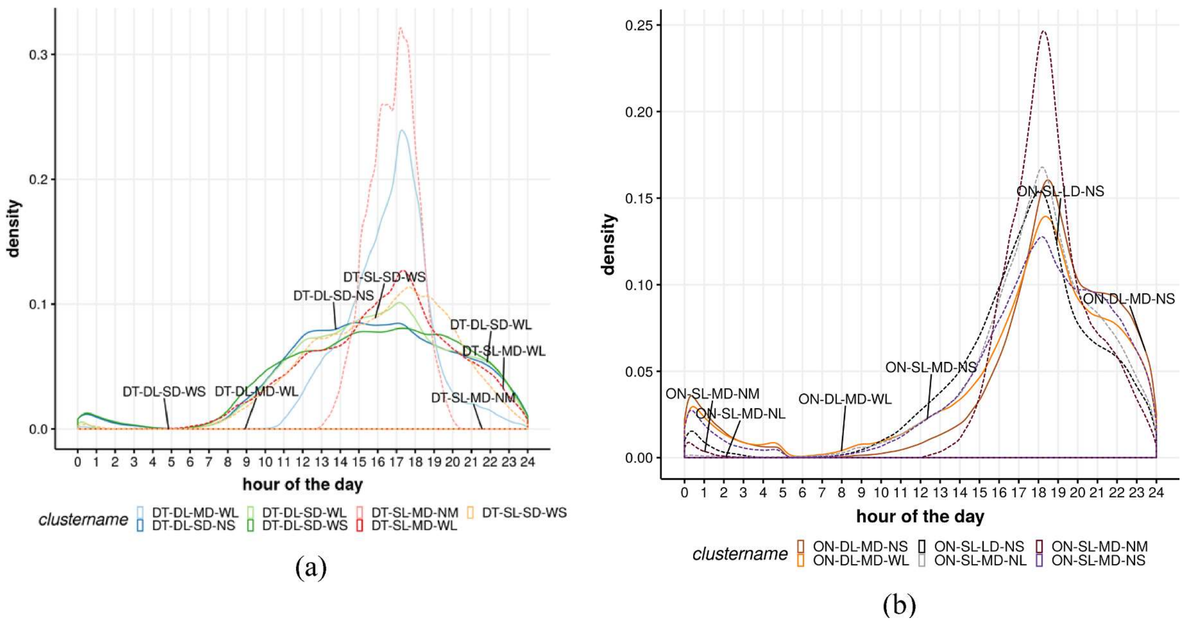


Fig. 6. Density plot of End connection time for Daytime (a) and Overnight (b).

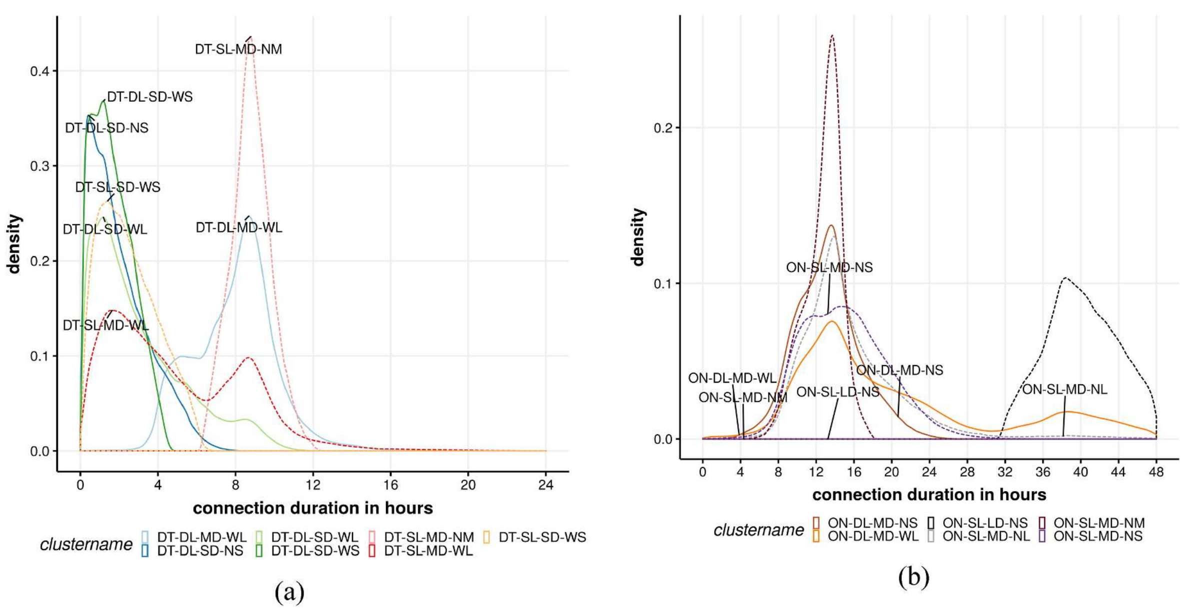


Fig. 7. Density plot of Connection duration for Daytime (a) and nighttime (b).

Pricing of public charging in the Netherlands consists of a flat fee per use (kWh) with a price cap of the Charging Point Provider (CPO) of at €0,28 per kWh. This allows us to investigate charging decisions without influence of tariffs.

Municipalities, being the owner of public charging infrastructure, have the power to collect charging transaction data including charging card information of the subgroups taxis and car sharing fleets. The IDs of car sharing fleets as well as taxis were made available to the municipality by specific owners as part of contract arrangements; and provided us with the opportunity to pinpoint them in our dataset. The data includes start and end time, charging card number, connection duration and energy uptake (in kWh). The dataset was cleaned using a set of filters as applied in earlier research resulting in removal of 132 thousand transactions with incorrect durations or negative values (extensively described in (Van Den Hoed et al., 2014)).

Previous research has shown that a dataset of charging sessions contains many users with very few number of sessions (e.g. infrequent visitors) (Kim et al., 2017; Helmus et al., 2018a, 2018b). It is difficult to set rules for behavior on user level if there is

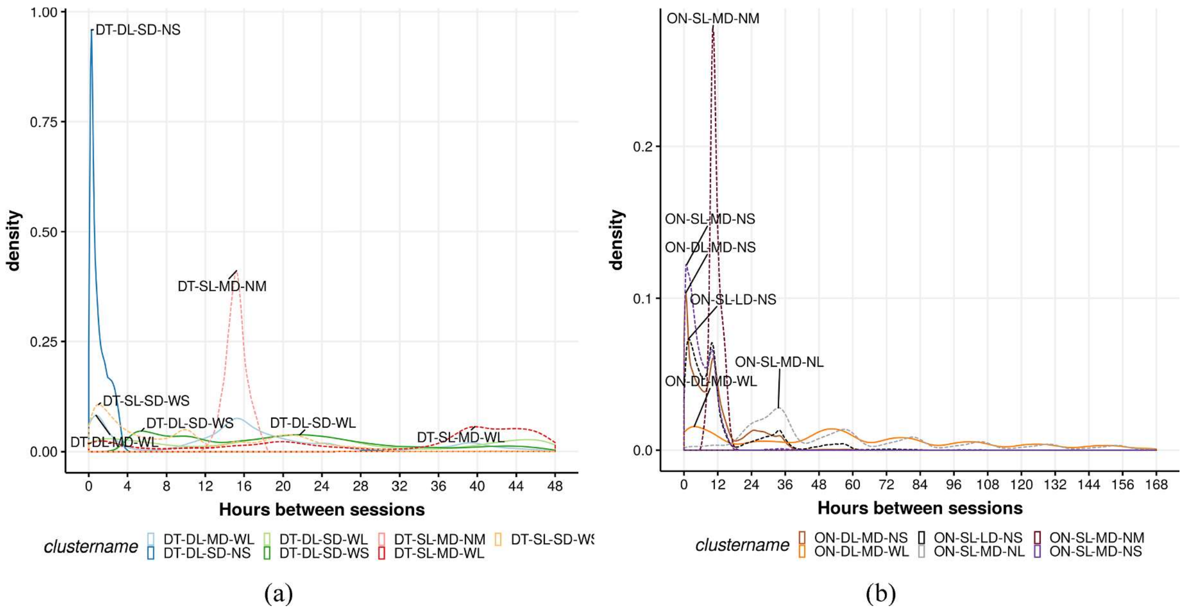


Fig. 8. Density plot of Hours Between Sessions (HBS) for Daytime (a) and nighttime (b).

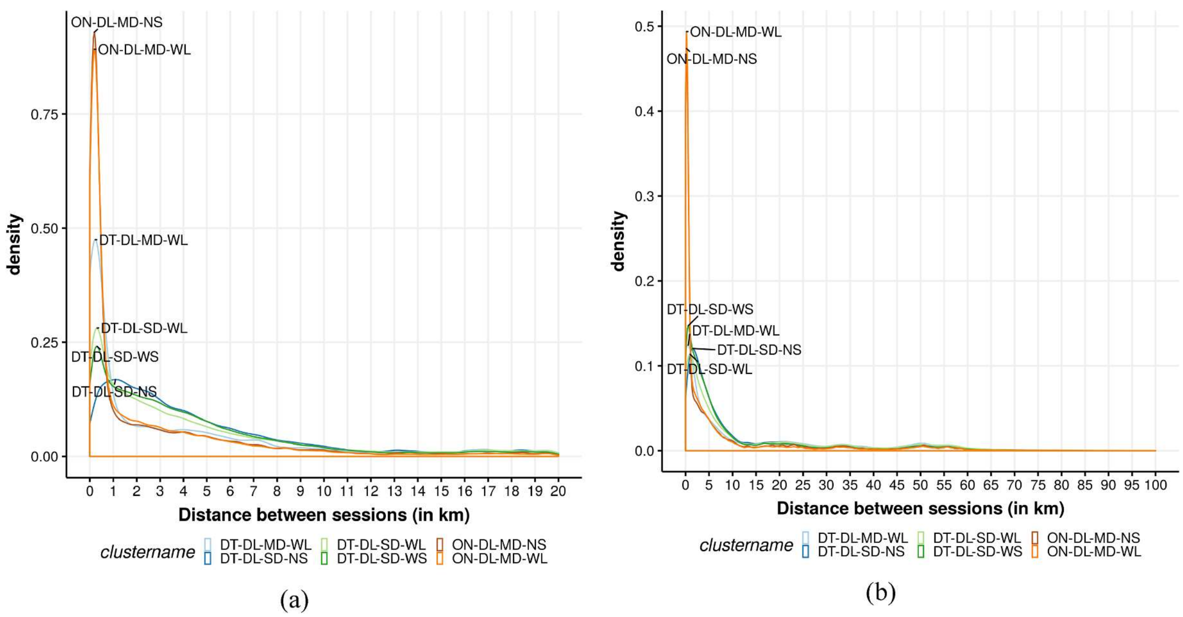


Fig. 9. Density plot of Distance Between Sessions DBS in kilometer (km) for 0–20 km (a) and 0–100 (km) (b).

limited data available for a user, since the rules for behavior cannot be accurately divided over the portfolio. For this reason, users with a limited amount of sessions were excluded from the analysis. To determine a threshold for filtering based on the minimum number sessions we analyzed behavioral stability using the following factors; (i) the number of locations visited (ii) mean and minimum charging frequency and (iii) variation the number of arrival hours. We investigated the effect of putting a threshold on the number of sessions for an EV user to be included in the analysis on these three factors (see Appendix A). We found that the first two factors were significantly affected by the threshold and that the effect diminished while increasing the threshold. We aimed to include as many users as possible, while having the least amount of noise in the dataset. Based on the analysis a threshold of 40 sessions per user was applied to avoid inferior clustering results (see Figs. B2–B5 in Appendix B). This result in removing 1.55 million sessions and 105 thousand users from the dataset.

After filtering using the threshold, a dataset of 4,270,573 charging sessions (75% of initial dataset) of 28,668 users (21% of initial dataset) was used for further analysis (see also Fig. B6 and B1 in Appendix B). The users in this dataset are considered regular (private) users (96.5%), complemented by in total 443 Shared Fleet cars (1,5% of users) and 552 taxis (2.0%).

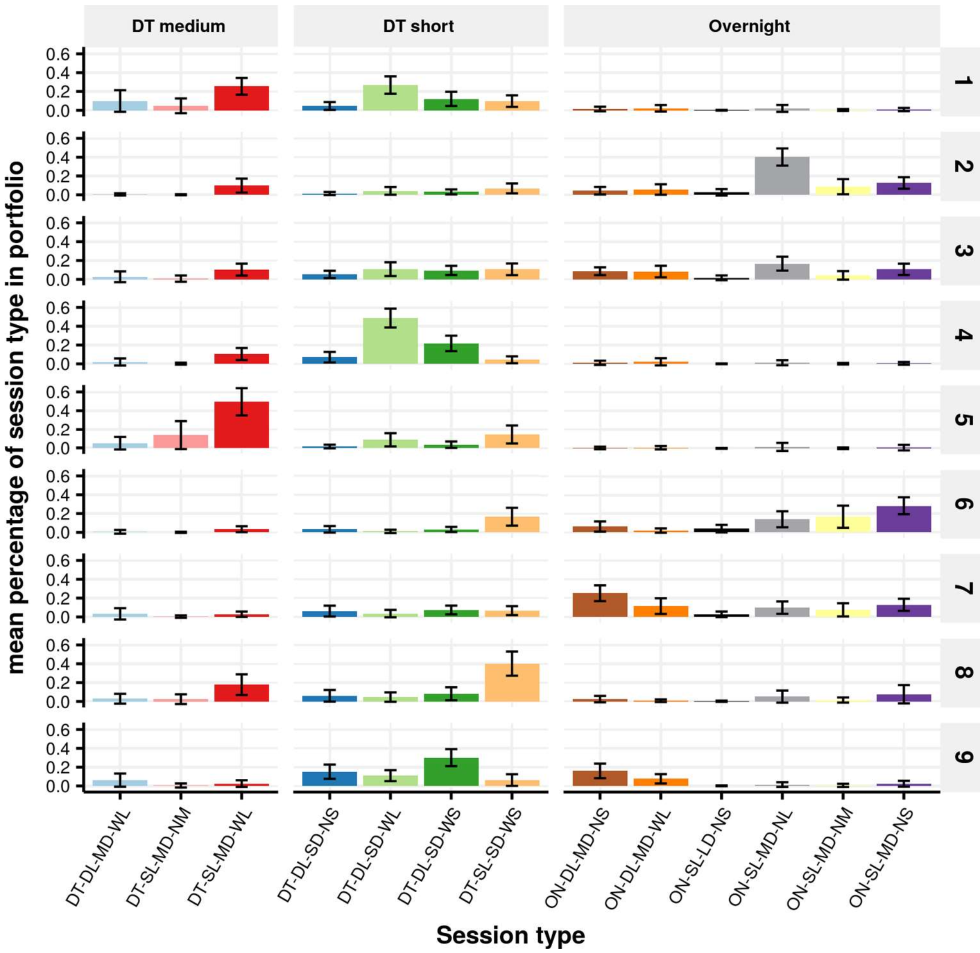


Fig. 10. Portfolio of charging sessions per user type (bars are based on the mean ratio of session type per user in the portfolio, error bars display the standard deviation per session types over all users in the cluster).

Table 3
User type distribution and suplementary properties per user type.

User type	Size	Percentage	# car sharing	# taxis	Weekly charging sessions	Mean nr CPs used	Mean estimated Battery size	Mean transaction Volume
1	2391	8.9	0	5	1.42	16.85	15.12	8.17
2	3570	13.2	0	38	1.86	8.32	23.48	14.07
3	3818	14.1	0	57	2.17	17.70	23.67	12.37
4	2281	8.4	0	5	0.94	29.43	13.53	6.22
5	2236	8.3	0	17	1.70	7.86	12.81	8.25
6	5676	21.0	1	163	4.60	11.02	15.71	8.88
7	4003	14.8	0	194	3.68	18.11	24.58	13.16
8	1525	5.7	0	25	4.14	13.57	12.68	7.13
9	1514	5.6	398	27	4.78	121.56	19.55	9.19

3.2. Conceptualization of charging sessions

We regard a charging session as a decision of an EV user to connect their vehicle at a particular time and place for a specific duration in relation to a previous charging transaction. We relate each charging session to a session's predecessor in time (time between the current and the previous session) and space (distance between the current and the previous session). Each session is defined by of four characterizing features: (i) start time of the connection, (ii) connection duration, (iii) distance between two sessions in meters (DBS) and (iv) Hours Between Sessions (HBS).

Although including the energy uptake (in kWh) may seem obvious, we chose not to include energy uptake as primary feature during clustering, but as supplementary feature for several reasons. Energy uptake is seen as a mediating variable between battery

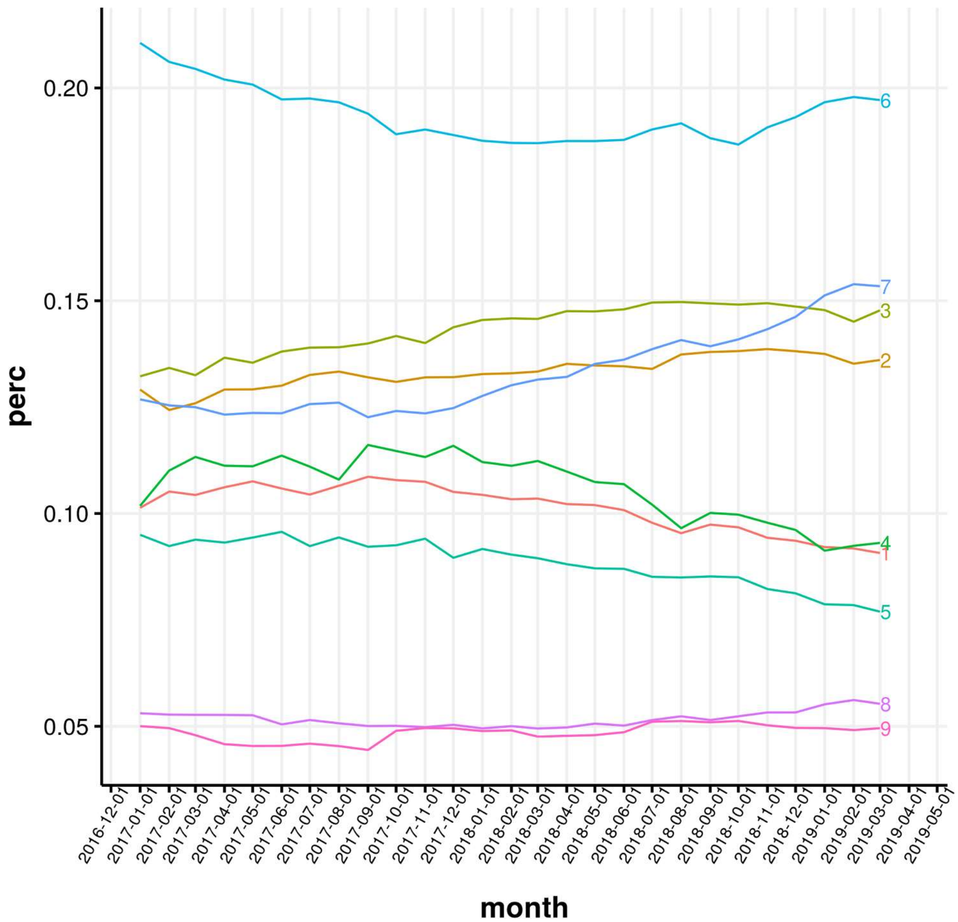


Fig. 11. Development of user types over time; each line is a percentage of monthly active users.

size and observed charging behavior, rather than a feature of the behavior itself. The energy uptake per transaction is the result of (i) the connection time *user related* (ii) the max energy uptake being the difference between arrival SOC and total battery size *EV related* and (iii) the charging speed of the charging point (Mies et al., 2017). Adding the kWh as primary feature in the clustering may skew the clustering towards transaction sizes rather than arrival time and skew towards EV types EV type (Plugin Hybrid EV (PHEV) versus large full Battery EVs (BEV)) rather than EV user types. The energy uptake is included in the session type distributions (Appendix E). In the description and analysis of the user types the energy uptake was used as a possible determinant for particular EV user types (see Section 4.2).

The charging location is an important feature of a charging sessions. In this research, the traveling distance in meters between two consecutive sessions (distance between sessions; DBS) was used as an indicator of the type of displacement. For example, a typical trip from a home location to a shopping location may be 7 km. The charging location (as GPS location in longitude and latitude) was not used as GPS locations may result large geographic clusters of sessions that may be related from a location perspective but not from a functional perspective.

For start time and connection duration, hours and minutes were chosen as level of detail. The exact date was left out and the day of the week was treated as secondary feature. The connection duration is known to contain a long tail of sessions with (extremely) long sessions (charging station hogging) up to 29 days (Wolbertus and van den Hoed, 2017). During clustering connection durations of more than 48 h (2 full days connection) (1.6% of sessions) were recalculated by taking the modulo of 48 h. Given that the longest duration under normal conditions would be a duration of a full weekend (48 h) we set the boundary to 48 h.

The Distance Between Sessions (DBS) was calculated using the distance as the crow flies between one charging session and its predecessor measured in meters. The DBS feature varies between 0 m and 153 km (which is approximately the largest distance between any set of CPs in the dataset). The distribution contains 61% of sessions occurs at the same location as the previous that have a value 0. The distribution of DBS values larger than zero has a median of 3.3 km and a mean of 11 km. In the cluster analysis a $\log_{10}$transform on the DBS in meters was applied to (i) avoid broad difficult to interpret Gaussian distributions as result of clustering due to sparsity of data at high distances (ii) accommodate to the human nonlinear perception of distance (e.g. the there is a larger perception of distance difference between 150 m and 1500 m than between 40.150 km and 41.500 km). As a consequence, during clustering the minimum distance is shifted from 0 to 10^0 (10) meters.

The Hours Between Sessions (HBS) was calculated by the difference in start time of a session and the end time of its predecessor in units of hours with two decimals. We assumed that the HBS relates to different activity patterns. For instance, a short HBS (<2 h) may point a new charging session directly after traveling whereas a long HBS (>7 days) may indicate that a user may charge at facilities not present in the data. Values of HBS larger than a full week (8% of sessions) were recalculated by taking the modulo of 168 h. The full week preserves the relation with the start connection time and the day of the week as supplementary variable.

3.3. Clustering of charging sessions

The four features of sessions types are shown in Fig. 1(a–d). Start times were normalized from 0 to 1, 1 meaning midnight (23:59 hrs), with clear morning and evening peaks. The connection duration was normalized from 0 to 48 h. It can be seen in Fig. 1(b) that the distribution contains four distinct peaks from short connections durations up to long connection durations larger than one day (>0.5). The HBS value was normalized from 0 to 196. The HBS feature in Fig. 1(c) contains less clear peaks. This figure is cut off at 0.35 (48 h) to improve clarity, since the tails tends to flatten out (8% of sessions). For the Distances Between Sessions (DBS) first the $\log_{10}$ taken on the DBS in meters and then the DBS was normalized from 0 to $\log_{10}152.99$ (km). In Fig. 1(d) normalized log-DBS is shown where values of 0 are left out (61% of sessions) to improve clarity of the distribution shapes.

