

Article

Charging Point Usage in Germany—Automated Retrieval, Analysis, and Usage Types Explained

Philipp A. Friese ^{1,*}, Wibke Michalk ², Markus Fischer ², Cornelius Hardt ² and Klaus Bogenberger ²¹ Department of Informatics, Technical University of Munich, 85748 Garching, Germany² Chair of Traffic Engineering and Control, Technical University of Munich, 80333 Munich, Germany; ge32lic@mytum.de (W.M.); markus.dieter.fischer@tum.de (M.F.); cornelius.hardt@tum.de (C.H.); klaus.bogenberger@tum.de (K.B.)

* Correspondence: philipp.friese@in.tum.de

Abstract: This study presents an approach to collect and classify usage data of public charging infrastructure in order to predict usage based on socio-demographic data within a city. The approach comprises data acquisition and a two-step machine learning approach, classifying and predicting usage behavior. Data is acquired by gathering information on charging points from publicly available sources. The first machine learning step identifies four relevant usage patterns from the gathered data using an agglomerative clustering approach. The second step utilizes a Random Forest Classification to predict usage patterns from socio-demographic factors in a spatial context. This approach allows to predict usage behavior at locations for potential new charging points. Applying the presented approach to Munich, a large city in Germany, results confirm the adaptability in complex urban environments. Visualizing the spatial distribution of the predicted usage patterns shows the prevalence of different patterns throughout the city. The presented approach helps municipalities and charging infrastructure operators to identify areas with certain usage patterns and, hence different technical requirements, to optimize the charging infrastructure in order to help meeting the increasing demand of electric mobility.

Keywords: charging infrastructure; electric mobility; usage types; analysis; automated retrieval; clustering; machine learning; socio-demographic data; usage prediction



Citation: Friese, P.A.; Michalk, W.; Fischer, M.; Hardt, C.; Bogenberger, K. Charging Point Usage in Germany—Automated Retrieval, Analysis, and Usage Types Explained. *Sustainability* **2021**, *13*, 13046. <https://doi.org/10.3390/su132313046>

Academic Editors: Davide Giglio and Alice Consilvio

Received: 29 October 2021
Accepted: 19 November 2021
Published: 25 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rise of electric mobility, the task to provide charging infrastructure (CI) has gained importance over the last few years. Whereas in the beginning, CI had to be deployed without a priori knowledge about potential utilization, today, a growing network of CI allows for data-based decisions concerning the amount and type of CI to be built. However, the quality of decisions depends on the availability of data. As most operators of CI keep (at least parts of) the information on charging events (CEs) private, a big challenge is the scarcity of information. Hence, the paper at hand deals with the information that can be retrieved from publicly available sources on the utilization of CI. Therefore, we follow a multi-step approach, including data gathering, usage behavior classification, and usage prediction:

1. We present a steady approach to gather data from a public website about CEs.
2. Using the data, we calculate the utilization per charging point (CP).
3. Based on the utilization, we employ an agglomerative clustering in order to identify usage patterns.
4. In a final step, we use a machine learning approach to predict a CP's usage pattern based on socio-demographic data of its environment.

By implementing the presented approach, operators of CI, as well as city administrations, can be supported in their decisions about the type and amount of CI to be built up in

the future. Due to the availability of socio-demographic data, we chose the city of Munich, Germany, to present our approach.

The remainder of this paper is structured as follows: Section 2 discusses relevant literature and highlights the research gap. The collected data is described in Section 3. Section 4 describes the calculation of utilization, as well as the first part of our machine learning approach that derives usage patterns. The second part of our machine learning approach is presented in Section 5 and predicts usage patterns from available socio-demographic information. Section 6 discusses the results, and Section 7 concludes the paper.

2. State of the Art and Related Work

The following section reviews the relevant literature from which a research gap can be derived. This study addresses this gap in the research for CIs. The charging behavior of Plug-In Electric Vehicles (PEVs) has already been the subject of several studies and has been investigated using various data sources and methodological approaches. For a better understanding, the following review is structured into three parts, with the data basis and the methodological approach used as the sorting criterion. The first part looks at studies that analyze the mobility and charging behavior of PEVs based on vehicle and User data. The second part evaluates publications that deal with user behavior at public charging stations (CS) based on *charging infrastructure data*. The first two parts, thus, describe the usage and charging behavior of PEVs based on real trips and CEs. The last part deals with relevant literature where charging behavior of PEVs has been researched based on *simulation models*.

Table 1 provides an overview of the relevant literature. In addition to the *Authors*, the table contains further information about the *Data Classification Criteria* in the review (EV = vehicle data, Survey, CID = charging infrastructure data, and SM = simulation model), the time *Period* over which the data of the study was collected, the *Number of Charging Points/Vehicles* collected, and whether the *Access to the Data set* is public or not. In addition, the main results of the investigations are summarized.

Table 1. Overview of related literature.

Authors	Data	Period	No. CP/EV	Data Public	Results
Kessler and Bogenberger [1]	EV	2015	40	-	<ul style="list-style-type: none"> • Analysis of the mobility and charging behavior of first-generation PEVs • Derivation of three user types
Krause et al. [2]	Survey	2015–2016	117	-	<ul style="list-style-type: none"> • Empirical charging behavior, thresholds for residual SOC, detour acceptance • Estimates of energy consumption at CE
Philipsen et al. [3]	Survey	2017–2018	1021	-	<ul style="list-style-type: none"> • Charging behavior, influence of personal traits • Comparison of in refueling/recharging behavior PEV and conventional
van den Hoed et al. [4]	CID	2012–2013	>500	-	<ul style="list-style-type: none"> • Numerical evaluation of usage behavior of public CS • Transmitted energy, utilization and idle duration
Wolbertus et al. [5]	CID	2014–2015	>5600	-	<ul style="list-style-type: none"> • Comparison of usage behavior of public CS for different areas
Wolbertus and Van den Hoed [6]	CI	2013–2016	158	-	<ul style="list-style-type: none"> • Analyze effects of use parking spots at CS overnight by all vehicle types • Evaluation of the effects by means of a natural experiment
Wolbertus and van den Hoed [7]	CID	2016	>5600	-	<ul style="list-style-type: none"> • Quantification of idle connection duration
Wolbertus et al. [8]	CID	2014–2016	>5600	-	<ul style="list-style-type: none"> • Classification of CEs according to connection duration • Multinomial regression analysis with the classes formed

Table 1. Cont.

Authors	Data	Period	No. CP/EV	Data Public	Results
Gerritsma et al. [9]	CID	2017–2018	42	-	<ul style="list-style-type: none"> Quantification of the potential for shifting CEs
Almaghrebi et al. [10]	CID	2013–2018	90	-	<ul style="list-style-type: none"> Classification of CEs according to connection duration Multinomial regression analysis with the classes formed
Olk et al. [11]	CID	2016–2018	-	✓	<ul style="list-style-type: none"> Numerical evaluation of usage behavior of public CS Quantification of usage probabilities
Almaghrebi et al. [12]	CID	2013–2019	97	-	<ul style="list-style-type: none"> Evaluation and classification of transferred energy Multinomial regression analysis with the classes formed
Almaghrebi et al. [13]	CID	2013–2019	97	-	<ul style="list-style-type: none"> Prediction of energy demand due to CEs
van der Kam et al. [14]	CID	2016–2018	24,955	-	<ul style="list-style-type: none"> Roll-Out strategy for charging infrastructure Decision tree for structuring recommendations for actions
Hecht et al. [15]	CID	2019–2020	26,951	✓	<ul style="list-style-type: none"> Numerical evaluation of usage behavior of public CS Classification of the results according to CP power rating
Fischer et al. [16]	CID	2020	1156	-	<ul style="list-style-type: none"> Quantification of idle connection duration Correlation analysis share of idle duration with location factors
Anderson et al. [17]	SM	2008	-	✓	<ul style="list-style-type: none"> Quantification of the required number of charging points for PEVs in Germany Agent-based simulation based on German National Travel Survey
Adenaw and Lienkamp [18]	SM	-	-	-	<ul style="list-style-type: none"> Prediction of PEVs charging behavior and charging locations Agent-based simulation