In this research the Gaussian Mixture Model (GMM) was used for clustering. The GMM is a clustering method that performs clustering based on fitting multiple Gaussian (normal) distributions on a dataset such that the sum of these distributions fits the original shape of the data. As a result, each cluster is shaped as a multivariate normal distribution that describes the input features. We hypothesized that our dataset was drawn from a mixture of probability distributions, each from an independent subpopulation of charging sessions relating to types of behavior.

In GMM given a dataset with observations $X = (x_1, \cdots, x_n)$ with $Xin\mathbb{R}^d$ ($d = 4$ in our data) are assumed to be generated by a mixture of G components. Each component reflects a type of session in our dataset. The total density function is described as

$$f(x) = \sum_{k=1}^{G} \pi_k f_k(x_i|\theta_k) \tag{1}$$

In Eq. (1), the function f_k is the multivariate normal distribution with θ_k as parameters,π_k is weight of the component in the mix of the kth component ($\sum_{k=1}^{G} \pi_k$=1) and G is the number of mixture components (Scrucca et al., 2016). For our multivariate four dimensional data, the density function is given by Eq. (2) where $\theta_k = (\mu_k, \Sigma_k)$, $\mu_k$is the mean and $\Sigma_k$is the covariance matrix. Based on the covariance matrix the geometry (volume, shape and orientation) of each multivariate normal distribution is determined (see Table D2 GMM Model overview (adapted from Scrucca et al., 2016) in Appendix D).

$$f_k(x_i|\theta_k) = \phi(x_i|\mu_k, \Sigma_k) = 1/(2\pi)^{d/2}|\Sigma_k|^{-1/2}\exp\left\{-\frac{1}{2}(x_i - \mu_k)^T\Sigma_k^{-1}(x_i - \mu_k)\right\} \tag{2}$$

For a given number of session types, G, the mixture model parameters θ_k are unknown and must be estimated. This is in most cases performed by a two-step Expectation and Maximization (EM) procedure that maximizes the likelihood of the parameters (Fraley and Raftery, 2007). To do so we defined the describe that x_i comes from Gaussian k which we write as $z_{ik}= p(z_k=1|x_i)$. First, the EM algorithm makes an estimate of parameters π_k and θ_k from which the probability is evaluated that datapoint x_i belongs to component k using Bayes' Rule (Eq. (3) left-hand side). The expectation is then evaluated over Z having $\mathbb{E}[z_{ik}]$ which is $\sum_{k=1}^{G} z_k\gamma(z_{ik}) = \gamma(z_{ik})$.

$$p(z_k=1|x_i) = \frac{\pi_k f_k(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{G} \pi_j f_k(x_i|\mu_j, \Sigma_j)} = \gamma(z_{ik}) \tag{3}$$

Second, having found the Expectation, the Maximization step seeks to maximize the likelihood of the data given values of π_k and θ_k. To do so, the loglikelihood over our data X and latent variable Z is calculated using Eq. (4). The E-step and M-step are iterated until convergence is reached. The result of the EM is a set of probabilities $S = (p_{1,k=1}\cdots p_{n,k=G})$ for $(x_1, \cdots, x_n)$ from which the highest probability is assigned to be the component or cluster to which the observation belong to.

$$\ln p(X, Z|\theta_k^*) = \sum_{i=1}^{N}\sum_{k=1}^{G} z_{ik}[\ln\pi_k + \ln f_k(x_i|\mu_k, \Sigma_k)] \tag{4}$$

A proper strategy for GMM model selection is to run GMM on $G = 1. .G_{max}$and thereafter evaluate the best model performance by a value that balances the loglikelihood (Eq. (4), as described in (Fraley, 1998)) of the model with the number of parameters$M$ and the number of observations$n$. Two often used values are the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). We choose to use the BIC (Eq. (5)) over the AIC as the BIC penalizes the number of parameters more than AIC. This prevents overfitting of the model resulting in too many clusters. The number of independent parameters M is found given by the degrees of freedom provided in the covariance matrix (see Table D1 in Appendix D). The minimum BIC value over clustering from 1 to G_{max} with different covariance matrices was chosen as the best fitted model.

$$BIC = \ln(n)M - 2\ln p(X, Z|\theta, \pi) \tag{5}$$

Four different clustering models (see Table D1 in Appendix D) were tested on ($G = 1\cdots30$) clusters assuming either normal or noisy conditions (Fraley and Raftery, 2007). Both the VVV and EVV model did not converge under either condition. Therefore, Bayesian regularization as proposed by Fraley & Raftery (Fraley and Raftery, 2007) was executed resulting in convergence for the VVV model and significant improvement of the BIC, see Fig. 2(a). The VVV model did not converge for values larger than 13, except for 22 which appeared unstable. All results larger than 13 appeared to have limited stability (Adjusted Rand Index (ARI) values lower than 0.6 (Hubert and Arabie, 1985)). In our research we strive for a stable (measured in ARI) and interpretable clustering of session types. From cluster 10–13 we found marginal, yet important differences in clustering. We chose 13 clusters over 12 as the DayTime Medium duration sessions appeared better explainable.

A deeper analysis of the clustering similarities was performed using the Adjusted Rand Index (ARI) (Hubert and Arabie, 1985). The ARI (Eq. (6)) measures the similarity of two clusterings $U = (U_1, \cdots, U_r)$ and $V = (V_1, \cdots, V_s)$ of the same set of $X = (x_1, \cdots, x_n)$ given the expectation of change using a contingency table $\mathscr{M}$. In Eq. (6) is n_{ij} is the number elements that are both in U_i and V_j, $a_i = \sum n_{i1}. ..n_{is}$ is the row sum of the contingency table and $b_i = \sum n_{1i}. ..n_{ri}$ is the column sum of the contingency table. The ARI value varies between 0 for complete randomness and 1 for similar clusterings. In Fig. 2(b) the ARI values for each combination from $G = 7\cdots14$ is displayed. It can be seen that the clusterings show similarities and that G13 is a refinement of G12 (ARI 0.85).

$$\text{ARI} = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}} \tag{6}$$

All sessions were assigned to a partition out of $K = (k_1, \cdots, k_{13})$ based on ($\text{argmax}(p(z_k = 1)|x_n)$) Eq. (3). Yet, there may exist sessions of which the $\max(p(z_k = 1)|x_n)$ is still a low value due to anomalous properties of that session. For instance, a session starting at 02:00 with a 1-hour duration has a low likelihood for any of the typical session types. Every user type is expected to display some fraction of sessions with a low $\max(p(z_k = 1)|x_n)$. The objective of our study is to derive rules for behavior. We have therefore chosen to set apart sessions of random behavior to a dedicated *noise* cluster of which each user has a specific percentage in its portfolio.

We set a threshold on $\max(p(z_k = 1)|x_n)$ to reach an average probability $p(z_k = 1)|x_n$ of 0.95, which was found after removing sessions with a ($\max(p(z_k = 1)|x_n) < 0.7$) 7.7% of the sessions). Sessions with a low probability of belonging to any of the clusters were labeled as the noise cluster (k_{noise}).

The result of this is a robust and distinct clustering configuration containing 13 distinct clusters of charging sessions and that this is good basis for starting to understand the different EV user types. Analysis and textual descriptions of the session types can be found in Section 4.1.

3.4. Clustering of portfolios

Having developed a typology of charging sessions, the portfolio of sessions per user was made by taking the ratio of sessions per session type in the portfolio. Having s_{ik} as the number of sessions of user i in cluster k, then the ratio u_{ik} of sessions of user i in cluster k is given by

$$u_{ik} = \frac{s_{ik}}{\sum_{k=1}^{K} s_{ik}} \tag{7}$$

This results in a set of $U = (u_1, ..u_n)$ with $U in \mathbb{R}^d$ where $d = |K| = 13$. After removing the noise cluster (k_{noise}), users with less than 40 sessions ($\sum_{k=1}^{G} s_{ik}$)were removed in the second clustering. This led to an additional removal of 15% of users (the remaining 28,208 users represent 92% charging sessions).

To identify the EV user types we clustered the user portfolios. The values of u_{ik} cannot be considered as an independent and identically distributed random variable (i.i.d.) because $\sum_{k=1}^{G} u_{ik}$ by definition sums up to 1. Therefore, GMM clustering could not be used for the second clustering. Instead, our approach was to try four clustering techniques (K-means, Partitioning Around Medoid (PAM), hierarchical Ward and hierarchical Complete) with a Euclidian distance function (Eq. (8)) while varying the number of clusters from 1 to 15. This gives a total of 60 potential combinations of clustering method and value of K. The most appropriate clustering was chosen based on a selection of clustering quality indices.

$$d(i, j) = \sqrt{(u_{i1} - u_{j1})^2 + \cdots + (u_{i13} - u_{j13})^2} \tag{8}$$

Clustering quality challenges have been widely addresses in literature and a wide variety of quality indices have been proposed each with own perspective of clustering quality (Milligan, 1981; Ghazzali, 2014). We assessed each of these combinations by using 30 clustering quality indices (specified in Table D2 in Appendix D). For each clustering method we tested the 30 quality indices, each index indicated the value of K that provided the best clustering according to its criteria. This effectively meant each index *voted* for the most appropriate value of K for a given clustering method. This was then repeated for each clustering method resulting in 120 votes over all combinations. The optimal clustering configuration (method and number of clusters) was chosen based on (1) the fraction of indices that indicates K as the optimal clustering (2) relative differences between different clustering configurations based on ARI, (3) high contingency with known UseType labels (as shown in Table 1) based on ARI results and (4) mean ARI value for 10 iterations the same clustering configuration.

The index scores of (1) define feasible K values for a specific clustering configuration. The ARI comparison in (2) ensures a realistic clustering since, it is expected that realistic clusterings display similarities, whereas deviant clustering results are less similar. The comparison of the clustering results with known user types (3) sets a condition of the clustering quality against known UseType labels (e.g. car sharing) are expected to end up in in the same cluster. The mean ARI for 10 iterations of the same clustering configuration (4) ensures robustness of clustering results.

In Fig. 3(a) the number of quality indices that address K as optimal clustering is shown for each clustering method. In this figure PAM K= {9} has ths highest fraction of votes (25%). Alternative combinations are K= {6, 15} for Complete clustering, $K = \{6, 12\}$ for K-means, K= {6, 7, 9} for hierarchical Ward and $K = \{6, 8, 12\}$ for PAM. These clustering configurations were analyzed in the ARI comparison.

From Fig. 3(b) it can be seen that the ARI values are relatively low compared to Fig. 2(b), which is due to the fact that the different clustering methods result in different partitions of the data. It can be seen that hierarchical clusterings are substantially different than all other clustering results and show only similarity with other hierarchical clusterings. This similarity is in line with expectations, since hierarchical clustering of $k + 1$ is a sub clustering of$k$.

The bottom row of Fig. 3(b) contains the ARI values on the subset of the known UseTypes labels. While these ARI indices appear low, they are in fact reasonably accurate, since the UseType label contains users from multiple taxi or car sharing companies, where each company may prescribe different policies for the use of public charging infrastructure. The ARI values for hierarchical clustering are low, even negative, which points at a similarity comparable with near random partitioning of the data.

Based on these results we concluded that hierarchical clustering is not suitable for user type clustering. The PAM $K = \{9\}$ has the highest ARI value, followed by K-means $k = \{12\}$. Yet, the K-means clustering appeared to have limited clustering robustness due to convergence issues, whereas PAM has an ARI of 1 meaning that the same clustering results were retrieved at each run. Based on the clustering robustness and comparable results for ARI on UseType, the PAM with $K = \{9\}$ was chosen as the best option for clustering user types.

4. Results

4.1. Typology of charging sessions

From the Gaussian mixture Model 13 distinct session types were revealed. In order to derive meaning from the clustering a naming system was devised that balances the specifics of each session type while retaining conciseness, see Table 2. To end up with least complex names for each of the four features we defined a name label that result in the largest split for that specific feature. The labels were ordered according to their distinctiveness.

The start time itself appeared to have little effect on splitting the data, yet from exploring the activity patterns of the session we found that the distinction between daytime (DT) and overnight (ON) was the largest distinction between the types of sessions. The second largest distinction was based on distance, having either the Same location or Different Location as the former session (SL or DL respectively). The duration of the appeared less distinctive and required three categories to set a split short, medium and long duration (SD, MD, or LD respectively). The last feature, HBS, appeared limitedly distinctive and required a label for both the type of distribution wide or narrow (W,N) and a label for the value of the mean (Short, Medium and Long duration).

A graphical overview of all features can be found in Appendix E that also includes the distribution of the previous type of session and energy uptake (kWh). Any figures that contain the letter E (e.g. Fig. E1) refer to figures in Appendix E. The data of distributions (supplementary information, e.g. State of Charging) upon which the figures are made are available online.

Fig. 4 displays the cluster sizes of the 7 types of daytime charging sessions (4 short, 3 medium duration) and 6 types of overnight charging sessions. Two things can be observed, (1) the total number of daytime sessions and overnight sessions seem quite balanced and (2) there are three session types (DT-DL-MD-WL, DT-SL-MD-NM, ON-SL-LD-NS) which are not common (in terms of frequency), yet they are sufficiently distinct from other sessions that the GMM clusters in separate clusters.

In the following sections we consider three groups of session types. In the analysis the clusters were separated into three types of sessions (i) long daytime sessions (Section 4.1.1) and (ii) short daytime sessions (Section 4.1.2) and (iii) overnight sessions (Section 4.1.3). In Figs. 5–9 the density plots are shown for all features relevant to the session types. In these figures the lines are drawn based on the location feature, a dash is used for same location (SL) and a line for different location (DL).

4.1.1. Daytime charging sessions with medium duration

The most stereotypical daytime session during office hours is captured in the DayTime-SameLocation-MediumDuration-

NarrowMediumHBS (DT-SL-MD-NM) cluster (Fig. E1). Sessions of this type display sharp peaks in both start (8:00) (Fig. 5(a)) and end connection time (18:00) (Fig. 6(a)) and are typically present during weekdays (Fig. E1 (a)). The locations of these sessions are always the same as the previous location. The Hours Between Sessions (HBS) (Fig. 8(a)) displays a sharp peak around 14 h, which is exactly the time window between end connection time at 18:00 and start connection time at 8:00. The combination of these features indicate that this session type is closely related to working activities. This is further corroborated by the fact that the cluster include zero sessions on Monday (Fig. E1(a)), since cars parked over the weekend exceed this HBS value. While this session type has often been described in charging infrastructure literature as an important and common type of session, only a small fraction of sessions (1.4%) of this type was found in the whole dataset.

The DayTime-DifferentLocation-MediumDuration-WideLongHBS (DT-DL-MD-WL) session type (representing 2.6% of total sessions) has an arrival and departure pattern similar to DT-SL-MD-NM and a similar tendency for weekdays (Fig. E2), yet it differs on the DBS and HBS features (Figs. 8(a) and 9(a)). A closer look at the HBS (Figs. 8(a) and E2) shows three distinct peaks. The first peak around 2 h (indicating the previous session ended two hours earlier), suggests that the previous session was an overnight (ON) session. This is confirmed by the fact that 35% of cases these sessions are a successor of an overnight session (see Fig. E2). The second peak around 16 h indicates that the previous session ended the day before between 17:00 and 18:00. The third peak at 40 (16 + 24) hours indicates that the previous session was a daytime session ending between 17:00 and 18:00 two days prior. An analysis of the distance distributions for these peaks reveals that indeed the first peak involves large distances, while the second and third peak often involve charging at the same location (Fig. E3).

Another office hours charging session is DayTime-SameLocation-MediumDuration-WideLongHBS (DT-SL-MD-WL) (10.1%) (Fig. E4). This session type is specifically different on the HBS feature as it had a broad distribution with several peaks (see Fig. 8(a)). The highest peak of this session type is around 40 h (16 + 24). Looking at the distribution of sessions over the days of the week (Fig. 4) it can be seen that this session type peaks on a Monday. All the other peaks are located at 16 h + N-days*24 h (see Fig. E4). The connection time distribution is bimodal having 36% longer than 6 h and 64% shorter than 6 h (Fig. 7(a)). An analysis of start and end connection times for sessions shorter and longer than 6 h revealed that sessions longer than 6 h display sharp peaks at start and end of office hours, and 79% of the sessions with a connection duration shorter than 6 h take place within the boundaries of office hours (Fig. E13 and E14). This session may be considered the first daytime session (at the office) after N-days of not charging.

4.1.2. Daytime sessions with short duration

Four daytime sessions with short durations were found. In literature these types have been called opportunity charging (De Gennaro et al., 2014) as they are the result of an opportunity to charge rather than a necessity to charge (e.g. range anxiety). An estimation of the State of Charge (SOC) of the battery at the start of the charging session was made to check whether the battery of the EV was empty or not. For daytime sessions with short duration the estimated start SOC was between 53% and 63%, while this value is significantly lower for e.g. overnight sessions 32–43%. A battery level larger than 50% suggests that the necessity to charge for daytime short duration sessions is low. This is consistent with our initial hypothesis that these charging session types are related to parking activities.

The cluster which best represents this type of charging is DayTime-DifferentLocation-ShortDuration-NarrowShortHBS (DT-DL-SD-NS) (5% of sessions) (see Fig. E7). These sessions have a connection duration of 1–2 h and start anywhere between 8:00 and 21:00. An important aspect of this type of session is its relation to the previous session (see Fig. E7). These sessions typically occur within 2 h of the previous session (80% of the sessions within this type). The distance distribution of DT-DL-SD-NS displays a different shape compared to others (see Fig. 9) as the distribution does not peak at very short distances. The time between both sessions was shown to correspond to the driving time plus a limited amount of time for unplugging and plugging the vehicle. This was validated for sessions with an HBS lower than 2 h (80% of DT-DL-SD-NS), as the traveling speed and time approach HBS as the estimated time to travel the DBS.

The DayTime-DifferentLocation-ShortDuration-WideShortHBS (DT-DL-SD-WS) (8.3%) is a charging session that appears particularly related to car sharing activities (see user type 9 in Fig. 10). Contrary to other session types, this session cannot be directly related to a specific activity type. It has a short connection duration that peaks before 2 h connection time (Fig. 7(a)). Yet, it shows a multimodal pattern with modes around 5 (33% of sessions within this type), 21 (42% of sessions within this type) and 42 (23% of sessions within this type) hours on the HBS feature (Figs. 8(a) and E8). The HBS does not correlate with the DBS. An analysis of three modes in HBS (Figs. E9 and E10) revealed that the first HBS mode is evening sessions starting at around 18:00 and ending just before 20:00. The other two HBS modes are normally distributed around 13:00 with a standard deviation of 3 h.

The DayTime-DifferentLocation-ShortDuration-WideLongHBS (DT-DL-SD-WL) (8.3% of total) (Fig. E11) session is a short session at a different location than the previous session with a start and end time distribution that is quite similar to the previous session (DT-DL-SD-WS). The HBS significantly deviates from all other short types of short sessions. The HBS in the original data shows a pattern of hours between sessions at exactly 3 to 7 days (3–7 times 24 h) repeating for 4 weeks. This suggests that this session occurs with a specific regular pattern. An analysis at the user level revealed that these sessions indeed show a regular frequency of use. This is confirmed by Fig. E11(h) that shows that this is session type is followed by a session of the same type in 35% of cases. The long times between sessions indicate that there are sessions in between DT-DL-SD-WL not present in the data, for instance at the (semi) private home charging locations. An analysis on the geospatial properties of charging points, where this session type occurs most, reveals that these were all nearby Points Of Interest (POIs) that induce regular returns (e.g. sport clubs).

Finally, DayTime-SameLocation-ShortDuration-WideShortHBS (DT-SL-SD-WS) (13.2% of total) (Fig. E12), which differs from all other short sessions, since this session occurs at the same location as the previous session. The DT-SL-SD-WS is a session that returns to its previous location for a short time during daytime without overnight charging. The HBS feature displays three modes (i) between 1 and 7 h (45% sessions), (ii) around 10 h (22% sessions) and (iii) around 21 h (33% sessions). A deeper analysis of the start and end connection times of these modes was performed (see Figs. E13 and E14). The start time and end time of the first mode are normally distributed from 07:00 to 20:00 with a mean around noon for start times and a mean at around 16:00. A large percentage of sessions (35%) of this mode occurs during the weekend. The start connection time of the second mode is sharply (72% of sessions) distributed between 16:00–19:00 and the end connection time is sharply distributed 17:00 and 21:00 (66% of sessions). The last mode is distributed between 07:00 and 19:00 and occurs during weekdays. For each mode a manual analysis of charging points that contain a large fraction of DT-SL-SD-WS sessions was performed. The first and second mode typically occur at residential areas, which points at short residential charging for instance after or during a day of work. The third peak typically is located nearby POIs, which may point at a short visit at a POI after having visited this before.

4.1.3. Overnight charging sessions

There were six overnight charging session types found in the dataset of which OverNight-SameLocation-MediumDuration-NarrowMediumHBS (ON-SL-MD-NM) (8% of all sessions) is the most stereotypical. This session type typically relates to overnight charging (at home locations) during weekdays. It has a very specific start time, at around 18:00 and end time at around 8:00. This results in a very narrow connection distribution of around 14 h. The HBS peaks at 10 h which implies that the previous session ended at around 8:00. The previous session is in 99% of the sessions another overnight session (Fig. E15(h)) and 35% is of the same type. There is no distance between sessions in this session type meaning that the previous session was at the same location.

The OverNight-SameLocation-LongDuration-NarrowShortHBS (ON-SL-LD-NS) (2.2%) session type shows approximately the same start and end time properties but differs on the connection time. This session type is an overnight charging type that lasts typically for 40 h, which is an overnight session plus one day. The distribution of sessions over the weekdays shows that this session typically starts on Saturday, which means that it lasts until Monday morning the first day of the work week. This session type has been referred to in literature (Wolbertus and van den Hoed, 2017; Wolbertus et al., 2018a) as hogging of charging stations, in which commuters connect their EV on Friday or Saturday and do not disconnect until Monday morning.

The OverNight-SameLocation-MediumDuration-NarrowLongHBS (ON-SL-MD-NL) (12.4%) (Fig. E17) displays a long disconnection duration which distinguishes this session from the other overnight types. This session type has a pattern of start and end times similar to ON-SL-LD-NS. The HBS feature appears to skip a number of days 1 (44%), 2 (19%), 3 (11%), 4 (7%) and more (see Figs. 8 and E17). Interestingly, almost 30% of this session type follows a session of the same type (Fig. E17(h)), which indicates a regular pattern of successive sessions that skip several days.

The OverNight-SameLocation-MediumDuration-NarrowShortHBS (ON-SL-MD-NS) (14.4%) (Fig. E21) on the other hand has a specific short duration between two sessions peaking at less than 2 h (see Fig. 8(b)). The arrival time of this session is also later than the standard overnight session. Having a short time between this session and its predecessor at the same location may indicate that an EV user comes home, connects at 17:00, say leaves at 19:30 to go shopping, then returns at 21:30 for overnight charge. This session does follow upon DaytTime-SameLocation-ShortDuration-WideShortHBS (DT-SL-SD-WS) in 25% of the sessions (Fig. E21(h)).