2.1. Vehicle

Research concerning electric mobility always reflected the development within the sector. Early research in the field focused on usage and user adoption of new vehicles, with a switch in later analyses towards usage types and comparisons to users of internal combustion engine (ICE) vehicles. In Reference [1], usage data from 40 BMW i3 vehicles over one year starting in 2015 is analyzed. This vehicle fleet also served as the data basis in the study of Reference [19] for the investigation of energy-efficient routing algorithms, in Reference [20] as the basis for estimating the potential for vehicle-to-grid concepts and in Reference [21] for evaluating public charging concepts for curb parking. As can be seen, vehicles were not utilized as a household's primary vehicle, as the lack of commuting patterns suggests. Especially the length and duration of seized trips hint towards a cautious usage of PEVs. Hence, the presented charging patterns and thresholds are a first rough estimation about future usage for PEVs, where CEs predominantly started in the afternoon at home with an average state-of-charge (SOC) of 54% and a traveled average distance of 50 km since the last CE, solely a fraction of the available range of the utilized vehicles. After all, such results highlight two problems: (1) early PEVs available were not perceived as a full substitute for ICE vehicles, and (2) pilot programs considering the topic only had data from early-adopters available, that did not necessarily reflect later usage of vehicles and infrastructure properly.

A small survey conducted among PEV users in 2015/2016 and concerning charging behavior is presented in Reference [2]. Users here, as well, stated to perform on average around 40% of CEs at public CI, while residual SOC when starting to look for a charging point (CP) averaged at 28% in rural areas and 19% in urban areas. These results indicate that users project between 30 to 40 km as a reserve to keep moving with their vehicles. Based on calculated values for SOC at CP arrival, realistic fast-charging CP usage was

stated around 63% of battery capacity, narrowing down the energy quantities necessary to be available. After all, the work gives insights into early EV usage. Comparing the results to Reference [1], it becomes clear that users tend to keep a safety threshold in terms of battery capacity to reach their next CP, reducing the available battery capacity. Comparing rural and urban area behavior, a continuously available network of CPs seems to reduce the safety threshold kept by users, allowing for more effective range to utilize. A later study, based on a survey of 1021 respondents in 2017/2018 with users of ICE and PEV [3], reflects parts of the charging patterns, already seen in Reference [1,2]. The main findings include differences, as well as similarities, in vehicle usage. PEV users tend to charge vehicles more often with smaller quantities of energy in a more habitual charging pattern at known CPs, especially when comparing to battery sizes. The share of CEs carried out at public CI is found to be around 40%, a result also found in other literature [2]. Furthermore, it is found that personality traits are an influential factor in refuel/recharge patterns for both types of vehicles.

Focusing on differences in the choice of CPs, a survey among 3201 users of PEV/Plug-In Hybrid Electric Vehicles (PHEVs) is utilized to build a Discrete Choice Model in Reference [22]. Results show that choosing between public charging, home charging, and work charging is predominantly related to housing situation, electricity prices, and vehicle range. While users living in single detached homes chose home charging in 37% of all cases, this number drops to 12% for all users living in apartments. PEV users with free charging at work choose this option in 44% of all cases, compared to 15% when it is not free. Finally, users with longer ranges show the tendency to charge solely at home or at work, while public CI is neglected. Alternative approaches utilizing in-vehicle data can be found in Reference [23], where data from on-board units of ICE vehicles are utilized to find initial locations for CPs, and Reference [24] where such data was incorporated to simulate demand for CPs in context of commercial vehicle usage.

2.2. Charging Infrastructure

In parallel, research was conducted on usage data from CI, especially usage data of CPs, allowing for insights of utilization patterns of the particular infrastructure and in the context of its spatial and demographic environment. By utilizing such data, it was possible to gain a more detailed understanding in terms of optimization, infrastructure requirements, and policy application in order to foster the distribution of PEVs.

The usage behavior of electric vehicle drivers at public CSs was investigated for the first time based on a comprehensive data set in Reference [4] and was followed up by Reference [5–8]. The research is based on several million CEs recorded and processed as part of the *Intelligent Data driven Optimization of CI (IDO-Laad)* research project at public CSs in the Amsterdam metropolitan region and the cities of Amsterdam, Rotterdam, The Hague, and Utrecht. In Reference [4], the usage behavior of PEVs at public CPs in Amsterdam was investigated from April 2012 to April 2013, thus being at an early stage of electric mobility. The data of the CEs was analyzed numerically to describe the usage behavior based on the number of CEs, the amount of energy transferred and the utilization of CSs on the level of different main districts, as well as for different user groups. Based on this, the usage behavior of public CSs in the Amsterdam metropolitan region and for the same four Dutch cities as before was analyzed in Reference [5] and compared within the framework of a benchmark analysis. The evaluations take place comparatively on the level of the investigated cities and regions, and represent a reference to different roll-out strategies for charging infrastructure. In Reference [6], the results of a natural experiment were documented in which CSs in areas with comparable parking pressure were equipped with and without a parking time regulation. Here, access to parking spaces at CSs was allowed exclusively for PEVs during the day from 10 a.m. to 7 p.m. and in a second period from 10 a.m. to 10 p.m. Outside this period and, thus when parking overnight, the parking spaces were accessible to all vehicle types. The results show the high influence of parking regulations on the occupancy behavior of the CSs. Thus, CSs in regions without parking

time regulations were occupied significantly less by charging PEVs in the evening than charging stations in regions with a corresponding regulation. In Reference [7], the focus was on quantifying the idle connection time of CEs. Here, the term *charging station hogging* was introduced, which describes an occupation of the CP by PEVs even after the charging phase has been completed. The CPs occupied by fully charged PEVs are no longer available for other vehicles. Hogging or idle time, thus, influences the potential energy sales of public CPs and should, therefore, be minimized. In Reference [8], CEs were divided into five different categories depending on the connection duration and analyzed in a multinomial regression analysis. Among other factors, four characteristics of the city were also taken into account. Overall, the investigations are based on a qualitatively and quantitatively high quality data set, which, however, is not freely available. The data was primarily processed by numerical evaluations and partly enriched with socio-demographic data of the cities. However, no reference was made to locations at the level of city districts or city quarters or the derivation of specific connection patterns at CSs of the same performance class.

Analyses based on comparatively smaller data sets were carried out in References [9,10,12,13,25]. The study presented in Reference [9] is based on CI data from 42 public CPs collected over a one-year period between 2017 and 2018. The high-quality data set contains information on the amount of energy transferred, as well as a customer and vehicle identification, in addition to the connection time. The data was used to investigate the potential to time-shift CEs with the aim of relieving the local distribution network during periods of increased power demand. Other studies are based on CI data collected from approximately 100 public CPs between 2013 and 2019 [10,12,13]. In Reference [10], charging behavior is divided into five different categories depending on the connection duration, comparable to Reference [8], and analyzed with a multinomial regression analysis. As location categories, the charging locations were divided into four groups (education, workspace, shopping center, and other public parking lots). In addition, four times of day were defined as temporal categories (morning, afternoon, evening, and night). The results show, for example, that short CEs between 0 and 2 h mainly took place in the afternoon and evening at shopping centers. In Reference [12], the same methodology was applied with four categories depending on the amount of energy transferred, and, in Reference [13], the existing data set was used to assess the prediction of energy demand by CEs at the recorded public CPs through different approaches. Based on the data of 711 CPs in Ireland from 2013–2015, in Reference [25], differences between fast-charging and standard CPs are distinguished, showing differences in CI sites can have an impact on usage. In particular, chargers were aggregated by five different location characteristics. Still, the largest differences in usage characteristics were found between fast-charging and standard CPs. Furthermore, the general distribution of CE-starts is presented for the data, resembling the general distributions found in this work, as well as in Reference [7,15,18,26].

In Reference [14,16], distinct recommendations for decision-makers were derived based on the evaluation of CI data. Central parameter in both studies was the proportion of idle connection time. In Reference [14], a decision tree was developed in order to be able to directly derive political measures for the control and optimization of public charging concepts on the basis of recorded charging behavior. For example, it is recommended to improve the cost efficiency of public CPs by integrating them more closely into the energy grid if there is a high proportion of idle connection time. Approaches to reduce energy costs and the load on the distribution network are discussed as specific measures. In Reference [16], a methodology is presented to quantify the idle connection time at public CPs on the basis of a qualitatively limited data set. For this purpose, the CEs at public CSs were analyzed with information on the amount of energy transferred and connection time within the framework of a scenario-based evaluation, and the results were linked to a structural data set of the charging locations. The results showed a moderately positive correlation between the share of idle connection time and the share of dense housing, which can be attributed to the frequent overnight charging of PHEVs in residential areas without private parking slots.

The studies listed above were based on data sets provided by operators or cities as part of research projects and are generally not freely accessible to third parties. References [11,15] are based on occupancy data of public CPs collected via public websites of several roaming platforms. Although the data is publicly accessible in this case, it only contains data on connection time and is, thus, qualitatively more limited compared to the data sets in the previously listed studies. In Reference [11], data was gathered from a website of the Irish Board of Electricity between 2016 and 2018 and peak periods in the use of public CSs in the study area were derived. A more detailed analysis for Germany was carried out by the same authors in Reference [15]. For this purpose, data from public websites were again collected over a period of 3 months between the years 2019 and 2020 and analyzed with five location types (urban, suburban, industrial, uninhabited, and non-fitting), among others, as part of a linear regression analysis. Results indicate usage of chargers in the data set is rather low with rarely above 20% occupation. Furthermore, differences in usage between chargers in industrial, urban, and suburban areas are detected with urban chargers adding a further peak in the early evening. Finally, the utilized regression model revealed that the most relevant factors in analyzing the usage of CPs is given by spatio-demographic data, while factors concerning specific hours or weekdays perform rather weak.

2.3. Simulative Approaches

A third approach introduced in literature is the application of simulations towards estimates of future CI requirements and potential usage. Based on survey data and basic assumptions of parking behavior and fleet composition, in Reference [17], an agent-based simulation is utilized to find an estimate for the number of CPs needed in Germany. Results show notable influences of on-street charging, home charging, and the share of PHEVs in the fleet towards the number of CPs needed. A more detailed analysis based on an agent-based simulation is presented in Reference [18], where charging behavior and location choice for charging are explicit inputs based on realistic assumptions. As a result of the conducted case study of Munich, Germany, usage-patterns, as well as occupancy patterns, of public chargers are generated and analyzed and further compared to patterns from Reference [7,15], showing only slight differences to real data. Such simulative approaches allow for evaluations of CI requirements in large scale networks, as well as expected usage for local CI, giving insights for further expansion by operators. Based on such approaches, even implications of operative measures can be estimated. Still, relevant inputs and correct causalities are vital for this method; therefore, the results presented in this paper can help to meet these needs for correct simulations.

2.4. Research Gap

The studies listed in the review have investigated the charging behavior of PEVs on the basis of various data sources. The accessibility of the data set presents a major hurdle for verification and further investigation by third parties. Among the studies listed, only References [11,15,17] are based on publicly accessible data sets. Furthermore, the charging behavior was mostly based on classifications of connection duration, charging duration, idle duration, power rating, or the amount of energy transferred. Furthermore, only in a few cases were the derived usage behavior and the land use of the CSs linked for further insights. However, this link represents an important basis for decision-making for the construction and operation of public CSs.

The work presented here closes this gap. For the first time, different connection patterns for CSs of the same power rating are formed and compared with criteria of charging locations by means of a two-step machine learning approach. The usage behavior is based on charging infrastructure data from a freely accessible website and can, thus, be reconstructed and transferred to other areas. The methodology for extracting and processing the connection data is explained in Section 3. The results enable a better understanding of the usage behavior of public CSs by electric vehicle drivers, as expected usage patterns for CSs of the same power rating can be determined and assigned to

objective location criteria in urban areas. Overall, this results in an improved decision-making basis for the selection, installation, and operation of public charging concepts in urban areas.

3. Gathering Publicly Available Data

Having highlighted the research gap in the previous section, this section introduces the first step of the proposed approach. We explain how data was gathered and prepared for the next step, which is presented in Section 4.

The data analyzed in this paper was collected from an online charging infrastructure status map (ISM) of an e-mobility service provider and comprises 2409 charging points in and around Munich between 6 May 2021 and 6 October 2021. Within the observation period, two outages occurred, leading to a loss of three days worth of data. Figure 1 shows the observed area and gathered CPs.

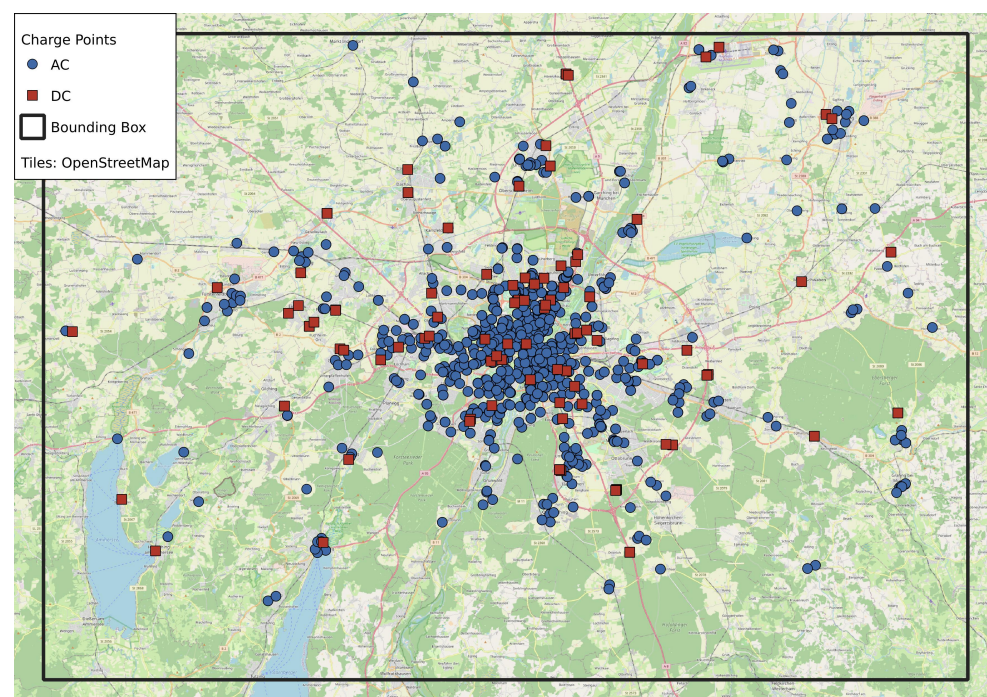


Figure 1. Observed charge points. Blue circles denote AC-only parks, red squares DC-only charge points. Black Box marks observation area with corners south-west at EPSG:4326 coordinates 47.90614, 11.05211 and north-east at 48.36607, 12.03568. Background tile credit: OpenStreetMap contributors.

3.1. Terminology

The Open Charge Point Protocol [27], which standardizes the communication between PEV and CPs, introduces a 3-tier model for classifying CI components. Figure 2 visualizes this model.

A Charging Station refers to a physical system used to charge EVs and potentially contains multiple Electric Vehicle Supply Equipments (EVSEs). An EVSE is defined as “a part of the CS that can deliver energy to one EV” [27]. An EVSE is usually associated with one CP, also referred to as connector, i.e., the physical outlet. An EVSE may comprise multiple CPs, for example, two electrical outlet types to support multiple plugs; however, only one CP of an EVSE may be actively used at any one time. An EVSE is associated with an EVSE identifier (EVSE-ID), which uniquely identifies the comprising CP(s). This paper operates on the layer of CPs, identified by their EVSE-ID.