Two overnight charging session types where the current charge station and previous charge station are at different locations are OverNight-DifferentLocaton-MediumDuration-NarrowShortHBS (ON-DL-MD-NS) (9%) (Fig. E21) and OverNight-DifferentLocation-MediumDuration-WideLongHBS (ON-DL-MD-WL) (4.9%) (Fig. E19). Both session types have a start connection time that is slightly skewed to later times compared to the most stereotypical overnight session at the same location (ON-SL-MD-NM) (see Fig. 5B). This may point at an early evening session at a different location than its predecessor. Both sessions have an identical distance distribution with significant peaks at distances less than 1500 (m). A reason for the displacement of an overnight charging session with a walking distance between the current session and previous may be the occupation of the previous station. A closer look at the data confirms that for distances less than 1000 m 60–70% alternative CPs to the current CP in the surroundings of 0 to 750 m were occupied at the time of arrival of the EV user. This means that in the majority of cases none of the CPs in the vicinity is available as an alternative to the EV user, whereas the occupation rate of stations for larger distances 5 km was significant lower. This session type may therefore be a proxy of competition between EV users for scarce CPs. For a more detailed analysis of CP scarcity we refer to (Glombek et al., 2018; Gorka et al., 2019; Helmus et al., 2019a, 2019b).

Another reason for displacement of the overnight session may be work-home or POI-home traveling. For ON-DL-MD-NS it was found that the first peak (less than 7 h) of the HBS feature typically relates to trips with large distance between sessions, whereas 70% of the sessions with HBS between 7 h and 15 h displayed a distance less than 1 km (see Fig. E22). The former is the last session of day after traveling with charging, whereas the latter relates to home-home charging with displacement due to occupancy. The same results were found for ON-DL-MD-WL with addition of the fact that users may skip N-days as with DT-DL-SD-WL.

In conclusion it can be said that we see three different main groups of session types each with a stereo typical session type present; (i) daytime charging sessions with medium duration (14% of all sessions) (ii) daytime sessions with short duration (35%) and (iii) overnight charging sessions (50%). Within the daytime medium-long sessions the DT-SL-D-WL type is most dominant (10%), while the DT-SL-MD-NM (1.4%) is most stereotypical. The daytime sessions with short duration have DT-SL-SD-WS as most dominant type (13.2%), which is not the most stereotypical session type (DT-DL-SD-NS). And for the overnight session types the most stereotypical session is ON-SL-MD-NM (8% of sessions), while a variant that arrives later after a short session ON-SL-MD-NS (14.4%) is most dominant. Derivatives from stereotypical types appear to relate to contextual factors such as displacement, weekend charging and CP scarcity. For all non-stereotypical session types, we found that subtypes based on the HBS revealed deeper insight in the context of a session.

4.2. Typology of EV users

Charging behavior in this research is seen as the collection of behavioral outcomes described in the session types of the previous section. The users with similar portfolios of charging sessions types were clustered, resulting in nine distinct clusters of user types each with characterizing distributions of session types.

Fig. 10 displays the mean percentage and standard deviation (error bar) of each session type (horizontal) per user type (vertical). Several aspects of the portfolios stand out. First, none of the user types have a single type of charging session that explains more than 60% of its behavior. This means that stereotypes of behavior (the typical office charger, resident or visitor) are significantly more complicated than found in literature. However, most user types do exhibit a dominant (or most common) session type, but these are complemented by a range of other typical charging types. The data confirms our initial hypothesis that charging behavior of individual EV drivers displays a combination of different charging activities, that may be explained by for instance weekend-week days, but also mirror different mobility patterns (going to work, home or sports club). From Fig. 10, we can distinguish user types in daytime-, overnight- and non-typical user types.

Second, the office charging stereotype appears less dominant in the portfolio than the short sessions or overnight sessions. For instance, user types dominant on office hours rarely show overnight sessions, whereas overnight user types do include daytime sessions in their portfolio.

Third, user types that have frequent overnight sessions in their portfolio can be categorised into two different groups according to the number of overnight sessions at the same location as previous or a different location; (i) a group that has a portfolio of sessions dominated by overnight sessions at the same location as the previous location (ii) a group that has a portfolio of sessions dominated by overnight sessions at different locations as the previous location.

Table 3 provides a summary of the nine clusters and includes how car sharing and taxis are distributed. This table shows that the PAM has correctly clustered special (identifiable) groups of users (car sharing) to the same user type (see user type 9, 6 and 7). Where car sharing EVs are largely linked to one particular cluster, electric taxis are more distributed among several clusters. This may be related to the fact that taxi companies may have different policies regarding public charging. In Table 3 several aggregation properties of user types can be found. Also, features of user types not directly derived from the session typology (e.g. weekly charging sessions) appear to differ significantly. Interestingly, while the estimated battery size tends to have a weak correlation with behavioral features, it has a stronger correlation with user types. In Fig. E1 in the Appendix E the mean percentage of Noise per user type is shown.

4.2.1. Daytime user types

In total three user types were found that are dominant on daytime charging (1, 5, 8). User type 5 is most related to the office charging stereotype. This user type has about 1.7 charging sessions per week of which more than 65% are during office hours and around 35% is typically opportunity charging. Only 4% of their sessions are overnight charging sessions. This user type uses the least number of charging points (1.7; see Table 3). We assume that this user type has access to (semi-) private home charging since, the estimated battery size and mean energy uptake (see Table 3) deviate, while the connection duration allows for full charging.

User type 1 is also a daytime user with an average of 1.42 sessions per week particularly at weekdays, uses twice as much CPs as user type 5. This is the result of a different portfolio of session types. This user type has 40% of its sessions strictly during office hours and about 40% of its sessions outside office hours, yet still during daytime. The 20% of sessions is during the weekend of which only 0.1% is overnight charging.

User type 8 is a daytime user with 60% of its sessions during office hours and 23% of its sessions at the same day outside office hours. Contrary to the previous types, this user type is dominant on short session instead of medium. The locations of the dominant CPs of these user types are typically in work areas (Table D1). Connection duration is 3.5 h (median) and the user typically returns to the fist location after several hours, which suggests this user type to have a job that requires traveling (sales managers).

4.2.2. Overnight charging user types

From the clustering, four overnight charging user types were found (6,2,7). User type 6 is significant in the total population (21%) and considered most related to the stereotype of a residential user. This user type has a high frequency of charging of 4.60 times per week and has on average 75% of overnight charging sessions in its portfolio. The other 25% are short session during daytime of which 30% is in the weekend.

Closely related to this user type is number 2, which can be considered as an infrequent residential charger as this user type only has 1.8 sessions per week. This infrequent charging pattern is caused by the dominance of the ON-SL-MD-NL session type with the long times between sessions feature. On average 72% of its sessions are overnight charging and 28% is short duration charging. Table 3 shows that average battery size for user type 6 is significantly larger than that of user type 2 (26 versus 15kWh) which likely explains differences in charging frequency.

User type 7 is has the same pattern of frequency as user type 2. Yet, user type 7 has a large fraction of sessions that occur at a different location than the previous session, which is not the case for user type 2. An analysis of overnight sessions that have different locations then the location of the previous session revealed that the distance is typically less than 1000 m, which may be regarded as walking distance. Moreover, as mentioned in the description of session type ON-DL-MD-SS and ON-DL-MD-WL, these sessions typically occur when the previous CP is occupied. A closer look at the locations of CPs of this user type revealed that this user type is typically present in areas with high EV uptake and charging infrastructure maturity, such as the Amsterdam South region. The distance between CPs in these areas is often within walking distance, which points at the existence of alternatives nearby. Nevertheless, user type 7 also shows significantly larger distances between sessions than other user types.

4.2.3. Non-typical user types

Three user types do not fit with the traditional stereotypes described in literature. First, user type 3 is typically an all-round user with a mixture of office and residential charging. This user type has on average 23% of its sessions during office hours, 47% of nighttime charging and 30% of opportunity charging. Contrary to what may be expected by the wide portfolio of behavior, is that this user type has on average 2 sessions per week with a energy uptake of half the battery size. We therefore assume this user charges at (semi-) private or fast charging locations as well.

User type 4 is typically related to frequent visiting of locations nearby public charging infrastructure. This user type has close to one weekly charging session of less than 4 h duration. The sessions are uniformly distributed over the days of the week. The large average number of charging stations used (Table 3) suggests multiple destinations (e.g. shopping, sports, health).

Lastly, user type 9 can be considered as a random user type with noisy behavior as this user type has little to no sessions at the same location as the previous session. The start and end times of this user type are multimodally distributed over the day without significant tendency towards a specific hour. All the known car sharing EVs are in clustered with this user type in this cluster (Table 3). Yet, given the size of this cluster and the total number of car sharing EVs in the Netherlands, it is likely be that this cluster contains shared EVs of other providers (public or even company pool cars) currently unknown on the dataset.

4.3. Analysis of behavioral development over time

The user type clustering allows to evaluate any changes in composition over time, for instance due to technological development of EVs (i.e. battery size) and market adoption (number of EV users and number of charging points). Both the typology of charging sessions and the users can be analyzed as function of time. During the timespan of the data three developments have taken place in the Dutch context (see also Fig. E1): (i) a large uptake of BEV vehicles (ii) a decrease (sum import and export) of PHEVs and (iii) an increase of public charging infrastructure at a slower pace than EV uptake. From this it is expected that (i) user types related to PHEVs decrease in the population, while (ii) users related to BEV increase in the population and (iii) charging point scarcity may increase due to a decreasing ratio of CPs over EVs for EV users of public charging infrastructure.

In Fig. 11 the user type population development over time is shown. It shows significant increase and decrease of user types. Particularly user type 7, 3, 2 display growth, while user type 1, 4, 5 show significant decrease, the others (6, 8 and 9) remain stable. The correlation between user groups and monthly BEV uptake user types 2, 3, 7 R^2 is 0.97, 0.88 and 0.79 which is significantly larger than all other user types (Fig. E2). This suggests that the uptake of BEV in Fig. E1 is related to these user types. Moreover, user types 2, 3, 7 also show larger increase in mean battery size of newly added users over time (from 22 to 38 kWh), while the other user types only show an increase from 13 to 20 kWh. From extrapolation of current results, we expect that (i) future uptake BEV will lead to changes in division of user type (user type 2, 3). The increase of user type 7 in the population points suggests that charging point scarcity is also increasing in certain areas.

While we see that the composition of user types develops over time (Fig. 11), we do expect that the types of behavior (session typology) remain stable. We expect that future developments will mostly affect the distribution of rules for behavior rather than the rules of behavior itself. The implications of shift towards BEV (user type 2, 3) may lead to improved efficiency of charging points due to less frequent charging sessions with larger energy uptake. This is in line with earlier research on transitions from PHEV to large battery BEV (Helmus et al., 2018a, 2018b).

5. Conclusion

Current literature lacks a clear set of rules to describe when, where and how long different types of EV users will charge. In this paper we have developed a typology of EV users, and their underlying behaviors as charging sessions based upon 4.9 million charging transactions from more than 27,000 users. The typology of charging sessions based upon a stable Gaussian Mixture Model revealed 13 distinct types of charging sessions based on the start connection time, duration, hours between sessions and distance between sessions. The current typology of charging sessions contains several session types that still display multi-modal distributions of features (for instance hours between sessions). A session type must be seen in the context of the whole portfolio of a user's charging sessions. A potential solution may have been to subdivide session types until each session type is unimodal, yet this comes at a price of the BIC value and clustering robustness. The user typology consists of 9 user types with a distinct different portfolio of session types and is based upon a stable (ARI = 1) Partitioning Around Medoids.

The session typology displayed three categories of session types (i) daytime session during office hours, (ii) daytime sessions with short connection duration and (iii) overnight sessions.

Based on the high state of charge at of the EV at the start of the session we found that daytime sessions with short connection duration did not shot low SOC at arrival. This would suggest that these sessions were driven by opportunity rather than necessity. This is consistent with our initial hypothesis that charging behavior is driven by parking behavior rather than range anxiety. For each session type a simplistic stereotypical session type was found though in limited numbers, while more complex variants appeared larger in larger numbers. This confirms our hypothesis that rules for charging behavior are more complex than current simplistic types of charging found in literature (e.g. office charging, opportunity charging). From this we suggest that policy makers should take non typical charging behavior into consideration when working on charging infrastructure rollout strategies.

We considered each user type to have a typical set (portfolio) of rules of which each rule results in typical types of charging behavior (charging sessions). Daytime (office hours) dominant user types (5, 1, 8), nighttime dominant user types (6, 2, 7) and non-typical user types (3, 4, 9) were found. The user typology has demonstrated that none of the user types display solely stereotypical behavior – the range of behaviors is more varied and more subtle. This emphasizes the need for improvement of charging behavior models, particularly sets of rules for user behavior in computational models.

Three user types (2, 3 and 7) show significant peaks at larger battery sizes, while all others have identical battery size distributions. Two out of these (2, 7) (large battery) user types have overnight charging as dominant type (with long times between sessions) combined with short daytime sessions.

We also found that the scarcity of CPs affects the behavioral portfolio (e.g. in user type 7). This suggests that searching for alternative stations in case of occupation is an important driver of behavior and should be explicitly captured in any model. This is especially important now as the effects of scarcity is likely to increase in the near future as public charging infrastructure in metropolitan areas matures and users owning large battery EVs (e.g. user type 2, 3, 7) start to increase (see Fig. 11). It is to be expected that behavioral shifts will be observed, and it is important that models are able to capture this transition in behavior.

Finally, we found that types of users with unregular charging patterns, either infrequent charging (user type 4) or frequent charging at many locations (e.g. car sharing) (user type 9) take up about 15% of the total population. These types of users have specific behavioral properties and as such require attention in simulation models. It may be relevant to simulate population compositions related to interventions. For example, adding a car sharing scheme (user type 9) to the system or increasing city attractiveness for EV visitors (e.g. user type 4) by providing public charging possibilities.

6. Limitations and future research

The current research focused on AC charging as data was provided on public AC charging infrastructure. In practice, charging infrastructure in metropolitan areas consists of a variation of charging points from public to private and from AC slow to DC fast charging. As such, generalizing results is more likely in countries where public charging is prominent. While charging behavior in this research is limited to public AC charging, it may well be that the EV users in the data use DC fast and or private AC charging as well. Future research on charging behavior typology may expand to all other forms of charging as well. Yet, the difficulty of combining AC and DC charging lies in the fact that charging cards, used as unique identifier of an EV user, for AC and DC charging come from different providers. This causes AC and DC to be identified as two different users, while they may in practice be the same user.

The dataset at hand contains a wide geographic area with different charging point densities and infrastructure maturities. This makes the results applicable for countries with strong public charging infra utilization; possibly with high density of public chargers and similar tariff structures. Tariffs and business models in may affect charging behavior connection durations. Future work may therefore also include an analysis of the applicability of charging behaviors in different contexts than the Netherlands.

In this research we deliberately treat the energy uptake and estimated battery size as supplementary aspects of behavior. In the typology, we found a week relation between user types and battery sizes. However, developments in battery size and the transition from PHEV to FEV may affect behavior, particularly the hours between sessions (Vermeulen et al., 2019). Future research may explore behavioral changes over time between EV users from early stages of EV uptake in 2014 until recent influenced by technological developments.

Also, charging infrastructure in cities is maturing over time as well. This may cause network formation, effects of competition and collaboration between EV users (Helmus et al., 2019a, 2019b). In this research we have found several session types and user types that relate to CP scarcity. Future work may not only include effects of EV maturity, but also treat a city or area as a portfolio of behaviors. This allows for comparing cities or areas in time and space. In addition, we plan to work on a typology of charging point types based on the portfolio of session types developed in this paper. This typology may include grid requirements, CP utilization and economics.

CrediT authorship contribution statement

Jurjen R. Helmus: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. **Michael H. Lees:** Supervision, Writing - review & editing. **Robert van den Hoed:** Supervision, Writing - review & editing, Funding acquisition.

Acknowledgements

This work is part of the doctoral grant for teachers with project number 023.009.011, which is (partly) financed by the Netherlands Organization for Scientific Research (NWO). The data is provided by the G4 cities (Amsterdam, Rotterdam, Utrecht, The Hague and the Metropolitan region of Amsterdam) of the Netherlands. Finally, this research is part of the IDOLAAD project funded by Stitching Innovatie Alliantie (SIA).

Appendix A. Figures on charging data

Fig. A1.

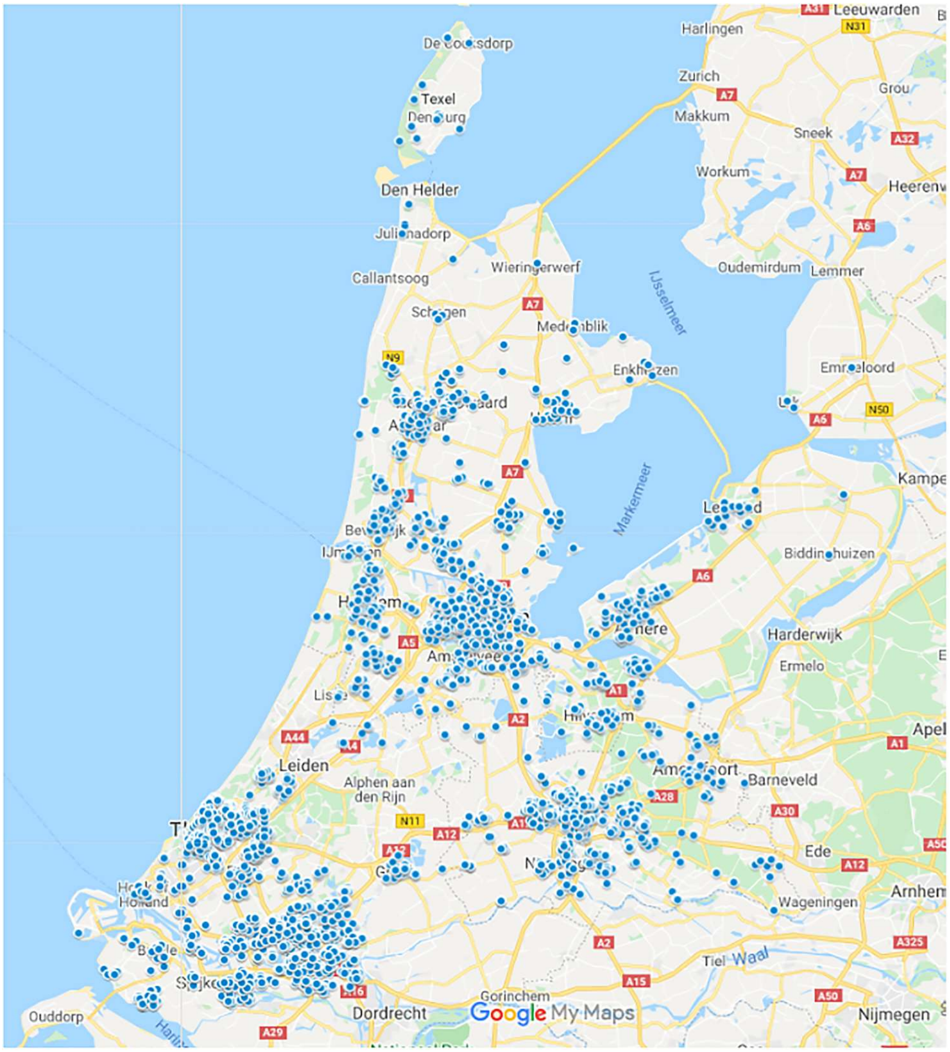


Fig. A1. Map of charging points in the dataset.

Appendix B. Figures on charging data

See Figs. B1–B7.

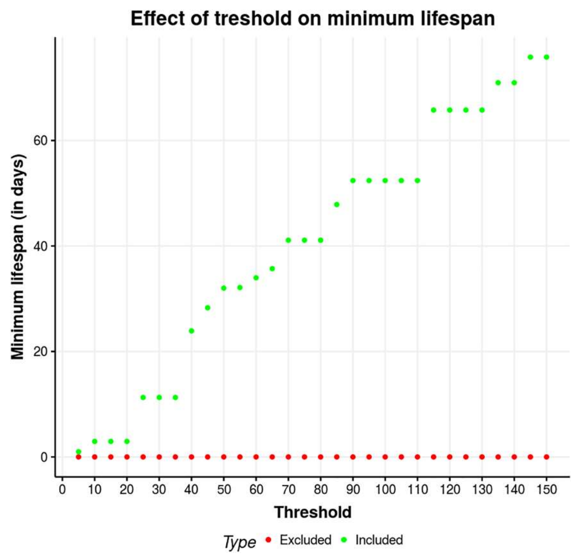


Fig. B1. Effect of applying a threshold on number of sessions on the minimum lifespan of an EV user in the population.

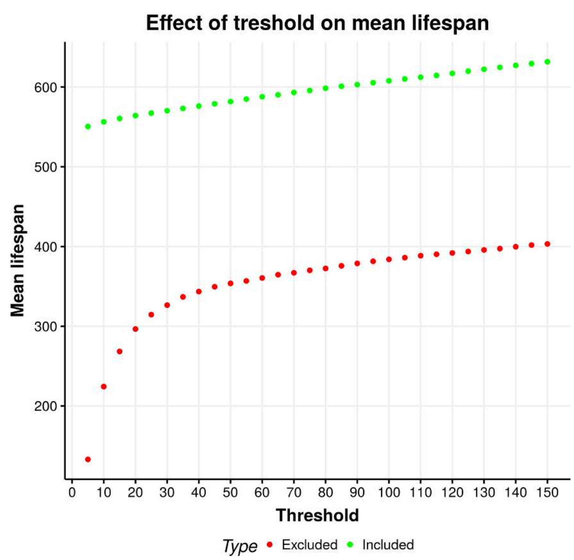


Fig. B2. Effect of applying a threshold on number of sessions on the mean lifespan of an EV user in the population.

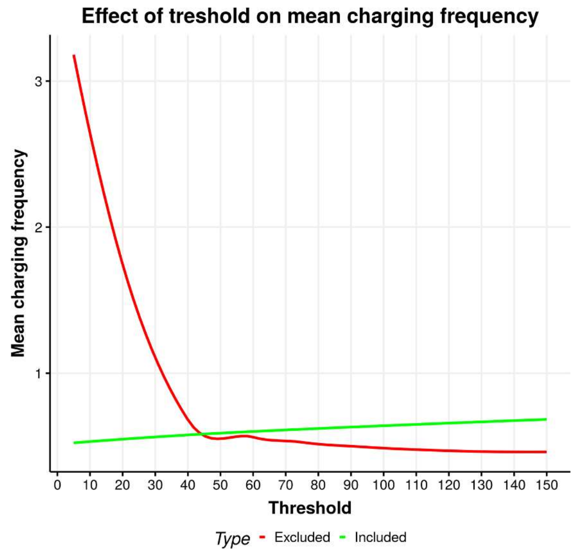


Fig. B3. Effect of applying a threshold on the mean charging frequency of EV users in both populations.