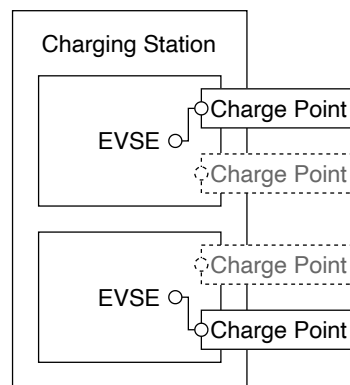


Figure 2. Three-tier model as defined by the Open Charge Point Protocol.

3.2. Retrieval

Data from the ISM was retrieved every 15 min and comprises the status per CP, the associated EVSE ID, and the timestamp since the observed status went into effect. The CP may be in status available, occupied, defect, or unknown. Status 'occupied' indicates a connected but not necessarily charging vehicle. Information on the charging time and transmitted power is not available. Table 2 describes the gathered data.

Table 2. Schema of gathered data points.

Name	Description	Unit
EVSE-ID	EVSE Identifier, identifies one CP	string
status	CP Status (<i>available, occupied, defect, or unknown</i>)	string
status timestamp	Timestamp since the status went into effect	timestamp

A Python system was developed, which repeatedly fetches the status information for each CS displayed on the ISM. Data gathering consists of two phases: First, an overview data-set is gathered in a singular request, which contains the fused status of each CS. A CS is assigned the fused status 'occupied' if at least one of the contained CPs is occupied; otherwise, it is assigned the status 'available'. Second, the detailed status information per CS is requested, which yields status information on the CP level. Figure 3 visualizes the two-phase data gathering.

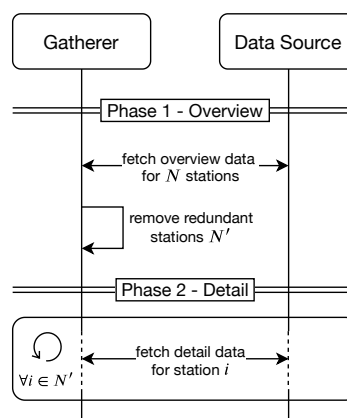


Figure 3. Overview of two-phase gathering sequence. Phase 1 fetches overview data for all stations in one request, and Phase 2 iterates through all stations and fetches detail data.

Since the detailed status information in Phase 2 has to be requested for each individual CS, several hundred requests have to be sent for a complete gathering run. In order to reduce the network and server load, the overview data gathered in Phase 1 is used to filter

the CSs requested in Phase 2. A request is considered redundant if both the previously recorded and currently gathered CS status is 'available'. Then, it is concluded that either no charge events have happened in between, or potential CEs have been missed. In either case, no new information may be gathered, and the Phase 2 request is filtered.

Depending on the time of day, this filtering approach reduces the number of requests per gathering run by 30% to 50%. In addition, the gathering system is rate-limiting the requests to further reduce network and server load.

3.3. Transformation

Using the provided status timestamp, the complete event intervals are reconstructed: The event start of, for example, a CE is set to the timestamp since the observed 'occupied' status went into effect. The event end is set to the status timestamp of the subsequent event. Assuming no events are missed, then, every event interval is accurately reconstructed. This transformation also captures events shorter than 15 min, assuming that the status was gathered during occupation. Events shorter than 15 min occurring entirely between two gathering runs, however, cannot be recorded.

In the event of a brief 'available' event shorter than 15 min, which separates two CEs, then, the separating 'available' event is not recorded. However, the two CEs are still distinguished, since their respective start timestamps differ. The first CE is set to end at the start of the second CE, and the separating 'available' event is, thus, considered as a part of the first CE. CE intervals are reconstructed with no error if the separating event is recorded, or with an error of, at most, 15 min, if the separating event is not recorded.

Table 3 shows the schema of the transformed data points.

Table 3. Schema of transformed data points.

Name	Description	Unit
EVSE-ID	EVSE Identifier, identifies one CP	string
status	CP Status (<i>available, occupied, defect, or unknown</i>)	string
status start	Timestamp marking status start	timestamp
status end	Timestamp marking status end	timestamp

4. Deriving Usage Patterns of Charge Points

Based on the data gathered according to Section 3, this section gives an overview of characteristics of the charging data and the calculation of utilization, as well as the first part of our machine learning-approach: identifying utilization patterns of public CI.

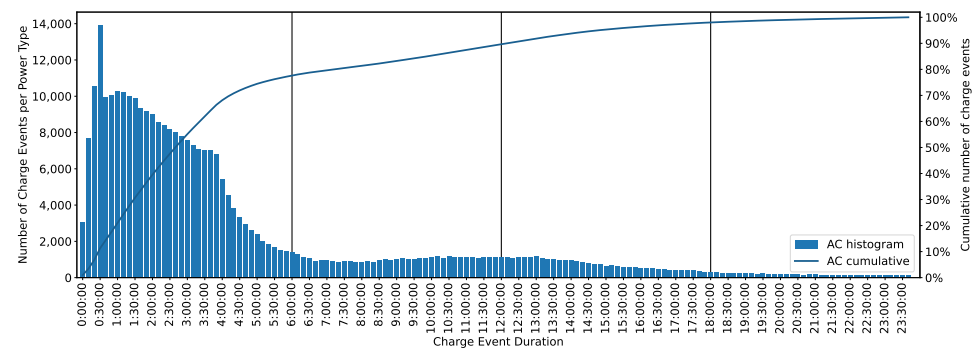
4.1. Charge Event Lengths

In a first step, the gathered data from public CI is described. Figure 4 shows a histogram of the gathered CEs by duration and power-type with a bin size of 10 min. Based on this depiction, several observations can be made:

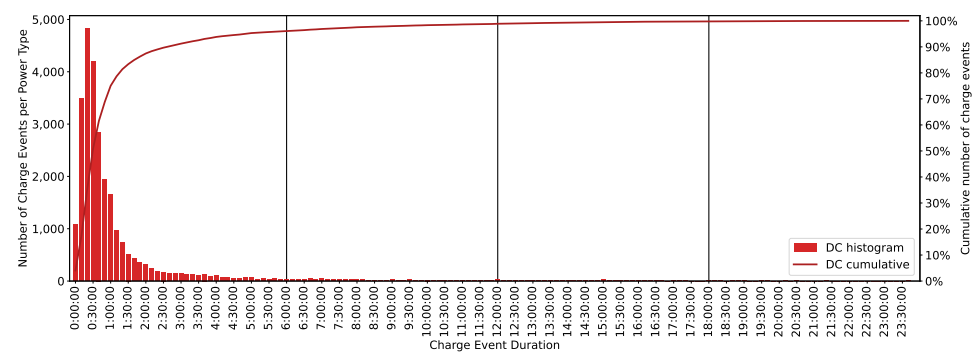
Figure 4a displays a visible decrease at the four-hour mark. The gathered data comprises CPs in and around the city of Munich, and 51% of all AC CPs are operated by the Stadtwerke München (SWM). At several AC CPs, SWM enforces a maximum charge duration of four hours between 8 a.m. and 8 p.m.

In addition, about 20% of all AC CEs are between 6 and 18 h long, which primarily originate from over-night charging. Figure 5 shows a histogram of CE starts and compares the duration groups 0 to 6 h (blue) and 6 to 18 h (orange). The latter group peaks in the evening hours, which represent overnight CEs.

Figure 4b displays the histogram of DC CPs and shows a peak at the 30 min mark. The variance is significantly lower as compared to Figure 4a, and over 90% of all DC CEs are, at most, 2 h long.



(a) Histogram for AC CPs



(b) Histogram for DC CPs

Figure 4. Histogram and cumulative charge interval lengths per power type with bin size 10 min, (a) shows AC CPs, and (b) shows DC CPs. Data from 6 May 2021 to 6 October 2021.

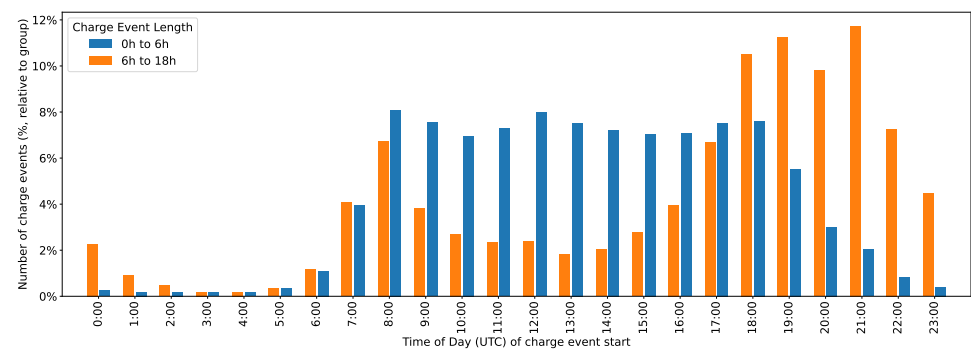


Figure 5. Histogram of AC CE starts for events shorter than 6 h and between 6 h and 18h, bin size one hour. CEs 6 to 18 h long primarily start in the evening hours. Data from 6 May 2021 to 6 October 2021.

4.2. Charge Event Anomaly

Figure 6 shows a histogram of CEs by length for AC CPs with a bin size of 30 s and truncated to, at most, 6 h in charge length for readability. In comparison to Figure 4a, which displays the same data but with a bin size of 10 min, three additional observations can be made.

First, a spike in the number of CEs at 30 min is visible. Of all 4576 CEs in this bin, 2620 originate from 20 CPs from one operator. All 20 CPs exhibit CEs with precisely 30 min in length, although not all CEs from these CPs display this behavior. This suggests that either some customers of this operator are permitted to charge, at most, for 30 min, or the operator automatically stops CEs after 30 min.

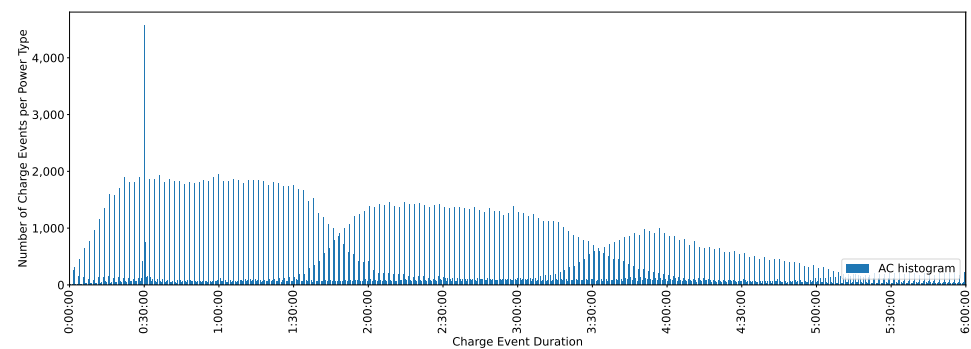


Figure 6. Histogram and cumulative charge interval lengths for AC CPs with bin size 30 s, truncated to, at most, 6 h. Data from 6 May 2021 to 6 October 2021.

Second, the histogram does not show a continuous distribution of CEs. Instead, a discretization pattern is visible with a frequency of two minutes, although not all events follow this pattern. It is speculated that some component in the CI polls data with a two-minute frequency. The section *J. MeterValues* in the Open Charge Point Protocol introduces the Transaction Meter Values, which describe “frequent (e.g., 1–5 min interval) meter readings taken and transmitted (usually in ‘real time’ to the CSMS [Charging Station Management System]) [...]” [27]. The frequency is controlled by two parameters denoting the “time (in seconds) between sampling of metering (or other) data”. This suggests that some operators have set these parameters to 2 min, resulting in the visible pattern.

Third, three distributions are visible, which display an interference-like pattern around 1 h, 45 min and again at 3 h, 30 min. Figure 7 shows a close-up of the first interference in Figure 6.

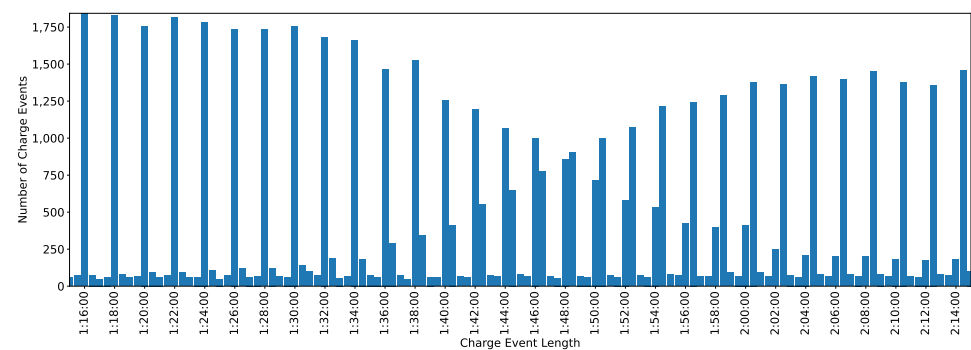


Figure 7. Close-up of charge interval lengths for AC CPs with bin size 30 s. Data from 6 May 2021 to 6 October 2021.

Assuming CPs dispatch status information every two minutes, then, the CE length will be a multiple of 2 min. Choosing a bin size of 30 s leads to a histogram with every fourth bin containing the majority of events. These bins can be described with the set $B = \{f * x \mid x \in \mathbb{N}_0^+\}$, with f as the polling interval in minutes. Elements $b \in B$ represent the left bin edge.

The observed interference pattern may be explained by a continuous drift in the charge station clock. This drift then causes delayed transaction meter readings: readings which should be taken at 1 h, 58 min may instead be taken at 1 h, 58 min, 30 s after starting the CE, which will subsequently be assigned the next-larger bin. For a dispatch frequency f of 2 min and a bin size b of 30 s, the bins with the majority of events can be described with the set $B^\alpha = \{f * x + \alpha b \mid x \in \mathbb{N}_0^+\}$, with $\alpha \in \mathbb{N}_0^+$ determining the phase shift or the number of bins a meter reading is shifted rightwards. For the given bin size and dispatch frequency, a packet can, at most, drift three bins before it is ‘in-phase’ again.

Figure A1a displays the complete histogram for AC CPs and shows additional interfering groups with larger α values. Figure A1b shows the 30-s histogram for DC CPs and displays a similar discretization pattern. Due to the limited length, however, interference patterns do not appear.

The source of this phase drift remains to be speculated. However, the observed interference patterns do not contain information with regards to the CP usage. Instead, they are a consequence of the observed two-minute discretization, assumed clock drift, and the employed visualization method. The smallest bin size that does not lead to interference patterns is two minutes, which will be used in the subsequent analyses.

4.3. Occupancy Analysis

In this paper, we define “occupancy” as the number of CEs relative to the total number of possible CEs, per time interval and per CP. An occupancy of 40% between 11:00 and 11:10 denotes that between 11:00 and 11:10 on any day, on average 40% of all CPs have been observed to be occupied. Calculating the occupancy for CEs, which comprise the start and end timestamp, is similar to a histogram raster approach, which is described below.

4.3.1. Calculating Occupancy

For a time interval f of ten minutes, a raster R is generated, containing 144 bins, each representing the number of CEs for one ten-minute slice. Next, all bins which are overlapped by a CE, denoted e , are determined and the counter at each bin is incremented by one. This process is repeated for all CEs.

We define raster R as a sorted, zero-initialized array, with length $N = \lceil \frac{60 \cdot 24}{f} \rceil$ denoting the number of f -sized bins in R , assuming R spans one day. Each element stores the counter of the associated bin. For a resolution f of 10 min, this gives an array R of length 144 with bins $[00:00, 00:10)$, $[00:10, 00:20)$, \dots , $[23:50, 00:00)$.