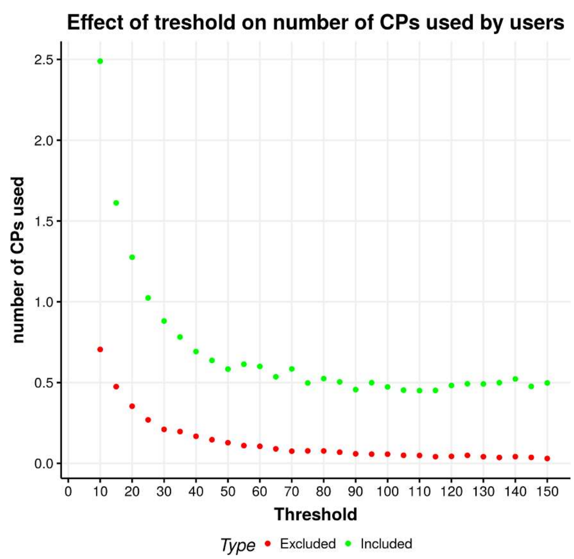


Fig. B4. Effect of applying a threshold on mean number of CPs used by EV users in both populations.

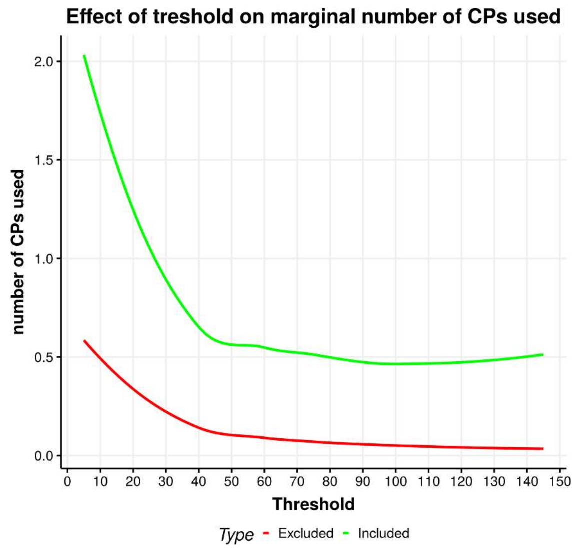


Fig. B5. . Marginal effect of applying a threshold on mean number of CPs used by EV users in both populations.

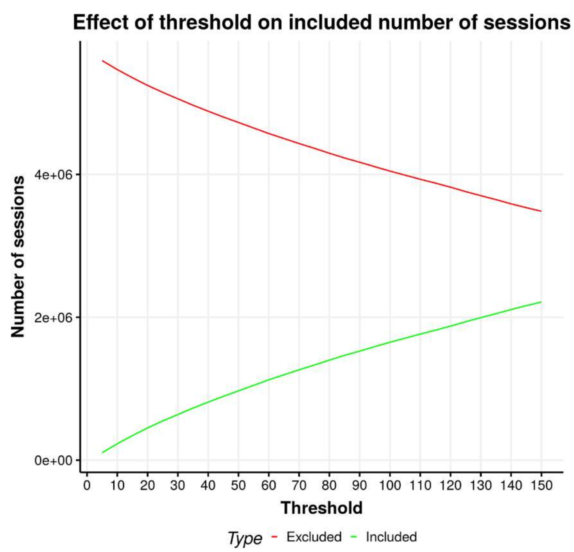


Fig. B6. Effect of applying a threshold on the number of sessions.

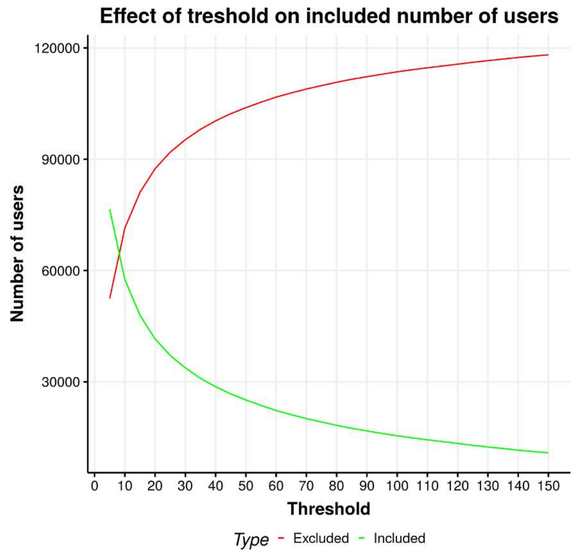


Fig. B7. Effect of applying a threshold on the number of users.

Appendix C. Literature on charging behavior choice modelling and data driven research on charging behavior

Research on charging behavior choice modelling

Early work on charging behavior mainly involves research focused on understanding psychological aspects of charging decisions (Caroll et al., 2010; Franke and Krems, 2013). Successive research has been developed that models dependent variables such as location choice (Nunes et al., 2015), charging routing and detours (Xu et al., 2020), charging interval or frequency (Daina et al., 2017a; Kim et al., 2017), charging mode (fast or slow) (Xu et al., 2017), or whether or not to charge on public charging infrastructure (Zheng et al., 2015; Wen et al., 2016; Ge et al., 2018).

. Other research has acknowledged the heterogeneity in user types and contextual situations using more complex models (Sun et al., 2015). The result of this research is typically a model that calculates the probability for a set of (charging) choices (dependent variables) based on independent variables.

Typical independent variables taken into account include State of Charge (SOC) of batteries (Zoepf et al., 2013), dwelling time or distance (Yang et al., 2016; Ge et al., 2018) and costs of transactions (Daina et al., 2017a; Ge et al., 2018). Beside these variables also socio-demographic variables have been used such as age, gender income and experience with EV charging. An overview of variables that have been used to describe and explain charging behavior is presented in (Li et al., 2017) and (Pan et al., 2019).

From this stream of research, we learned that arrival times, time between two charging sessions and type of charging locations are distinct features of charging behavior. A limiting factor of this research is that the models are often based on stated preferences under artificial experimental conditions, sometimes with non EV-owners.

Franke and Krems (2013) followed by others (Daina et al., 2015; Kim et al., 2017) bridged the gap between choice models and user charging data with an adaptive control model from the perspective of a single charging decision of a user. While the latest models of this type of research have become more sophisticated, the outcomes are limited to a discrete set of choices. As such we begin to understand more about individual charging behavior of EV drivers, there is still room for improvement on development of a typology for particular user types

Data driven research on charging behavior

In another line of literature charging behavior is modeled from charging data, rather than that a choice model is fitted on charging data. Charging transactions data are used to estimate rules for charging decisions. Charging behavior is modelled in terms of probability densities for diverse variables on a population level or session level. Due to the current state of EV adoption and limited examples of large scale roll out of EV infrastructure, the amount of data in such studies is still limited. Most data driven research on charging behavior uses either a relatively small real world dataset (Khoo et al., 2014; Kara et al., 2015; Morrissey et al., 2016; Xydas et al., 2016a, 2016b) or a dataset from a semi controlled environment (Yoo and Park, 2015), synthetic data (Guo et al., 2012; Momtazpour et al., 2012) or charging behavior data derived from GPS datasets of non EV users (Ashtari et al., 2012; De Gennaro et al., 2015; Gottwalt et al., 2015; Paffumi et al., 2015; Yang et al., 2016). An example of an exceptions on this is Wolbertus (Wolbertus et al., 2018a) who studied over 2.6 million charging transactions of 60,000 EV users to analyze factors that influence connection times.

A typical topic involves optimizing charging infrastructure dimensions or locations (Ashtari et al., 2012; De Gennaro et al., 2015; Gottwalt et al., 2015; Kara et al., 2015; Paffumi et al., 2015; Yang et al., 2016), which requires modelling charging behavior and calculating charging profiles as uptake of energy per unit of time. While charging profiles indeed help to set the grid requirements for charging infrastructure, they are dependent on user behavior, battery size and charging points properties. Connection profiles on the other hand are largely driven by traveling (and hence parking) behavior (Wolbertus et al., 2018a). In this research we are interested in the relation between charging behavior and activity patterns as it determines infrastructure usage (Kontou et al., 2019). We assume that recharging decision may be driven to activities rather than necessity.

Several cases of data driven research on charging behavior are still in an descriptive phase (Khoo et al., 2014; Morrissey et al., 2016; Triebke et al., 2016). Both Khoo and Triebke analyzed EV driver charging behavior aggregated over a population, in terms of time of day, duration, time between recharging events and the charging volume (kWh) to recharge the vehicle battery (Khoo et al., 2014; Triebke et al., 2016). These studies show how different locations (rural, business, residential) result in different charging profiles in terms of arrival times, duration and charging volume. Form this research it is also found that weekday and weekend profiles significantly differ. Morrisey et al. (2016) not only extended these variables with location properties, they also developed a models that fit the connection, arrival and duration distribution on a population level.

Wolbertus (Wolbertus et al., 2018a) provides a model that predicts the connection duration based on the arrival time and day, charging card owner type, city characteristics, type of charger (fast or level 2) and the price. From this research it appears that the moment of arrival and the type of charger (AC/DC) are the strongest determinants for predicting connection duration. Also modelling charging behavior in urban contexts proved to be difficult due to the variety of usage in parking and refueling. This work embraces the idea of distinguishing different EV user types based on their charging behavior at the same public charging infrastructure.

In sum, data driven literature on charging behavior has developed from statistical exploration towards modelling of specific properties of behavior. Despite statistical models have led to better understanding of distributions and probabilities of charging behavior, applying data analyzing to develop a thorough classification of EV user types is missing.

Appendix D. Additional information on clustering methods

See Fig. D1 and Tables D1 and D2.

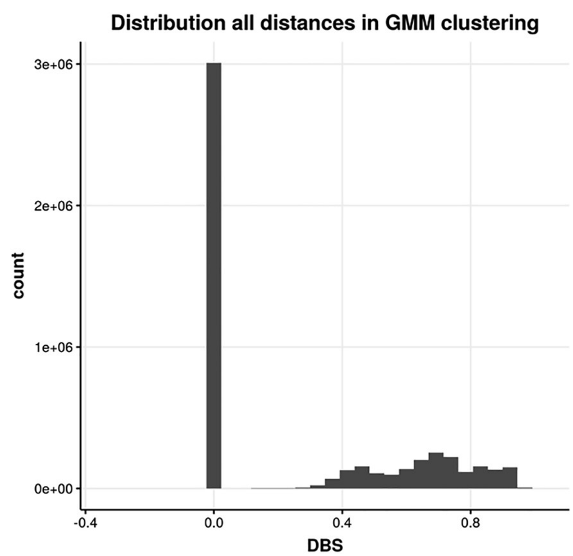


Fig. D1. Distribution of all distances in GMM clustering.

Table D1
GMM Model overview (adapted from Scrucca et al., 2016).

Clustering	Covariance matrix	Distribution	Volume	Shape	Orientation	Convergence
VEV	$\lambda_k D_k A D_k^T$	Ellipsoidal	Variable	Equal	Variable	No
VVV	$\lambda_k D_k A_k D_k^T$	Ellipsoidal	Variable	Variable	Variable	Yes, after Bayesian regularization
VEI	$\lambda_k A$	Spherical	Variable	Equal		Yes
EEV	$\lambda D_k A D_k^T$	Ellipsoidal	Equal	Variable	Variable	No

Table D2
Clustering Quality indices used in our research, ().
adapted from Ghazzali, 2014

Clustering quality index
Calinski and Harabasz (CH) index (Calinski
Duda index
Pseudot2 index
C-Index was
Gamma index
Beale index
Cubic Clustering Criterion (CCC) i
point-biserial correlation
Gplus index
Davies and Bouldin index
Frey index
Hartigan index
Tau index
Ratkowsky index
Scott index
Marriot index
Ball index
Trcovw index
Tracew index
Friedman index
McClain and Rao index
Rubin index
Krzanowski and Lai index
Silhouette index
estimated Gap statistic proposed
Dindex
Dunn index
Hubert's Γ statistic
SD validity index
SDbw validity index definition

Appendix E. Detailed information on session types

See Figs. E1–E23.

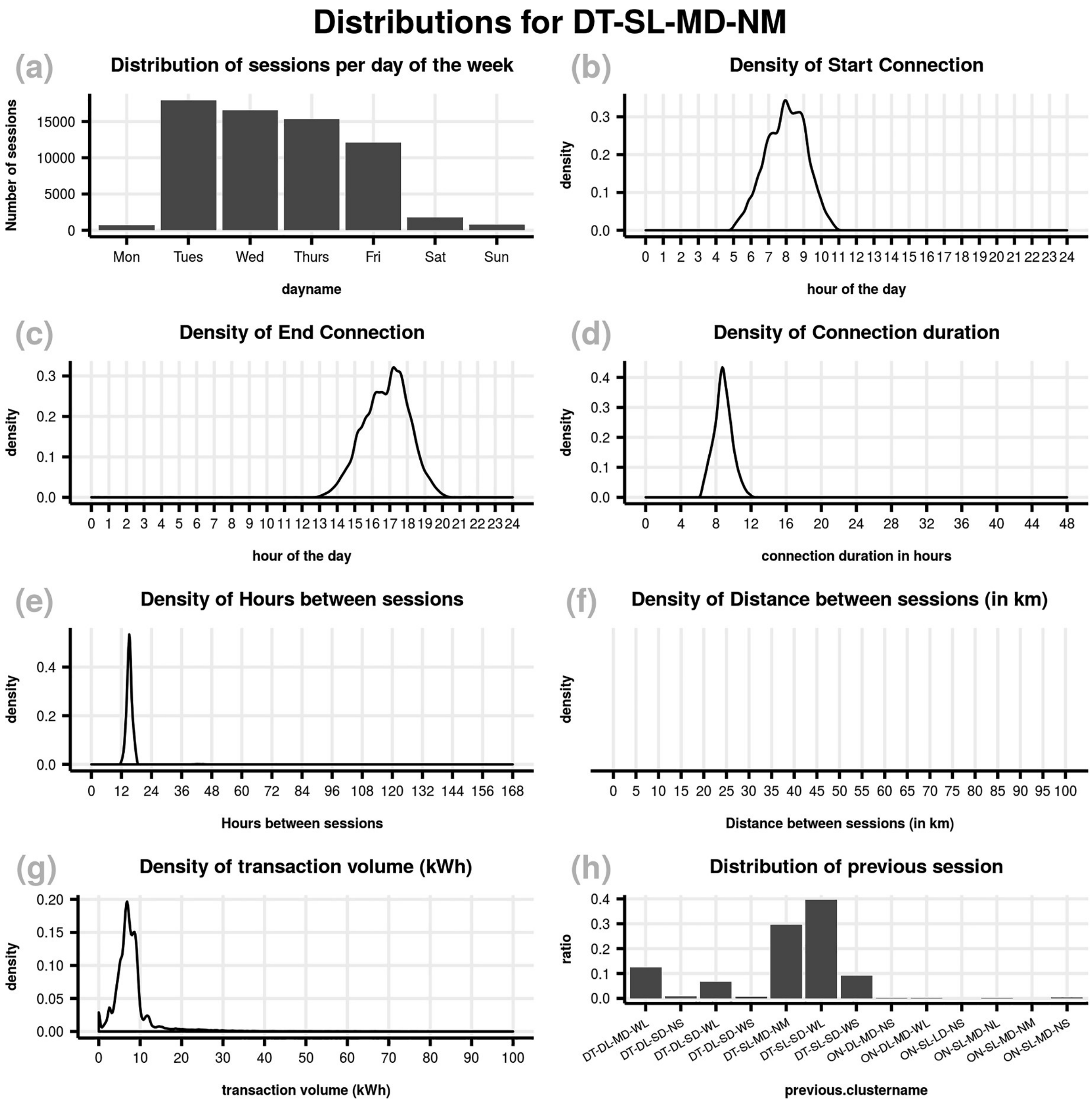


Fig. E1. Clustering Distributions for DT-SL-MD-NM.

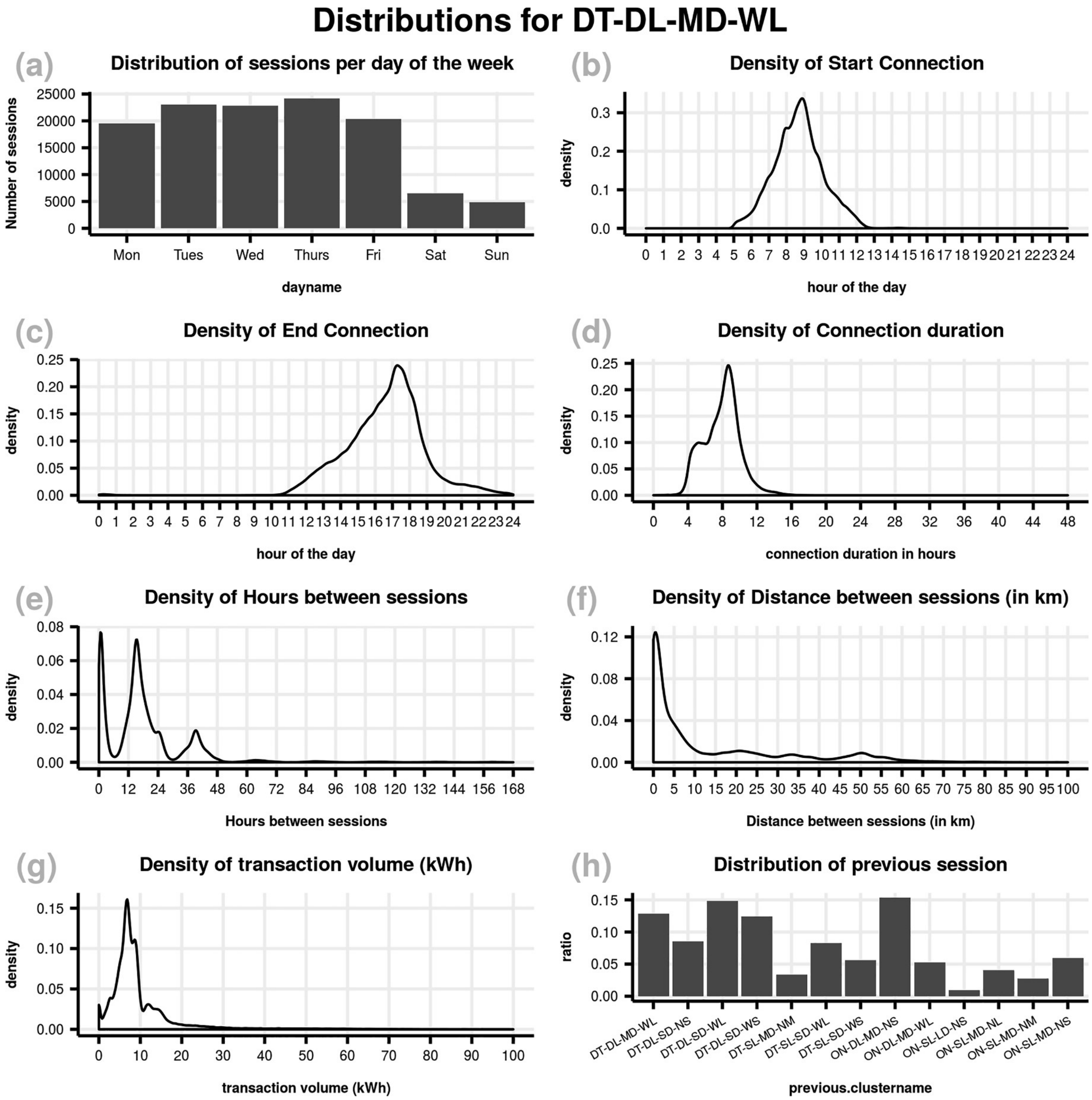


Fig. E2. Distributions for DT-DL-MD-WL.

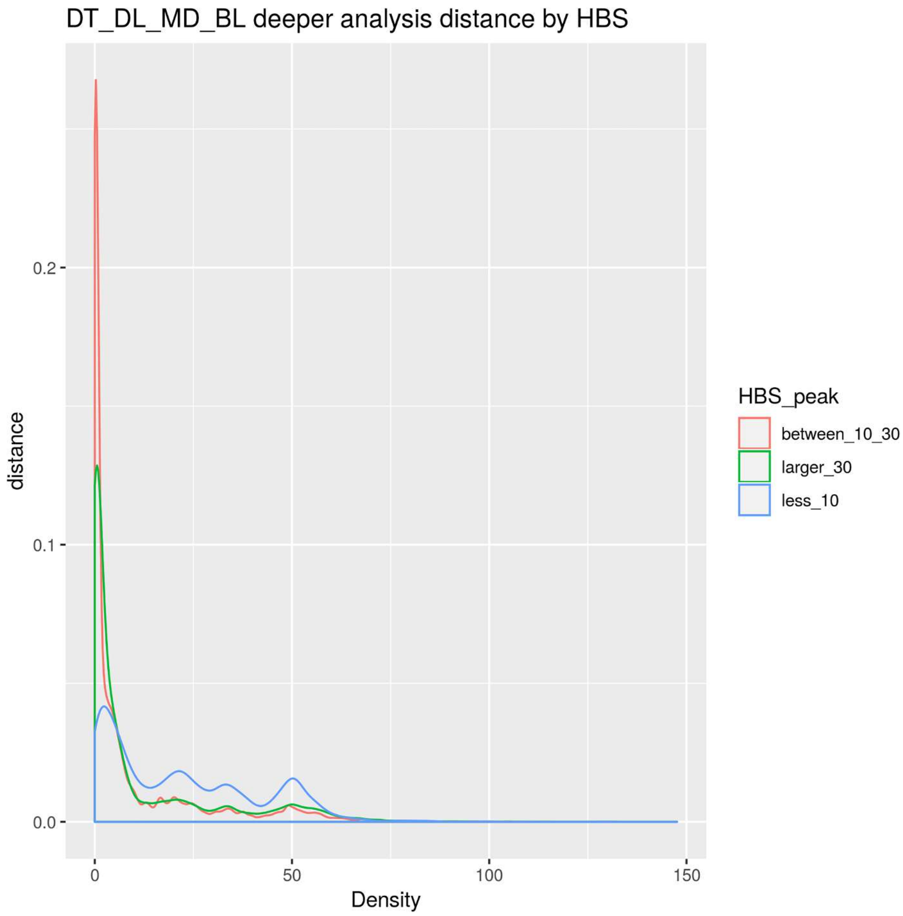


Fig. E3. Deeper analysis of distance between sessions by sub distribution of Hours Between Sessions for DT-DL-MD-WL.

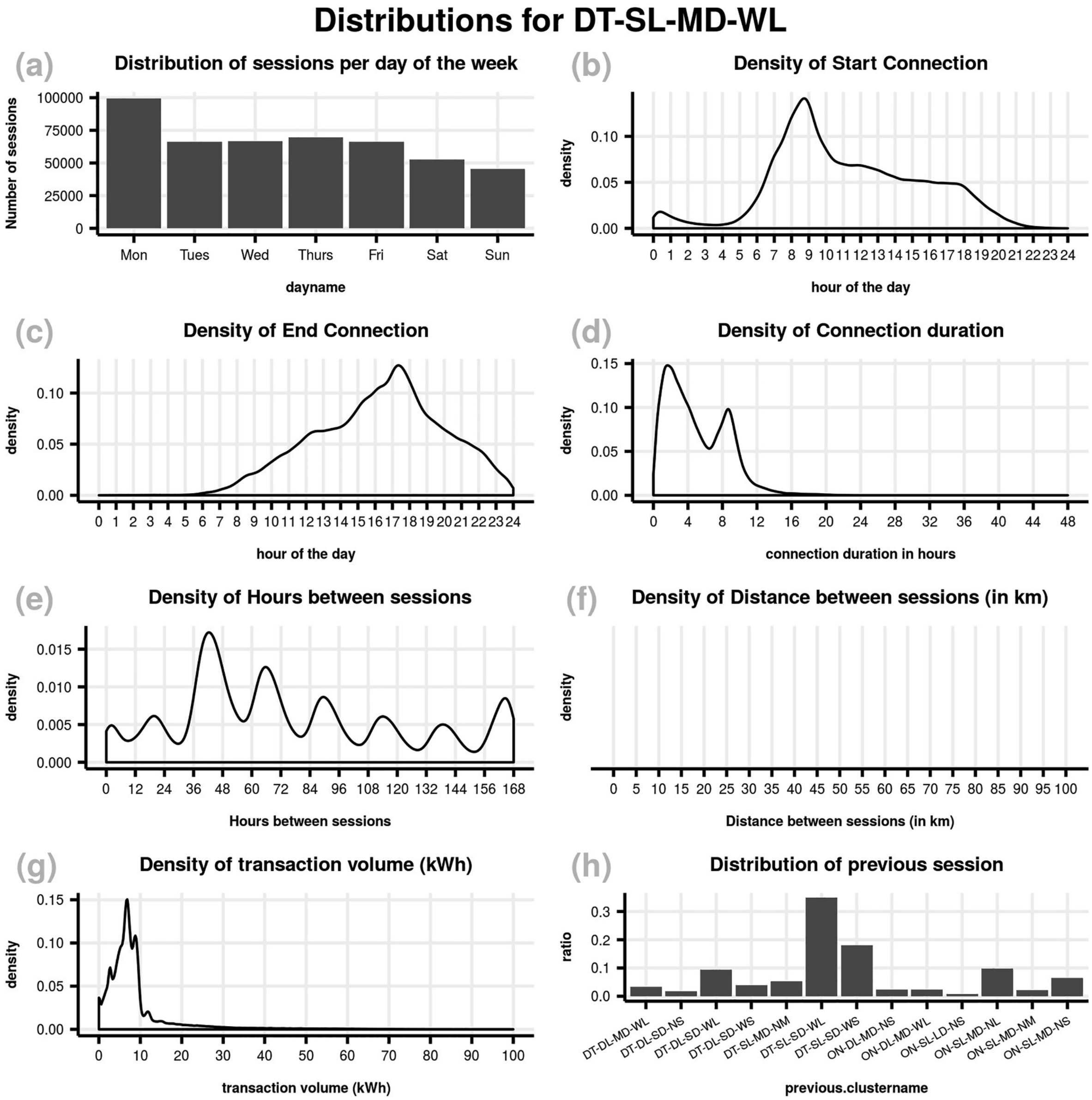


Fig. E4. Distributions for DT-SL-MD-BL.

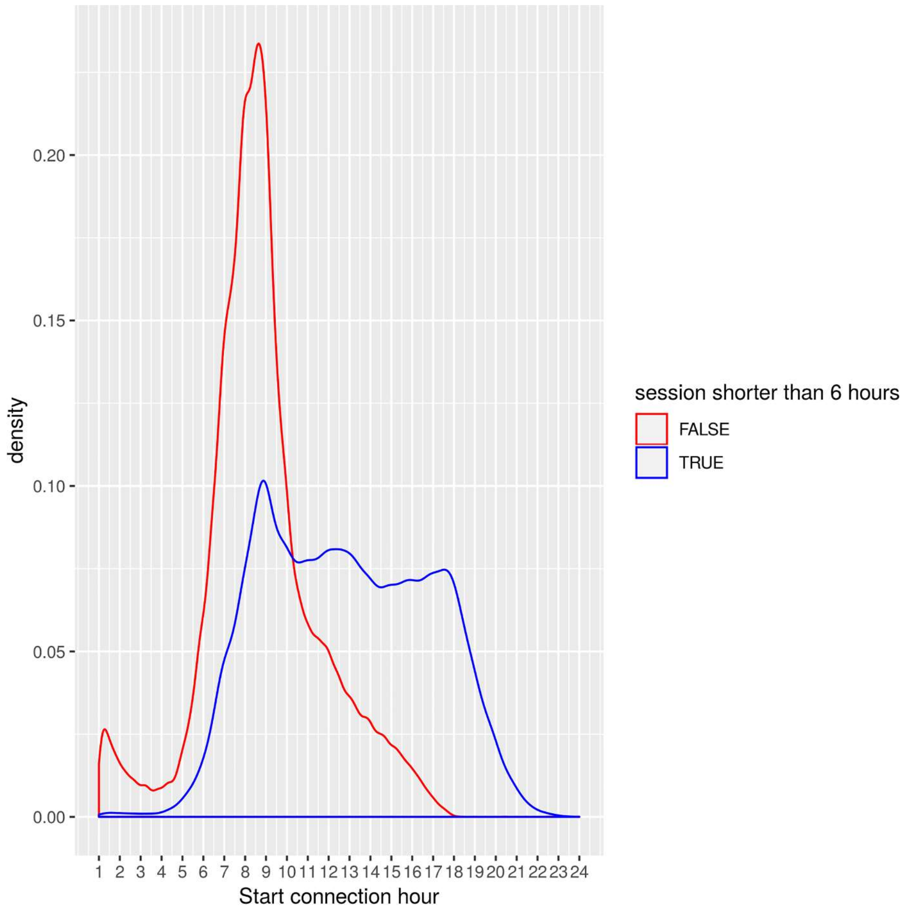


Fig. E5. Start connection distribution for sessions shorter and longer than 6 h for DT_SL_SD_WL.

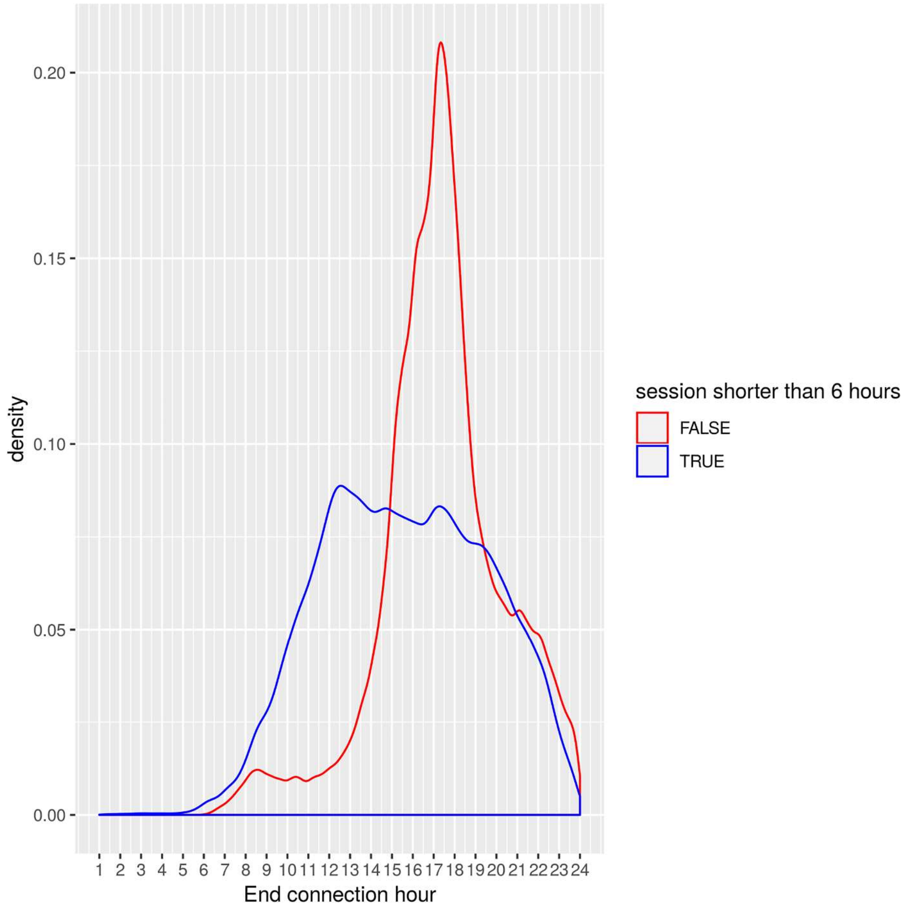


Fig. E6. End connection distribution for sessions shorter and longer than 6 h for DT_SL_SD_WL.

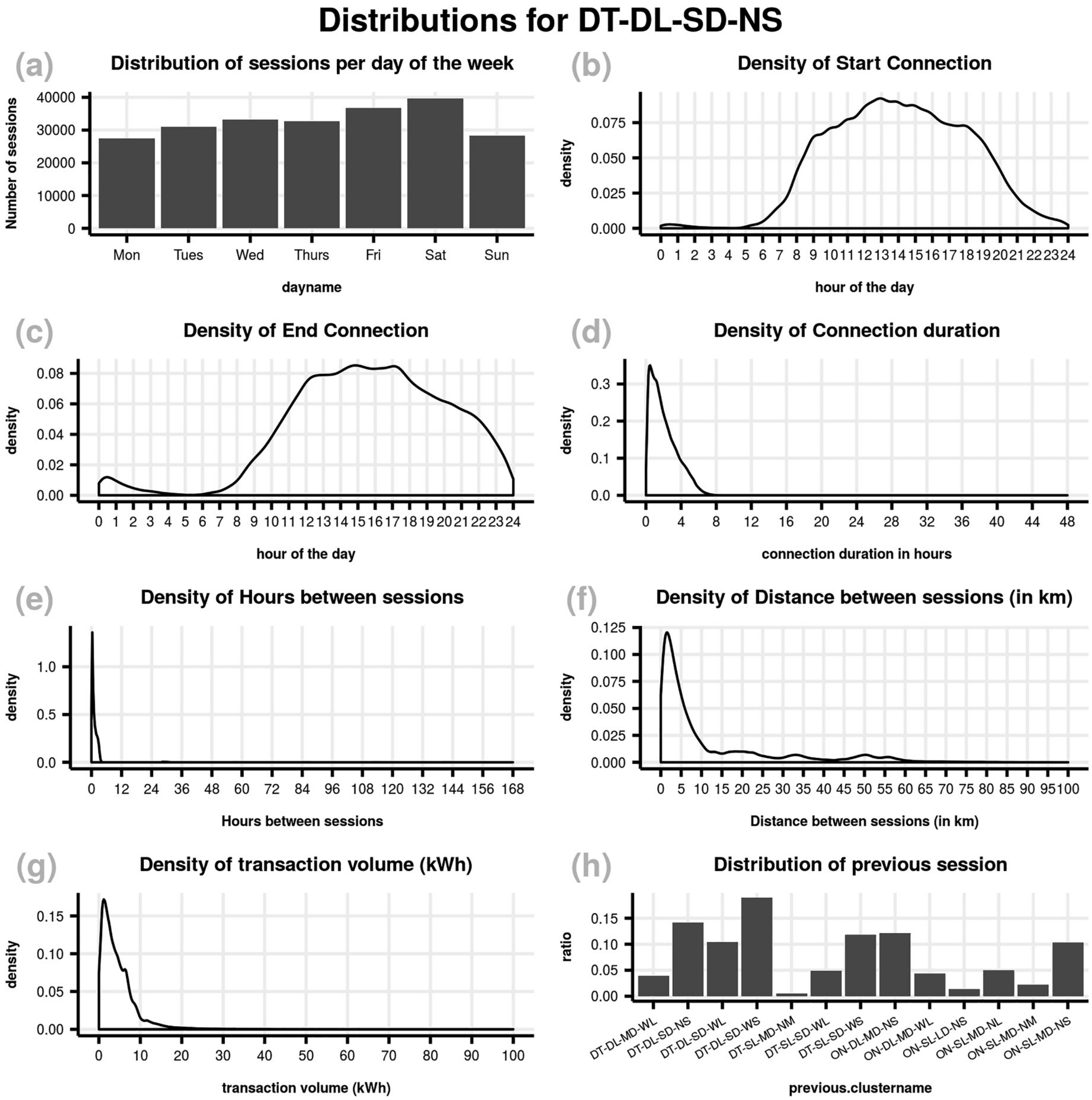


Fig. E7. Distributions for DT-DL-SD-NS.

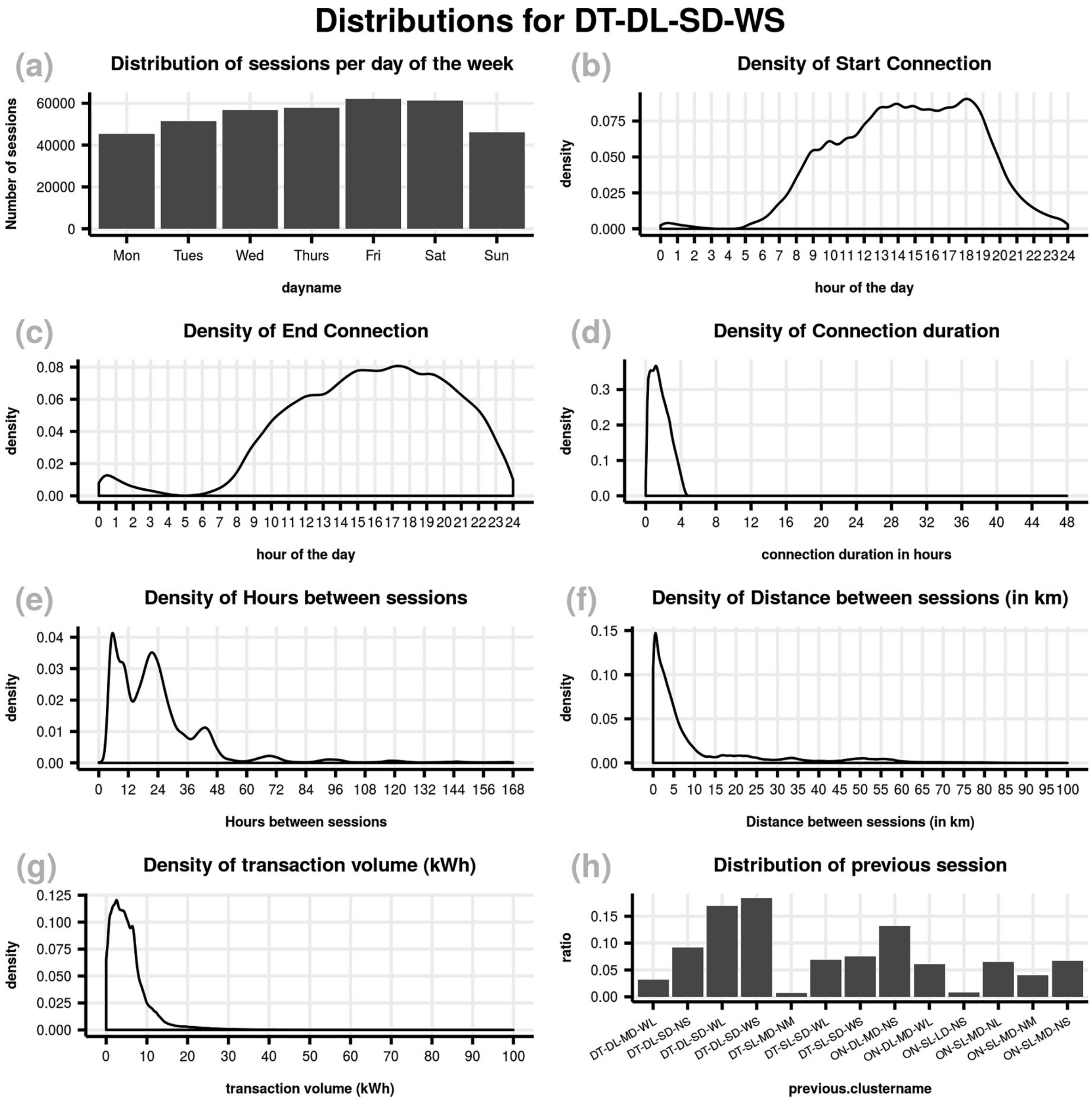


Fig. E8. Distributions for DT_DL_SD_WS.

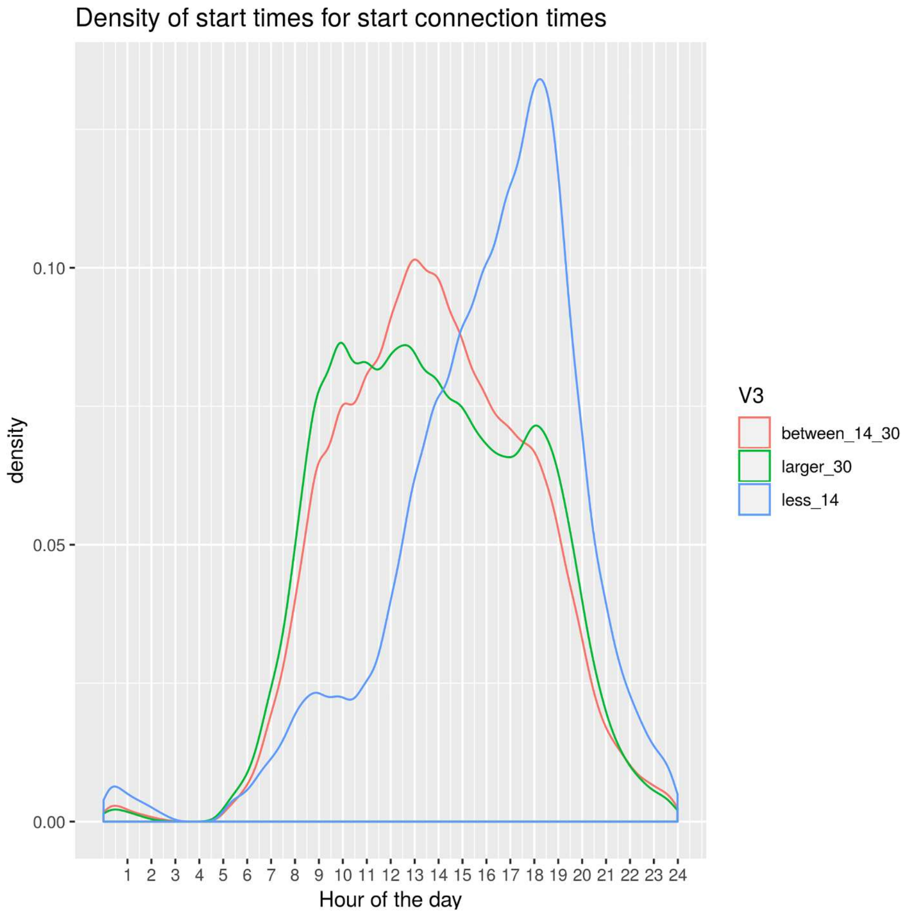


Fig. E9. Deeper analysis of start connection distribution for subgroups of HBS for DT_DL_SD_BS session type.

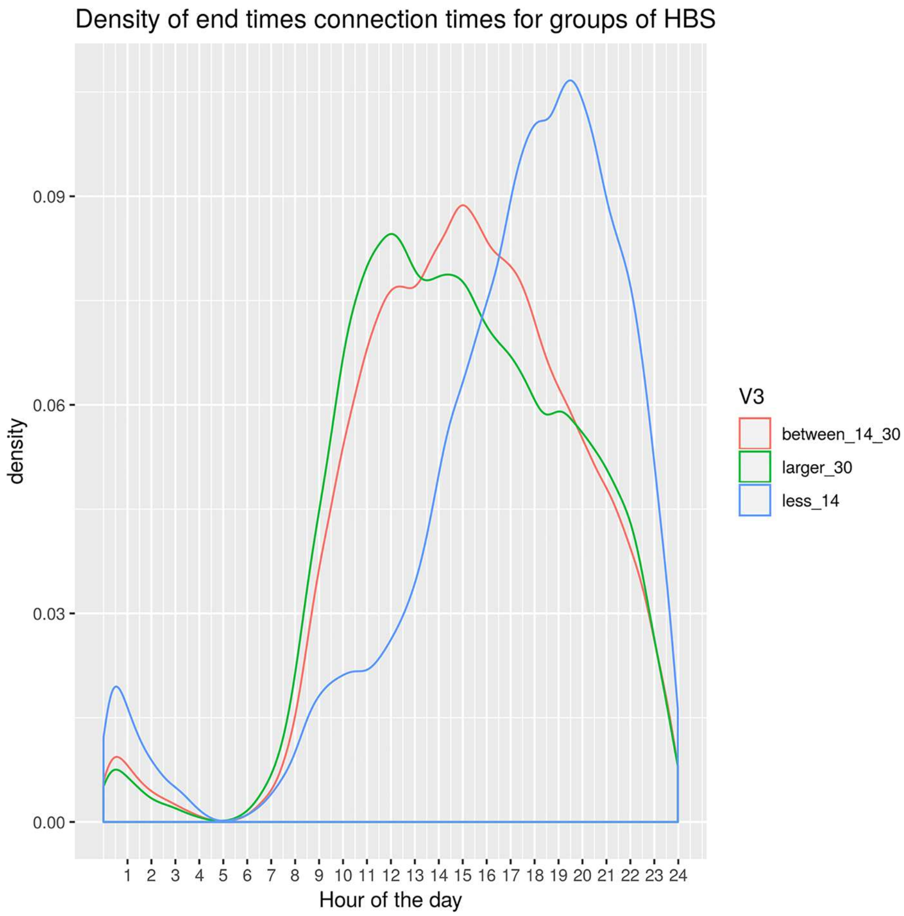


Fig. E10. Deeper analysis of end connection distribution for subgroups of HBS for DT_DL_SD_BS session type.