In order to determine the bins overlapped by event e , the first and last bin is determined by considering the time-of-day component of the event start and end timestamps e_{start} and e_{end} . Each bin $b \in R$ can be addressed by its index in R , starting at 0. Using timestamps e_{start} and e_{end} , the bin indices i_e^{start} and i_e^{end} are calculated, which denote the bin containing the CE start and end, respectively. Assuming that event e does not exceed a length of 24 h, it can then be concluded that all bins between i_e^{start} and i_e^{end} are overlapped by event e . In case of an event crossing midnight, index i_e^{end} will be smaller than i_e^{start} . In this case, the bins after i_e^{start} and before i_e^{end} are considered. Indices i_e^{start} and i_e^{end} for timestamps e_{start} and e_{end} are calculated via the formula:

$$i_e^{start/end} = \lfloor \frac{h(e_{start/end}) * 60 + \min(e_{start/end}) + \sec(e_{start/end})}{f} \rfloor$$

The functions $h()$, $\min()$, and $\sec()$ denote the hour, minute, and second component of $e_{start/end}$, f again denotes the time interval. Algorithm 1 shows the pseudo-code for the transformation approach.

The returned raster R is an array with length N , where each element denotes the number of CEs overlapping the associated bin. These values are then normalized with respect to the total number of possible CEs. The resulting fraction denotes the occupancy as initially defined.

Note that this approach requires CEs to be, at most, $24 \text{ h} - 2 * f$ long. As described in Section 4.2, the smallest time interval f not leading to interference patterns is 2 min. The following analyses, therefore, filter for CEs shorter than 23 h and 56 min. This filter excludes less than 1% of all recorded CEs.

4.3.2. Occupancy Analysis Results

Figure 8 shows the occupancy for all CPs per power type.

Algorithm 1: Create a histogram raster for a given list of charge events.**Function** IntervalToRaster($f, events$):

Data: f : time interval in minutes, $events$: list of charge events with start and end timestamp

$R \leftarrow \text{Array}[0, \lceil \frac{60 \cdot 24}{f} \rceil]$

for $e_{start}, e_{end} \in events$ **do**

$i_e^{start} \leftarrow \lfloor h(e_{start}) * 60 + \min(e_{start}) + \text{sec}(e_{start}) \div 60 \rfloor \div f$

$i_e^{end} \leftarrow \lfloor h(e_{end}) * 60 + \min(e_{end}) + \text{sec}(e_{end}) \div 60 \rfloor \div f$

if $i_e^{start} < i_e^{end}$ **then**

$R[i_e^{start} : i_e^{end}] \leftarrow R[i_e^{start} : i_e^{end}] + 1$

 /* $R[a : b]$ denotes all elements with index $a \leq i < b$ */

else

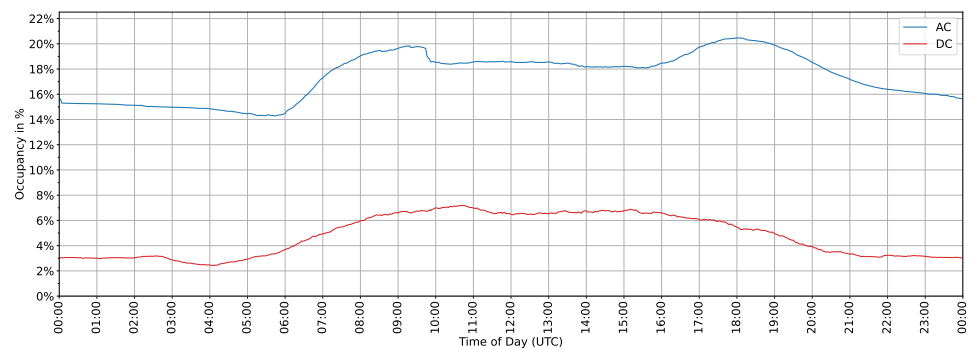
 /* event e crosses midnight */

$R[i_e^{start} :] \leftarrow R[i_e^{start} :] + 1$

$R[: i_e^{end}] \leftarrow R[: i_e^{end}] + 1$

end

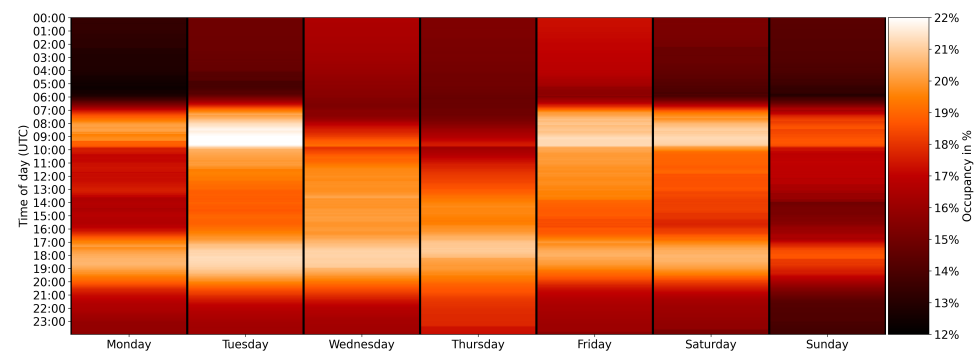
return R

**Figure 8.** Occupancy of all observed CPs, bin size 2 min. Data from 6 May 2021 to 6 October 2021.

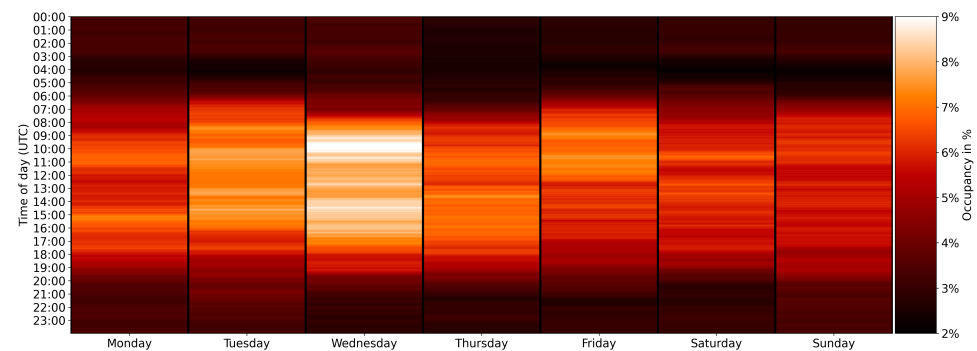
Several insights are provided by this analysis. DC CPs are overall less occupied and primarily used throughout the day. AC CPs display occupancy peaks at 08:00 to 10:00 UTC and 17:00 to 19:00 UTC. In addition, AC CPs are primarily used throughout the day but display a larger occupancy at night compared to DC CPs. Within the observation time frame, local time in Munich is given in CEST, which is UTC+2.

Figure 9 shows the occupancy for all CPs per power type and per weekday. Columns represent the weekdays, and occupancy is indicated by color brightness. The color is normalized by the minimum and maximum of AC and DC occupancy, respectively. Pure black indicates minimum occupancy rather than zero occupancy, while pure white indicates maximum occupancy rather than full occupancy.

This analysis provides further insight into differing occupancy per day of week. DC CPs display minor deviations and a low occupation in general of, at most, 9%. AC CPs display visible deviations in both time-of-day and weekday. Occupancy peaks in the morning and late afternoon, as observed in Figure 8 are visible throughout the week, but are most pronounced on Tuesday and Friday. Sunday displays the least occupancy, and Tuesday, Friday, and Saturday display the most. CEs throughout the night are minimal on Monday and Sunday mornings and maximal on Wednesday and Friday mornings.



(a) Weekday occupancy for AC CPs.



(b) Weekday occupancy for DC CPs.

Figure 9. Occupancy of all observed stations per weekday: (a) shows AC CPs, and (b) shows DC CPs. Bin size is 2 min, and color brightness indicates occupation. Color bar is normalized by minimum and maximum of AC and DC occupancy, respectively. Data from 6 May 2021 to 6 October 2021.

4.4. Extracting Usage Patterns

Based on the occupancy algorithm introduced above, we utilize the occupancy generated for all CPs and extract usage patterns using an agglomerative clustering approach.

4.4.1. Agglomerative Clustering

Agglomerative clustering is an unsupervised, bottom-up clustering algorithm [28]. Initially, each datum is assigned a singleton cluster. In this paper, the occupancy curve from one CP represents one datum. The algorithm then successively merges two clusters with maximum similarity until a termination condition is met.