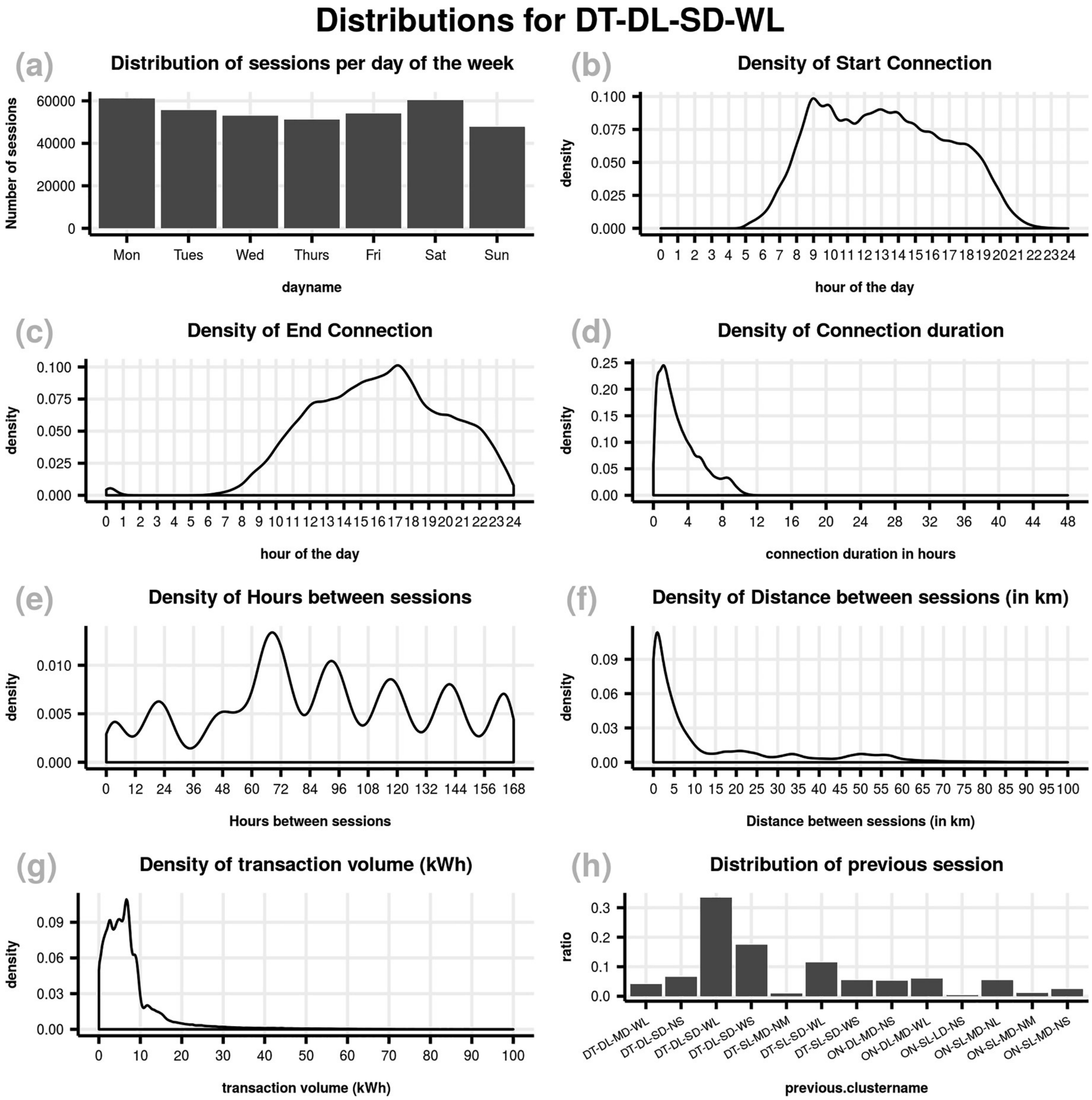


Fig. E11. Distributions for DT-DL-SD-WL.

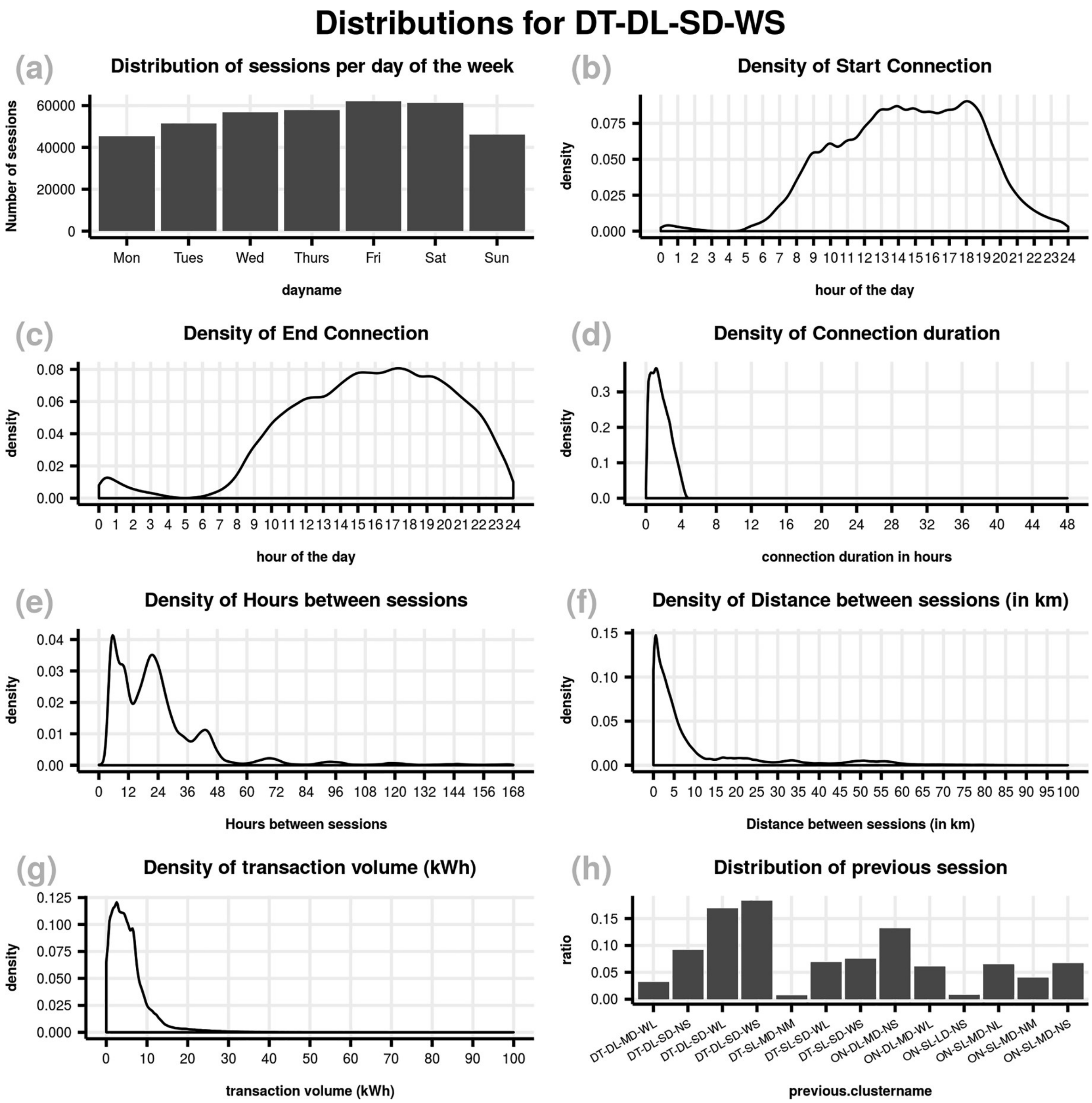


Fig. E12. Distributions for DT-SL-SD-WS.

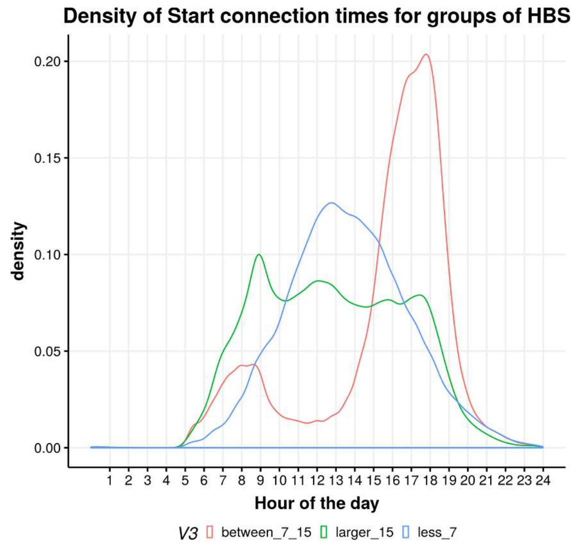


Fig. E13. Distribution of start connection times for subgroups of DT_SL_SD_BS.

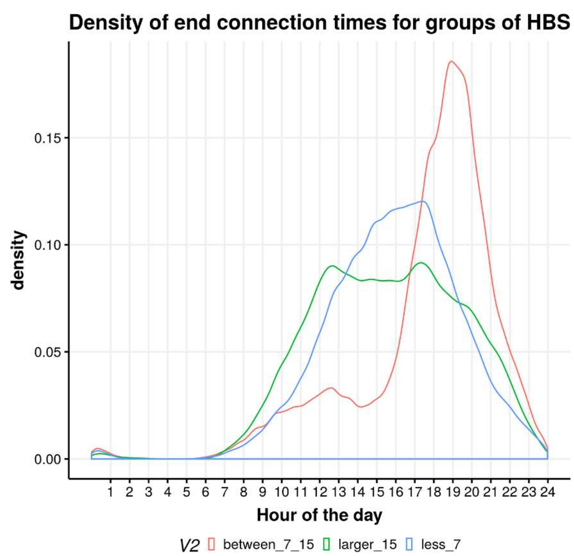


Fig. E14. Distribution of end connection times for subgroups of DT_SL_SD_BS.

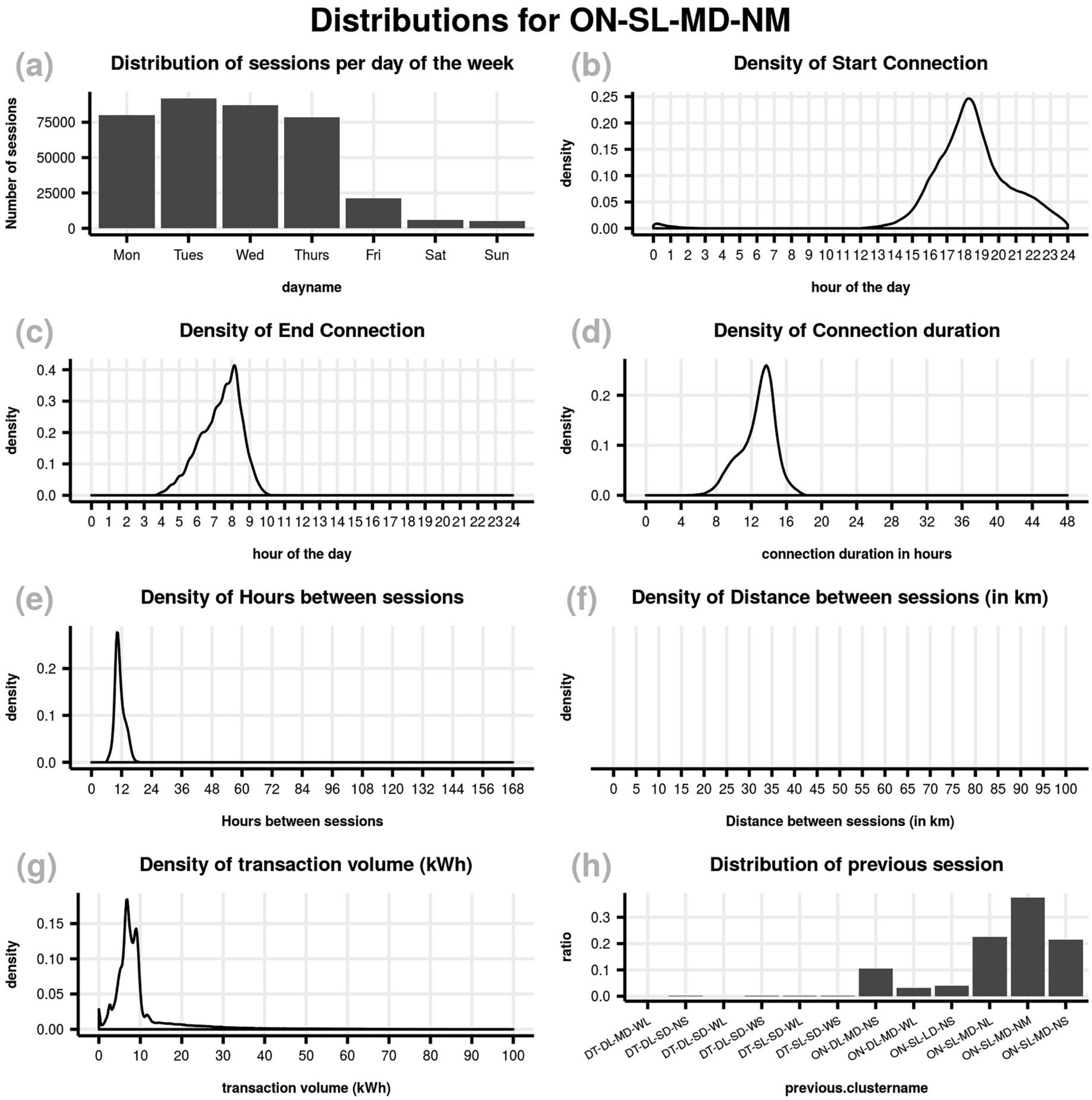


Fig. E15. Distributions for ON-SL-MD-NM.

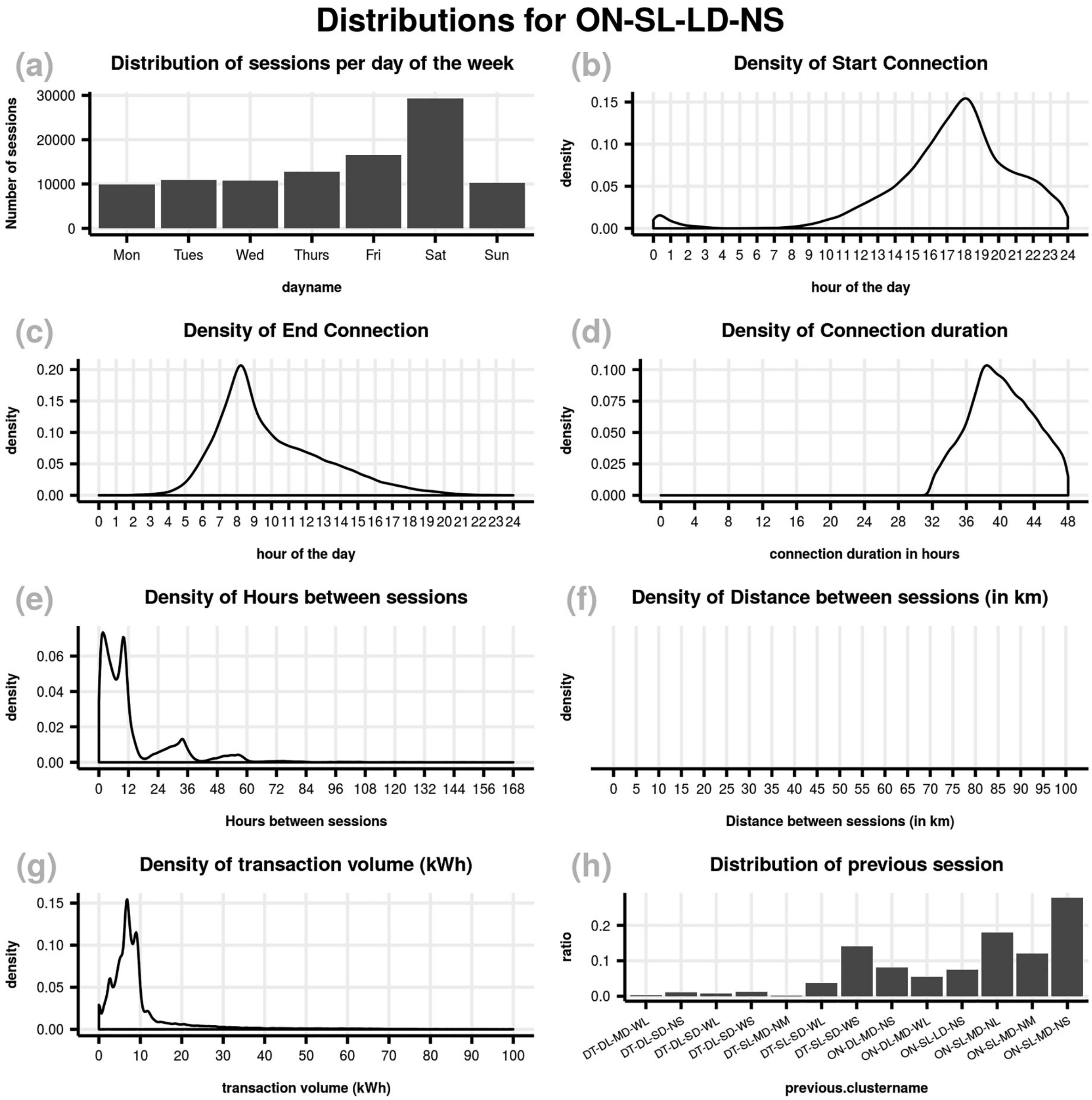


Fig. E16. Distributions for ON-SL-LD-NS.

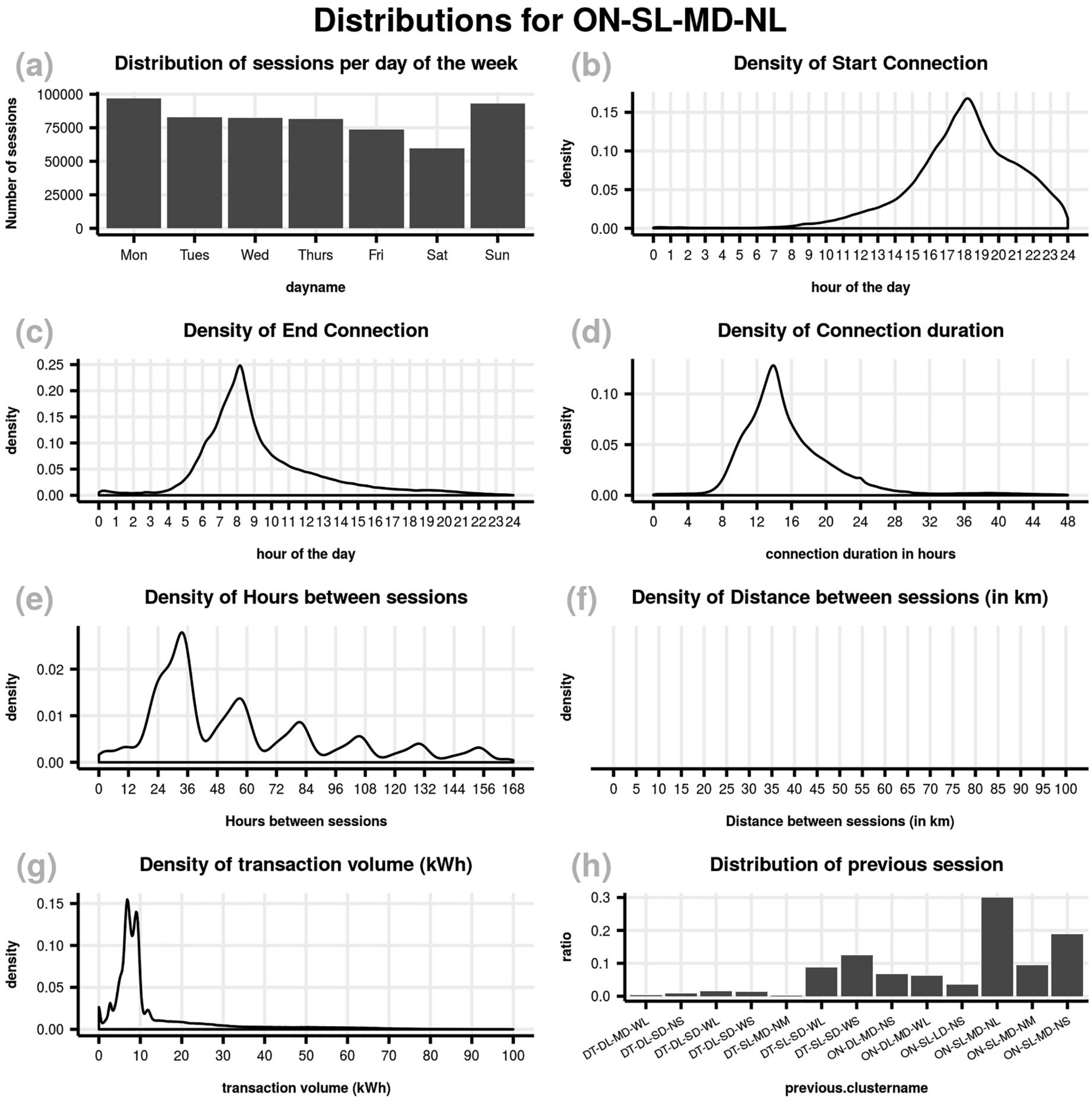


Fig. E17. Distributions for ON-SL-MD-NL.

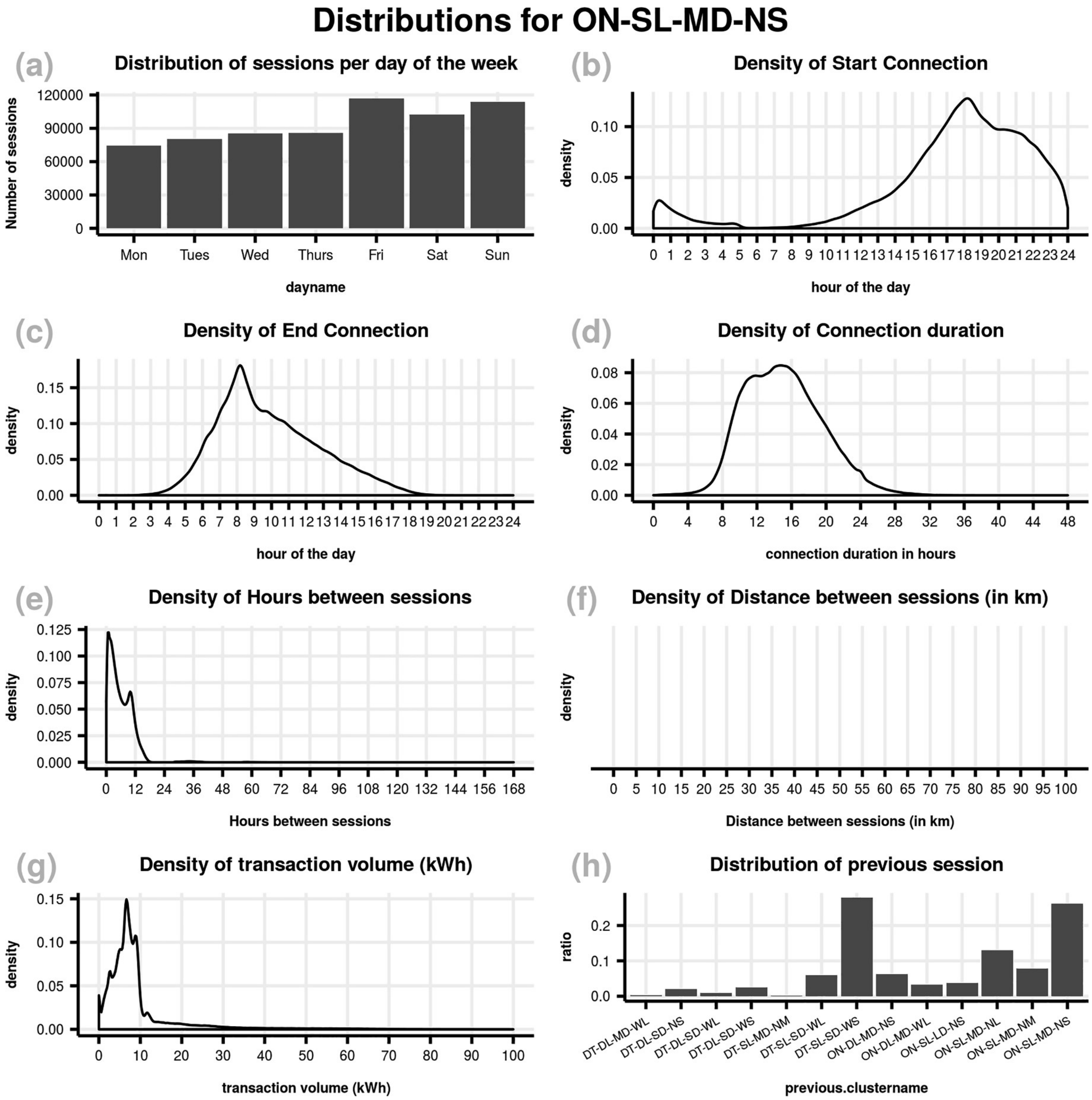


Fig. E18. Distributions for ON-DL-MD-NS.

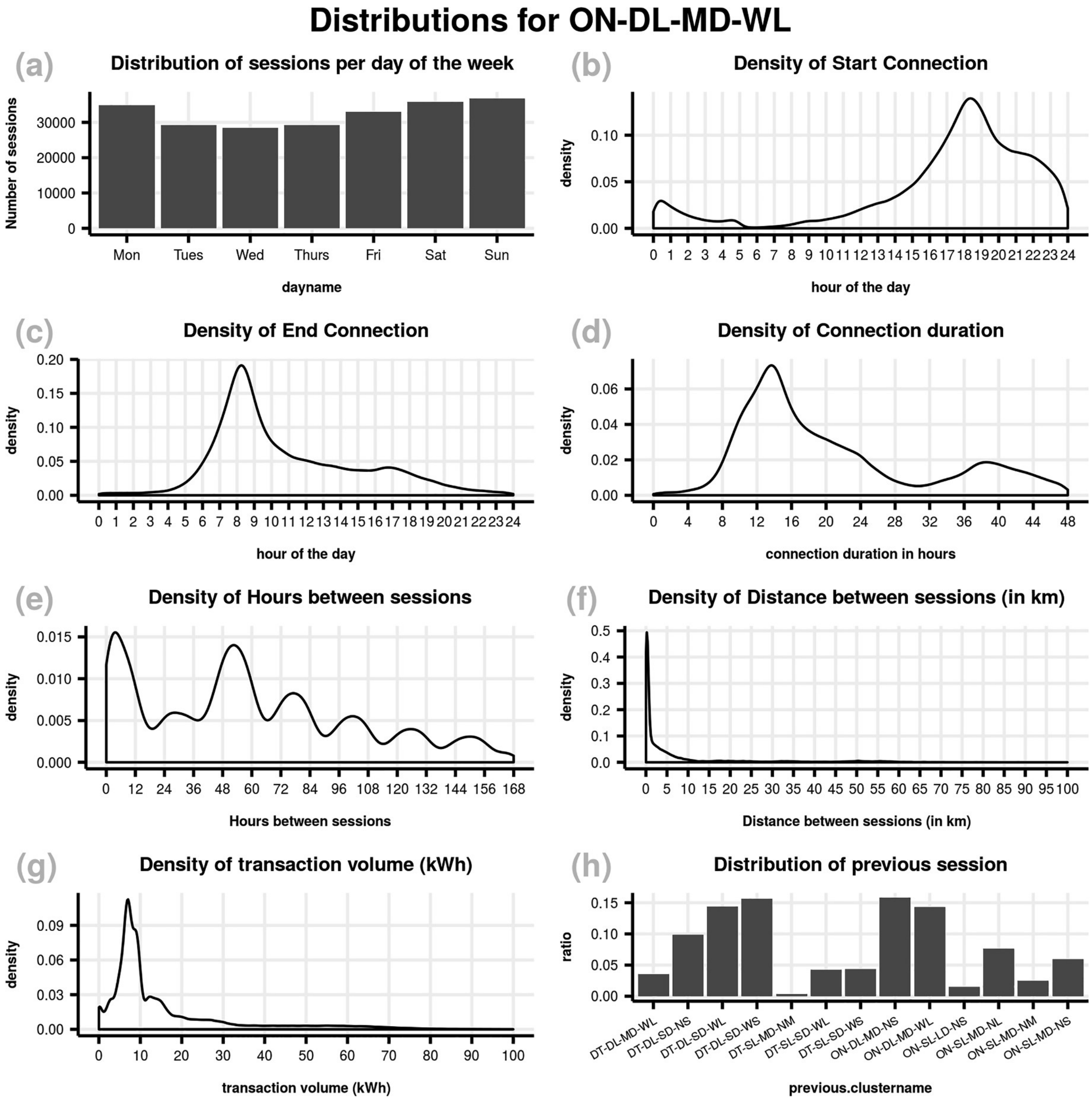


Fig. E19. Distributions for ON-DL-MD-WL.

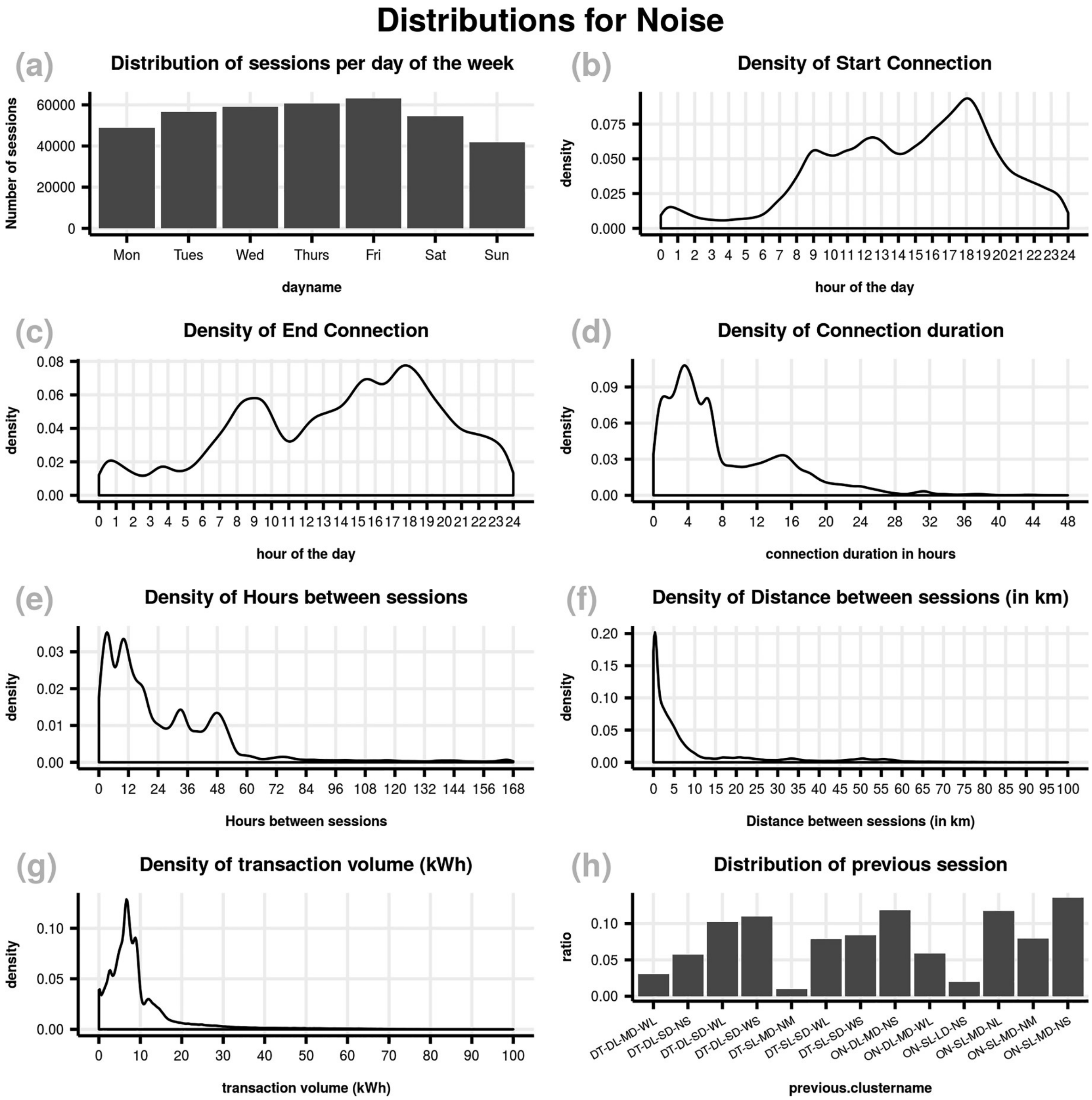


Fig. E20. Distributions for Noise.

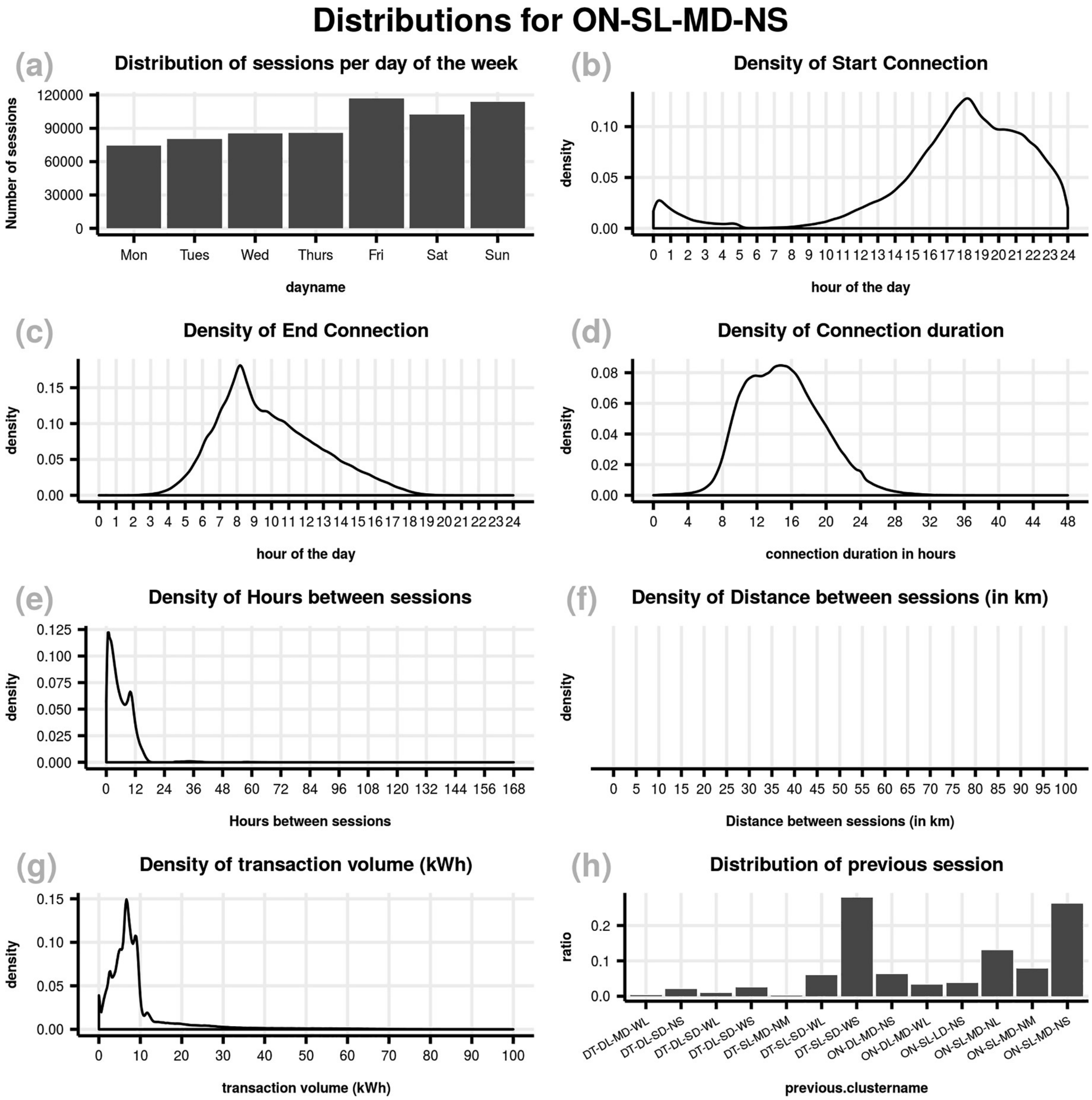


Fig. E21. Distributions for ON-SL-MD-NS.

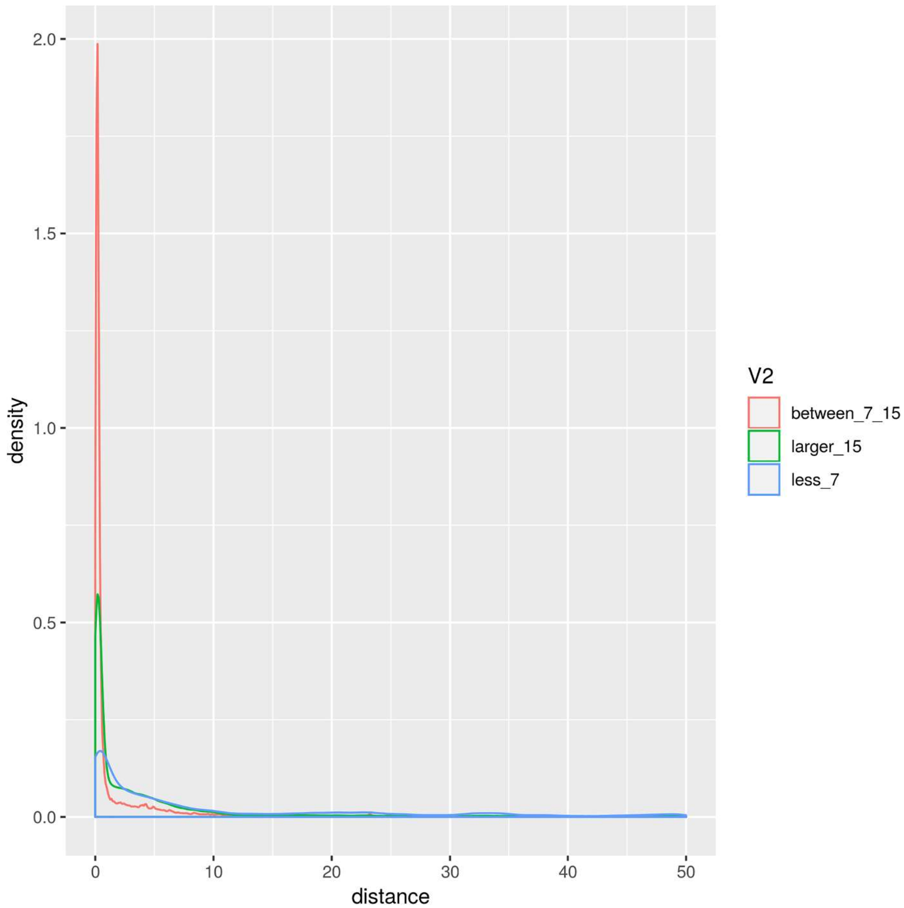


Fig. E22. Distance distribution for sessions with sub sets of HBS feature for ON-SL-MD-NS.

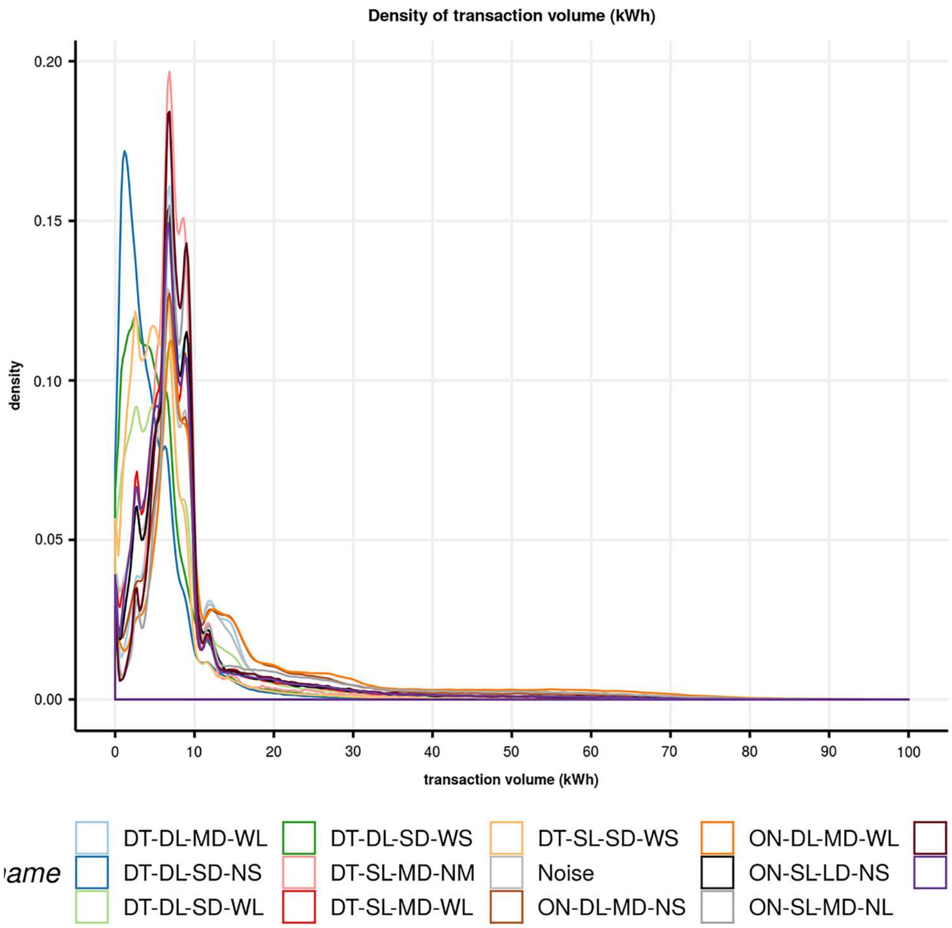


Fig. E23. kWh distribution for all sessions types.

Appendix F. Detailed information on user types

See Figs. F1 and F2 and Tables F1.

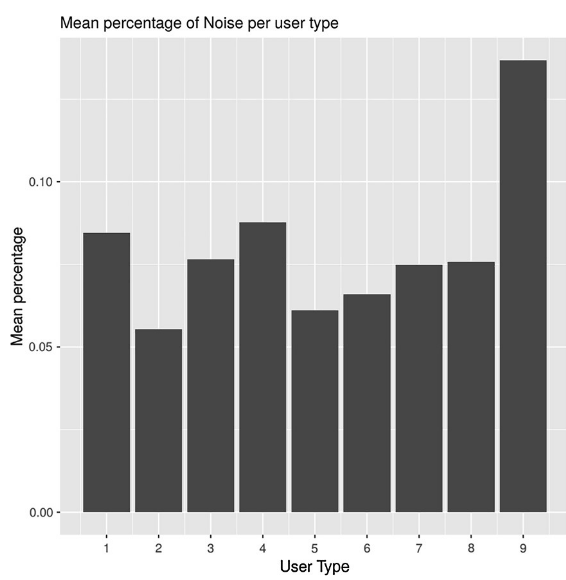


Fig. F1. mean percentage of noise per user type.

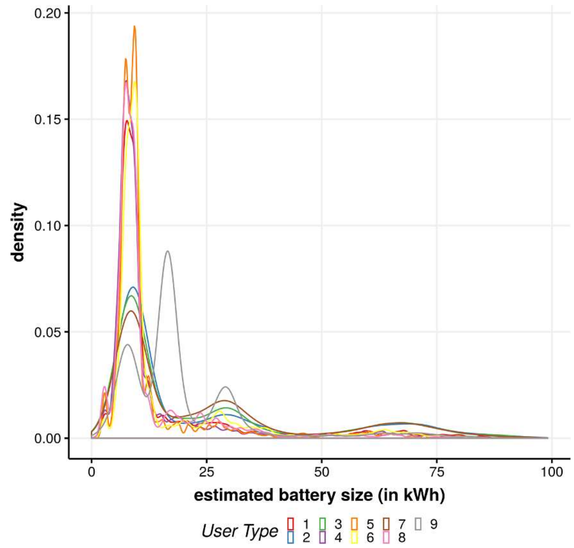


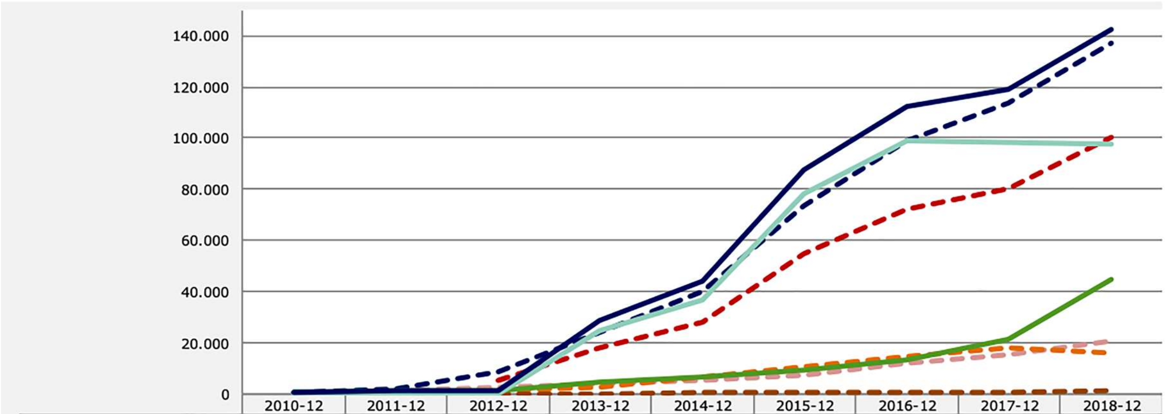
Fig. F2. Battery size distribution for user types.

Table F1
Overview CPs with high rate of User Type 8.