We measure the similarity of two occupancy curves X and Y using the *correlation* $\text{Corr}(X, Y)$, which is defined as follows:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

$\text{Cov}(X, Y)$ denotes the co-variance between X and Y , and $\sigma(X)$ denotes the standard deviation of X . Occupancy curves are represented as vectors, leading to the following definition

$$\text{Corr}(X, Y) = \frac{(X - \bar{X}) \cdot (Y - \bar{Y})}{\|X - \bar{X}\|_2 * \|Y - \bar{Y}\|_2}.$$

For vector A , \bar{A} denotes the vector mean, $A \cdot B$ denotes the vector dot product, and $\|A\|_2$ denotes the L2 vector norm.

Two occupancy vectors are merged via a weighted average, where the individual cluster weight considers the number of CEs. The clustering algorithm terminates if the maximum correlation falls short of a threshold.

As a result, a list of clusters is yielded, where each cluster comprises the averaged occupancy vectors and a list of all containing CPs. Each cluster represents one usage pattern.

Note that the occupancy vectors used for clustering are normalized by the vector sum instead of the maximum number of possible CEs. The resulting normalized vector can be interpreted as the occupancy density per CP. Normalizing by the maximum number of possible CEs is infeasible for clustering, as it would discriminate between two CPs with identical usage pattern but differing amounts of CEs.

4.4.2. Clustering Results

Figure 10 shows the four largest clusters for AC CPs, clustered with a correlation threshold of 0.8. Several thresholds have been tested, and a threshold of 0.8 was found to yield clusters with highest distinctiveness. The occupancy curves have a resolution of two minutes. Only CPs with at least 10 observed CEs in total are considered, which results in 1909 out of 2210 AC CPs. The four largest clusters contain 1430 out of 1909 (74.9%) AC CPs, for which 242,691 CEs have been recorded. Figure A3 shows each cluster and the individual CP occupancy curves. The remaining clusters which were not merged with the four largest clusters are primarily singleton-clusters and are considered outliers.

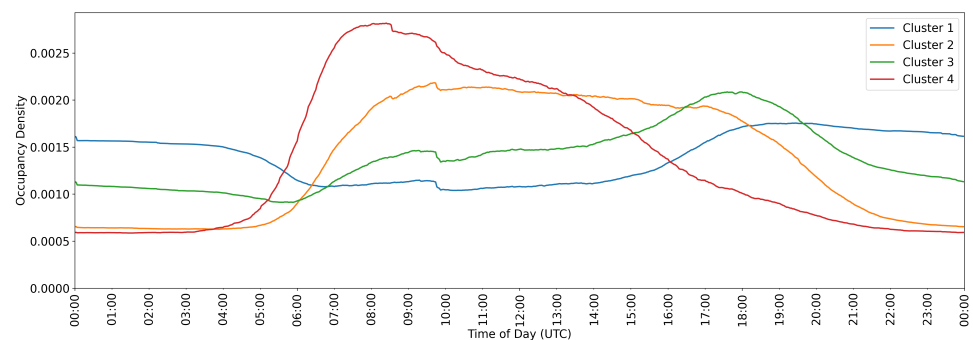


Figure 10. Occupation Density of four largest occupancy clusters for AC CPs. Contains CPs with at least 10 CEs. Similarity threshold is set to 0.8. Occupancy resolution of two minutes. Similarity threshold is set to 0.8. Data from 6 May 2021 to 6 October 2021.

Cluster 1 (blue, 445 CPs) can be interpreted as representing ‘night time’ CP usage, with rising occupancy in the evening hours and roughly constant occupancy over the whole night until 06:00 UTC. Occupancy during the day is low. This utilization pattern requires PEVs to be parked over the night and, hence, could indicate usage of CPs at locations where users spend their night, or where vehicles are parked over the night. While private vehicles tend to be parked near the home location, commercial vehicles are parked either near the home location of the user or near the location of the company, therefore generating such an occupancy curve.

Cluster 2 (orange, 312 CPs) represents ‘day time’ usage, where occupancy is largest between 08:00 UTC and 18:00 UTC. This utilization pattern is the inverse of Cluster 1; hence, PEVs are required to be parked at one spot throughout the day and could, thus, indicate charging during working-hours. Such occupancy curves could emerge in areas with larger amounts of workplaces, such as inner-city areas, business areas, or commercial zones, where employees approach a CP in the morning, attend their job and return to the vehicle after 6–10 h to leave the area.

Cluster 3 (green, 238 CPs) contains CPs primarily used in the evening between 16:00 UTC and 20:00 UTC, whereas utilization over night is low but increases throughout the day. Such areas may comprise locations of evening or night-time activities, such as shopping facilities, restaurants, fitness studios, cinemas, or other recreational facilities.

Finally, Cluster 4 (red, 435 CPs) represents CP usage throughout the day with emphasized usage in the late morning. In comparison to the other clusters, occupancy visibly decreases after 09:00 UTC. Hence, traveling to these CPs is rooted in use cases which occur mostly in the morning but in sum all over the day. Such areas could represent various commercial service locations, where customers travel to over the day and those, where employees have flexible working hours. Further usage of CPs in such areas is expected to arise from business related movements from and towards the areas.

DC CPs are not included in this analysis as there are not enough available CPs, leading to insufficient data for analysis.

5. Usage Patterns and Socio-Demographic Structures

This section extends the occupation analyses and cluster-based usage patterns introduced in Section 4 and analyzes whether the extracted usage patterns are reflected in socio-demographic structures.

5.1. Socio-Demographic Data

The socio-demographic data set used in this paper comprises 32 data points per geographic cell. Figure A2 gives an overview of the cell distribution for the city of Munich and the surrounding suburbs. Table 4 lists the data set schema.

Based on this data set, we use a weighted average approach for deriving socio-demographic data on a CP level. Each CP is assigned a buffer with a radius of 500 m in order to reflect willingness to walk [29] and, therefore, the catchment area of a CP. Next, all geographic cells which are overlapped by the buffer are determined, including the relative overlap area. The socio-demographic data for the CP is determined by the average of all overlapped cells, weighted by the relative overlap area. Figure 11 visualizes this approach. Three cells overlap the 500 m buffer around the CP. The weight of cell 1 (blue) is 50%, and cells 2 (green) and 3 (orange) are assigned weight 25%.

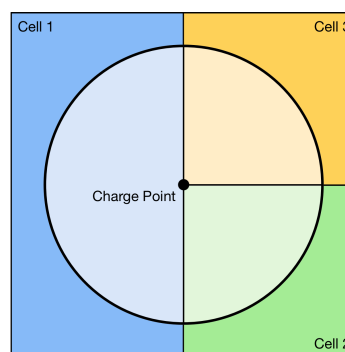


Figure 11. Example of weighted average approach. The center CP is encircled by a 500 m buffer, and cells 1, 2, and 3 are overlapped with a relative overlap area of 50%, 25%, and 25%, respectively (lightened areas).

Each CP is assigned a cluster as discussed in Section 4.4, which represents one of four usage patterns, and the above derived socio-demographic data. Next, we use a random forest classifier to predict the usage patterns based on socio-demographic data.

5.2. Random Forest Classification

Random Forest Classification is an ensemble machine learning technique which combines the classification result of multiple decision trees into a single classification [30]. Each decision tree is trained to a subset of the training data. The results of each decision tree in the ensemble is then merged using a majority vote. Random Forest Classification was chosen as it poses less risks of over-fitting and has reached the highest accuracy as compared to other classification methods, such as Support Vector Machines. The hyper-parameters of the classification model were tuned to yield the highest test accuracy.

Table 4. Schema of socio-demographic data set. Data set comprises data points for geographic cells each comprising residential blocks in urban and suburban areas and larger areas in rural regions. HH refers to households.