PostalCode	Address	City
1841 GB	Noordervaart 73	Alkmaar
3088GC	Albert Plesmanweg 107	Rotterdam
3207PE	Rijnlaan 29	Nissewaard
3454GT	Kleermakerslaan 30	Utrecht
3224BD	Dreef 112	Hellevoetsluis
2992VW	Schaatsbaan 130	Barendrecht
3431CX	Herenstraat 53	Nieuwegein
2991LT	Ebweg 66	Barendrecht
3451TH	Griendhoeve 8	Utrecht
2691JX	Greenpeacestraat 69	Westland
3136JE	Kraanvogellaan 174	Vlaardingen
2926 TB	Landgoed Sandenburg 17	Krimpen aan de IJssel
3032CG	Hofdijk 651	Rotterdam
2132KJ	Bernard Zweershof 1	Haarlemmermeer
4124AR	Vosstraat 7	Vianen
2262EM	Amstelhof 2	Leidschendam Voorburg
2678HB	Tuba 16	Westland
2516SH	Wiekstraat 64	Den Haag
2254BJ	Burgemeester Van Eijklaan 73	Voorschoten
2631VB	Gentiaan 33	Pijnacker Nootdorp
2272VV	Klaverweide 11	Leidschendam Voorburg
1823AB	Dr. Scheylaan 16	Alkmaar
3197LH	Shannonweg 80	Rotterdam
3062CD	ir. P. Kosterlaan 20	Rotterdam

Appendix G. Analysis of behavioral development over time

See Figs. G1–G3.



	2010-12	2011-12	2012-12	2013-12	2014-12	2015-12	2016-12	2017-12	2018-12
Regular public	400	1.250	2.782	3.521	5.421	7.395	11.768	15.288	20.288
Regular semi-public		576	829	2.249	6.439	10.391	14.320	17.587	15.666
Fast charging		15	63	106	254	465	612	755	1.083
Private			5.000	18.000	28.000	55.000	72.000	80.000	100.000
Total charging points	400	1.841	8.674	23.876	40.114	73.251	98.700	113.630	137.037
BEV Passenger car	266	1.124	1.252	4.161	6.825	9.368	13.105	21.115	44.984
PHEV Passenger car	-	17	65	24.512	36.937	78.163	98.903	98.217	97.702
BEV + PHEV	266	1.141	1.317	28.673	43.762	87.531	112.008	119.332	142.686

Fig. G1. Overview adoption of different types of EV and rollout of charging infrastructure (source RVO.nl, (Netherlands Enterprise Agency, 2018)).

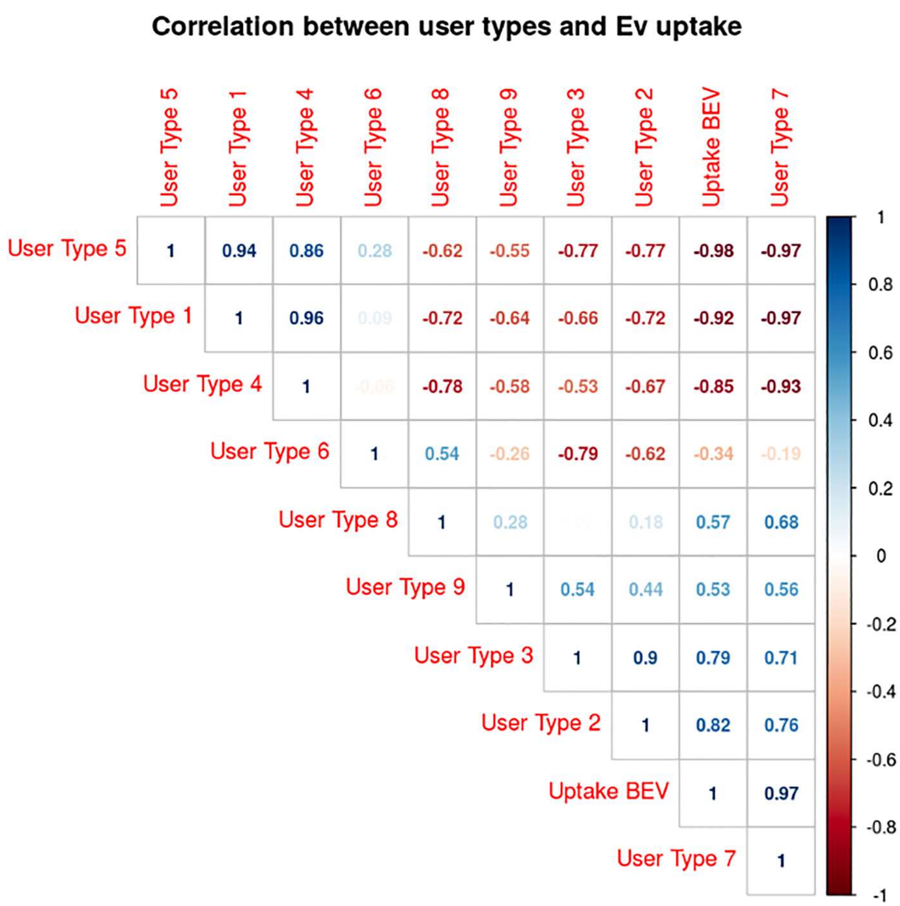


Fig. G2. Correlation between User types and BEV uptake.

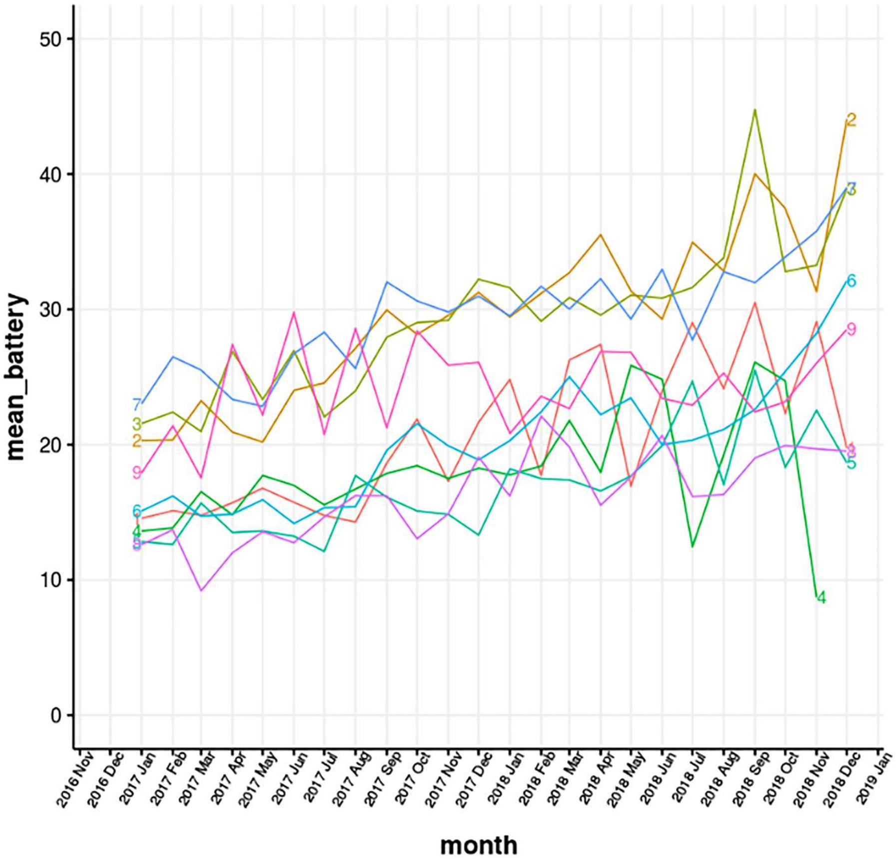


Fig. G3. Mean battery size of new users per user type.

Appendix H. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trc.2020.102637.

References

Ashtari, A., et al., 2012. PEV charging profile prediction and analysis based on vehicle usage data. IEEE Trans. Smart Grid 3 (1), 341–350. https://doi.org/10.1109/TSG.2011.2162009.

Azadfar, E., Sreeram, V., Harries, D., 2015. The investigation of the major factors influencing plug-in electric vehicle driving patterns and charging behaviour. Renew. Sustain. Energy Rev. 42, 1065–1076. https://doi.org/10.1016/j.rser.2014.10.058.

Bae, S., Kwasinski, A., 2012. Spatial and temporal model of electric vehicle charging demand. IEEE Trans. Smart Grid 3 (1), 394–403. https://doi.org/10.1109/TSG.2011.2159278.

Bonabeau, E., 2002. Agent-based modeling: methods and techniques for simulating human systems. Proc. Natl. Acad. Sci. 99 (Supplement 3), 7280–7287. https://doi.org/10.1073/pnas.082080899.

Caroll, S., Walsh, C. and Carroll, S., 2010. The smart move trial: Description and initial results. www.cenex.co.uk. Available at: http://www.cenex.co.uk/ (accessed: 11 September 2019).

Daina, N., Polak, J.W., Sivakumar, A., 2015. Patent and latent predictors of electric vehicle charging behavior. Transp. Res. Rec.: J. Transp. Res. Board 2502 (1), 116–123. https://doi.org/10.3141/2502-14. SAGE PublicationsSage CA Los Angeles, CA.

Daina, N., Sivakumar, A., Polak, J.W., 2017a. Electric vehicle charging choices: modelling and implications for smart charging services. Transp. Res. Part C: Emerg. Technol. 81, 36–56. https://doi.org/10.1016/j.trc.2017.05.006. Pergamon.

Daina, N., Sivakumar, A., Polak, J.W., 2017b. Modelling electric vehicles use: a survey on the methods. Renew. Sustain. Energy Rev 68, 447–460. https://doi.org/10.1016/j.rser.2016.10.005. Pergamon.

Fraley, C., 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput. J. 41 (8), 578–588. https://doi.org/10.1093/comjnl/41.8.578.

Fraley, C., Raftery, A., 2007a. Model-based methods of classification: using the mclust software in chemometrics. J. Stat. Softw. 18 (6), 1–13. https://doi.org/10.18637/jss.v018.i06.

Fraley, C., Raftery, A.E., 2007b. Bayesian regularization for normal mixture estimation. J. Classif. 24, 155–181. https://doi.org/10.1007/s00357-007-0004-z.

Franke, T., Krems, J.F., 2013. Understanding charging behaviour of electric vehicle users. Transp. Res. Part F: Traffic Psychol. Behav. 21, 75–89. https://doi.org/10.1016/j.trf.2013.09.002.

Ge, Y., MacKenzie, D., Keith, D.R., 2018. Gas anxiety and the charging choices of plug-in hybrid electric vehicle drivers. Transp. Res. Part D: Transp.Environ. 64, 111–121. https://doi.org/10.1016/J.TRD.2017.08.021. Pergamon.

De Gennaro, M., et al., 2014. GIS-driven analysis of e-mobility in urban areas: an evaluation of the impact on the electric energy grid. Appl. Energy 124, 94–116. https://doi.org/10.1016/j.apenergy.2014.03.003. Elsevier Ltd.
De Gennaro, M., Paffumi, E., Martini, G., 2015. Customer-driven design of the recharge infrastructure and Vehicle-to-Grid in urban areas: a large-scale application for electric vehicles deployment. Energy 82, 294–311. https://doi.org/10.1016/j.energy.2015.01.039.
Ghazzali, N., 2014. 'NbClust : An R Package for Determining the', 61(6).
Glombek, M. et al., 2018. 'Vulnerability of charging infrastructure, a novel approach for improving charging station deployment'. In: Proceedings of the 2018 Transportation Research Arena.
Gorka, V.C., Helmus, J.R., Lees, M.H., 2019. EVS32 Symposium Simulation of Free-floating vehicle charging behaviour at public charging points. In: EVS32. Lyons, pp. 1–10.
Gottwalt, S., König, L., Schmeck, H., 2015. Data-based assessment of plug-in electric vehicle driving. In: Energy Informatics: 4th D-A-CH Conference, EI 2015 November 12–13, 2015 Proceedings. Karlsruhe, Germany, pp. 115–126. doi: 10.1007/978-3-319-25876-8.
Guo, Q., et al., 2012. Factor analysis of the aggregated electric vehicle load based on data mining. Energies 5 (6), 2053–2070. https://doi.org/10.3390/en5062053.
Haidar, A.M.A., Muttaqi, K.M., 2015. Behavioral characterization of electric vehicle charging loads in a distribution power grid through modeling of battery chargers. IEEE Trans. Indust. Appl. 2015 (1), 483–492. https://doi.org/10.1109/TIA.2015.2483705.
Helmus, J.R., et al., 2018a. Assessment of public charging infrastructure push and pull rollout strategies: the case of the Netherlands. Energy Policy 121, 35–47. https://doi.org/10.1016/j.enpol.2018.06.011.
Helmus, J.R., et al., 2018. Simulation of future electric vehicle charging behaviour – effects of transition from PHEV to FEV. (2), pp. 2–5.
Helmus, J.R. et al., 2019. SEVA: a data driven model of electric vehicle charging behavior, arxiv: 1904.08748. Available at: http://arxiv.org/abs/1904.08748 (accessed: 19 April 2019).
Helmus, J.R., Hoed, R. Van Den, Lees, M.H., 2019. 'Exploring a complex systems approach to charging Infrastructure : implications for researchers and policy makers. EVS32 Symposium (1), pp. 1–17.
Hoed, R. Van Den, et al., 2014. Data analysis on the public charge infrastructure in the city of Amsterdam, pp. 1–10. doi: 10.1109/EVS.2013.6915009.
Van Den Hoed, R., et al., 2014. Data analysis on the public charge infrastructure in the city of Amsterdam. In: 2013 World Electric Vehicle Symposium and Exhibition, EVS 2014, pp. 1–10. https://doi.org/10.1109/EVS.2013.6915009.
Hoekstra, A., Refa, N., 2017. Characteristics of Dutch EV drivers. In: EVS 2017 – 30th International Electric Vehicle Symposium and Exhibition, (October).
Hubert, L., Arabie, P., 1985. Comparing partitions. J. Classif. 2 (1), 193–218. https://doi.org/10.1007/BF01908075. Springer-Verlag.
Idaho National Laboratory, 2015. Plugged in: how Americans charge their electric vehicles, pp. 1–24.
Jabeen, F., et al., 2013. Electric vehicle battery charging behaviour: findings from a driver survey. In: Australasian Transport Research Forum, ATRF 2013 – Proceedings. Available at: http://www.patrec.org/atrf.aspx (accessed: 13 September 2019).
Kansen, M., Waard, J. Van Der, Savelberg, F., 2018. Sturen in parkeren.
Kara, E.C., et al., 2015. Estimating the benefits of electric vehicle smart charging at non-residential locations: a data-driven approach. Appl. Energy 155, 515–525. https://doi.org/10.1016/j.apenergy.2015.05.072.
Khoo, Y.B., et al., 2014. Statistical modeling of Electric Vehicle electricity consumption in the Victorian EV Trial, Australia. Transp. Res. Part D: Transp. Environ. 32, 263–277. https://doi.org/10.1016/j.trd.2014.08.017.
Kim, S., et al., 2017. Heterogeneous hazard model of PEV users charging intervals: analysis of four year charging transactions data. Transp. Res. Part C: Emerg. Technol. 82, 248–260. https://doi.org/10.1016/j.trc.2017.06.022.
Kontou, E., et al., 2019. Understanding the linkage between electric vehicle charging network coverage and charging opportunity using GPS travel data. Transp. Res. Part C: Emerg. Technol. 98, 1–13. https://doi.org/10.1016/J.TRC.2018.11.008. Pergamon.
Li, W., et al., 2017. A review of factors influencing consumer intentions to adopt battery electric vehicles. Renew. Sustain. Energy Rev. 78, 318–328. https://doi.org/10.1016/J.RSER.2017.04.076. Pergamon.
Mies, J., Helmus, J., Hoed, R. Van Den, 2017. Estimating the charging profile of individual charge sessions of Electric Vehicles in the Netherlands. In: Electronic Vehicle Symposium, vol. 30, pp. 1–12.
Milligan, G.W., 1981. A monte carlo study of thirty internal criterion measures for cluster analysis. Psychometrika 46 (2), 187–199. https://doi.org/10.1007/BF02293899. Springer-Verlag.
Momtazpour, M., et al., 2012. Coordinated clustering algorithms to support charging infrastructure design for electric vehicles. In: Proceedings of the ACM SIGKDD International Workshop on Urban Computing - UrbComp '12, p. 126. doi: 10.1145/2346496.2346517.
Morrissey, P., Weldon, P., O'Mahony, M., 2016. Future standard and fast charging infrastructure planning: an analysis of electric vehicle charging behaviour. Energy Policy 89, 257–270. https://doi.org/10.1016/j.enpol.2015.12.001.
Nederland Elektrisch, no date. Nederland Elektrisch - Cijfers en statistieken EV's in Nederland. Available at: https://nederlandelektrisch.nl/actueel/verkoopcijfers (accessed: 15 January 2020).
Netherlands Enterprise Agency, 2018. Available at: https://www.rvo.nl/sites/default/files/2018/04/Statistics-Electric-Vehicles-and-Charging-in-The-Netherlands-March2018.pdf (accessed: 24 April 2018).
Neubauer, J., Wood, E., 2014. The impact of range anxiety and home, workplace, and public charging infrastructure on simulated battery electric vehicle lifetime utility. J. Power Sources 257, 12–20. https://doi.org/10.1016/J.JPOWSOUR.2014.01.075. Elsevier.
Nunes, P., Farias, T., Brito, M.C., 2015. Enabling solar electricity with electric vehicles smart charging. Energy 87, 10–20. https://doi.org/10.1016/j.energy.2015.04.044. Elsevier Ltd.
Paffumi, E., et al., 2015. Assessment of the potential of electric vehicles and charging strategies to meet urban mobility requirements. Transportmetrica A: Transp. Sci. 11 (1), 22–60. https://doi.org/10.1080/23249935.2014.913732.
Pagani, M., et al., 2019. User behaviour and electric vehicle charging infrastructure: an agent-based model assessment. Appl. Energy 254, 113680. https://doi.org/10.1016/j.apenergy.2019.113680. Elsevier.
Pan, L., Yao, E., MacKenzie, D., 2019. Modeling EV charging choice considering risk attitudes and attribute non-attendance. Transp. Res. Part C: Emerg. Technol. 102, 60–72. https://doi.org/10.1016/j.trc.2019.03.007. Pergamon.
RVO, 2018. Cijfers elektrisch vervoer | RVO.nl | Rijksdienst. Available at: https://www.rvo.nl/onderwerpen/duurzaam-ondernemen/energie-en-milieu-innovaties/elektrisch-rijden/stand-van-zaken/cijfers (accessed: 23 October 2019).
Scrucca, L. et al., 2016. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. The R journal. NIH Public Access, 8(1), pp. 289–317. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27818791 (accessed: 11 January 2019).
Sun, X.-H.H., Yamamoto, T., Morikawa, T., 2015. Charge timing choice behavior of battery electric vehicle users. Transp. Res. Part D: Transp. Environ. 37, 97–107. https://doi.org/10.1016/j.trd.2015.04.007. Pergamon.
Sweda, T., Klabjan, D., 2011. An agent-based decision support system for electric vehicle charging infrastructure deployment. In: 2011 IEEE Vehicle Power and Propulsion Conference, pp. 1–5. doi: 10.1109/VPPC.2011.6043201.
Triebke, H., Göhler, G., Wagner, S., 2016. Data analysis of PEV charging events in rural and business environments–a load behaviour comparison. In: University Stuttgart and Fraunhofer IAO, Germany, 13th Symposium of hybrid and electric vehicles, Februrar 23rd and 24th.
Vermeulen, I., et al., 2019. Simulation of future electric vehicle charging behavior—effects of transition from PHEV to FEV. World Electric Veh. J. 10 (2), 42. https://doi.org/10.3390/wevj10020042. Multidisciplinary Digital Publishing Institute.
Weldon, P., Morrissey, P., O'Mahony, M., 2018. Long-term cost of ownership comparative analysis between electric vehicles and internal combustion engine vehicles. Sustain. Cities Soc. 39 (March), 578–591. https://doi.org/10.1016/j.scs.2018.02.024. Elsevier.
Wen, Y., MacKenzie, D., Keith, D.R., 2016. Modeling the charging choices of battery electric vehicle drivers by using stated preference data. Transp. Res. Rec.: J. Transp. Res. Board 2572 (1), 47–55. https://doi.org/10.3141/2572-06. SAGE PublicationsSage CA Los Angeles, CA.
Wolbertus, R., et al., 2018a. Fully charged: an empirical study into the factors that influence connection times at EV-charging stations. Energy Policy 123, 1–7. https://

doi.org/10.1016/J.ENPOL.2018.08.030. Elsevier.
Wolbertus, R., et al., 2018b. Fully charged: an empirical study into the factors that influence connection times at EV-charging stations. Energy Policy 123, 1–7. https://doi.org/10.1016/j.enpol.2018.08.030. Elsevier.
Wolbertus, R., van den Hoed, R., 2017. Charging station hogging : a data-driven analysis. Electric Veh. Symp. 30, 1–14.
Wolbertus, R., Van den Hoed, R., 2019a. Electric vehicle fast charging needs in cities and along corridors. World Electric Veh. J. 10 (2), 45. https://doi.org/10.3390/wevj10020045. Multidisciplinary Digital Publishing Institute.
Wolbertus, R., Van den Hoed, R., 2019b. Electric vehicle fast charging needs in cities and along corridors. World Electric Veh. J. 10 (2), 45. https://doi.org/10.3390/wevj10020045. Multidisciplinary Digital Publishing Institute.
Xu, M., et al., 2017. Joint charging mode and location choice model for battery electric vehicle users. Transp. Res. Part B: Methodological 103, 68–86. https://doi.org/10.1016/J.TRB.2017.03.004. Pergamon.
Xu, M., Yang, H., Wang, S., 2020. Mitigate the range anxiety: Siting battery charging stations for electric vehicle drivers. Transp. Res. Part C: Emerg. Technol. 114 (February), 164–188. https://doi.org/10.1016/j.trc.2020.02.001. Elsevier.
Xydas, E., et al., 2016a. A data-driven approach for characterising the charging demand of electric vehicles: a UK case study. Appl. Energy 162, 763–771. https://doi.org/10.1016/j.apenergy.2015.10.151. Elsevier Ltd.
Xydas, E., Marmaras, C., Cipcigan, L.M., 2016b. A multi-agent based scheduling algorithm for adaptive electric vehicles charging. Appl. Energy 177, 354–365. https://doi.org/10.1016/j.apenergy.2016.05.034.
Yang, Y., et al., 2016. Modeling the charging and route choice behavior of BEV drivers. Transp. Res. Part C: Emerg. Technol. 65, 190–204. https://doi.org/10.1016/j.trc.2015.09.008.
Yoo, H., Park, J., 2015. Public charging behavior of Korea's early EV customer. In: EVS28 International Electric Vehicle Symposium and Exhibition, pp. 1–21.
Yu, H., MacKenzie, D., 2016. Modeling charging choices of small-battery plug-in hybrid electric vehicle drivers by using instrumented vehicle data. Transp. Res. Rec. 2572 (1), 56–65. https://doi.org/10.3141/2572-07. SAGE PublicationsSage CA Los Angeles, CA.
Zheng, C. et al., 2015. An energy management strategy of hybrid energy storage systems for electric vehicles. In: EVS28 International Electric Vehicle Symposium and Exhibition, pp. 1–13.
Zoepf, S., et al., 2013. Charging choices and fuel displacement in a large-scale demonstration of plug-in hybrid electric vehicles. Transp. Res. Rec.: J. Transp. Res. Board 2385, 1–10. https://doi.org/10.3141/2385-01.