Name	Description	Unit
hh_e	HH owning	count
hh_m	HH renting	count
hh_1	single HH	count
hh_2	HH with two inhabitants	count
hh_3m	HH with three or more inhabitants	count
lcgchar_priv	detached or row house	count
lcgchar_priv_mult	multi-story buildings	count
lcgchar_comm	commercial buildings	count
lcgchar_sum	total buildings	count
ew	total population	count
ew_0014	population between 0 and 14 years	count
ew_1524	population between 15 and 24 years	count
ew_2549	population between 25 and 49 years	count
ew_5064	population between 50 and 64 years	count
ew_65	population older than 64 years	count
hh_0029	HH with householder between 0 and 29 years	count
hh_3044	HH with householder between 30 and 44 years	count
hh_4559	HH with householder between 45 and 59 years	count
hh_60m	HH with householder bolder than 59 years	count
hh_tit	HH with titulars	count
hh_ek900	HH with monthly net income < 900 EUR	count
hh_ek1500	HH with monthly net income from 900 EUR to 1.500 EUR	count
hh_ek2600	HH with monthly net income from 1.500 EUR to 2.600 EUR	count
hh_ek3600	HH with monthly net income from 2.600 EUR to 3.600 EUR	count
hh_ek5000	HH with monthly net income from 3.600 EUR to 5.000 EUR	count
hh_ek5000m	HH with monthly net income \geq 5.000 EUR	count
kfz_ges	total motor vehicles	count
pkw_ges	total automobiles	count
pkw_gew	commercial automobiles	count
pkw_priv	private automobiles	count
kk_mio	buying power	EUR
kk_ew	buying power per inhabitant	EUR

5.3. Classification Results

Using the 32 features listed in Table 4 as input and the usage pattern, represented by the cluster ID, as output data, the classification model explains the usage patterns with an overall accuracy of 0.897, although the test accuracy is lower with a score of 0.76.

Out of the 32 data points, we determined 5 features which primarily explain the usage patterns: lcgchar_priv representing the number of detached or row houses, lcgchar_comm representing the number of commercial buildings, lcgchar_sum representing the number of all buildings, hh_ek900 representing the number of households with a monthly net income of, at most, 900 EUR, and kk_ew representing the buying power per inhabitant. Based on these four features, the trained model predicts usage patterns with an overall accuracy of 0.889 and a test accuracy of 0.735.

Both classification models trained on all 32 features, and the 5 primary features have a test precision and recall of ± 0.07 .

5.4. Spatial Distribution of Usage Pattern Predictions

Figure 12 visualizes the spatial distribution of the predicted usage patterns and was created using the following approach: First, a grid of sample points with a spacing of 100 m around the center of Munich was created. For all sample points, socio-demographic data points have been extracted using the above described weighted average approach. Next, the trained model using all 32 features was applied to all sample points to infer the primary

usage pattern and the probability of all four usage patterns. The probabilities for each usage pattern were then interpolated using Triangular Irregular Network (TIN) interpolation. Each usage pattern is displayed as a separate raster layer and colored according to the prediction probability: Intense colors indicate high prediction probability, and vice versa. To avoid overlay effects of the four layers, the color opacity scales linearly from 0% to 100%, starting at the prediction probability of 0.25; interpolated values with a prediction probability of ≤ 0.25 are not displayed.

Several observations can be made. First, the usage pattern 'night-time charging' (Pattern 1, blue) primarily occurs in residential areas, whereas the usage pattern 'day-time charging' (Pattern 2, orange) primarily occurs in the city center. The usage patterns 'evening charging' (Pattern 3, green) and 'morning charging' (Pattern 4, red) occur less frequently and are primarily located between residential and commercial areas.

Second, the probability for all four clusters in general is largest in the city center and decreases towards the outskirts. We expect that this is a reflection of the CP density, which is highest in the city center, leading to more accurate predictions. Figure A4 additionally shows the distribution of CPs and their true usage pattern.

Third, the top right shows a patch with high probability for usage pattern 4, 'morning charging'. This location contains a large charging park with 48 CPs, which is located near a commercial area. All CPs are clustered to usage pattern 4, leading to the observed high probability patch.

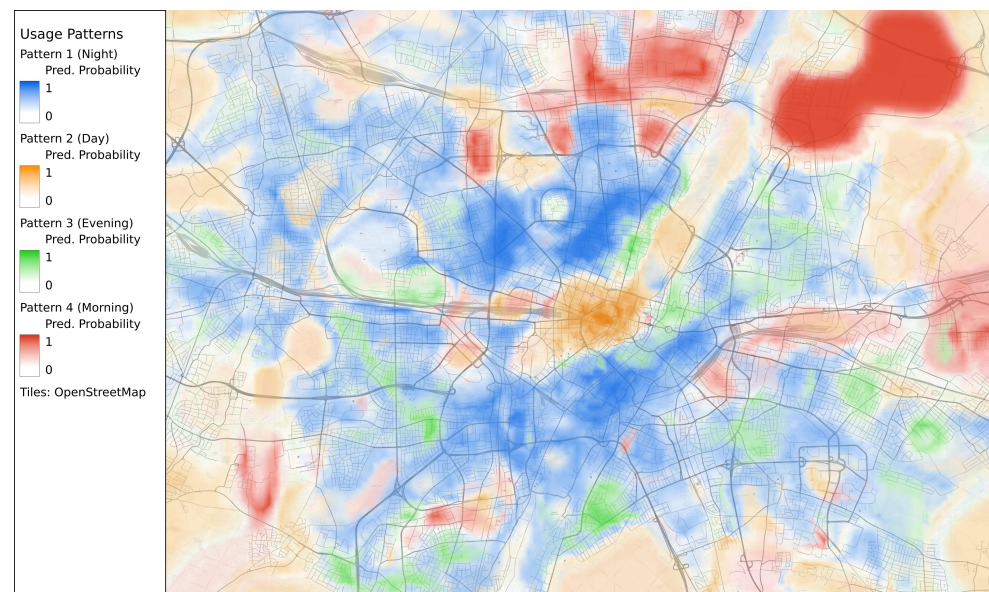


Figure 12. Spatial distribution of predicted usage patterns. Based on interpolation of grid with 100 m spacing. Probabilities of all four clusters at all sample points have been interpolated using TIN interpolation. color intensity indicates prediction probability. Map Projection: EPSG:25832. Background tile credit: OpenStreetMap contributors.

6. Discussion

In this work, we gathered data from a publicly available data on CI status information. This data was gathered with a frequency of 15 min, while precise timestamps of occupancy start and end were derived. The gathered data captures the occupation duration of CPs; the charging duration and transmitted energy are not publicly available and could not be gathered. Applying a data set containing the charging duration and transmitted energy to the analyses presented in this work may yield further insights into CI usage.

From this data, we extracted the occupancy on a CP level and observed an anomalous distribution of CE lengths. We expect this pattern to originate from CPs sampling data every two minutes. Based on this finding, we conclude that the maximum possible resolution for occupancy analysis of the gathered data set is two minutes.

Using the extracted CP occupancy, we then presented a two-step machine learning approach.

First, we employed an agglomerative clustering of the CP occupancy using correlation as a similarity metric. This approach extracts occupancy patterns shared by multiple CPs. While this clustering approach yields reasonable results, it may be valuable to employ other similarity metrics to further reduce outliers. Different clustering approaches may also yield more accurate usage patterns.

Second, we used a random forest classifier to predict usage patterns derived from the clustering from socio-demographic structures. The data set used in this work provides socio-demographic structures per geographic cell. In order to assign this structure data to CPs, we used a weighted averaging approach based on the relative overlap area. In comparison to assigning data from the cell containing a CP, this approach more accurately captures socio-demographic structures in heterogeneous areas with many small cells, such as suburban or urban cities, as neighboring cells are considered. However, further refinements on sampling structural data for CPs, such as different buffer sizes or a different, e.g., non-linear, averaging approach, may lead to further improvements in data accuracy. Using the weighted socio-demographic structures on a CP level, we applied a random forest classification model to predict usage patterns. The train and overall accuracy is reasonable. The test accuracy is a subject for further improvements. A longer observation time frame or observing a larger area may be beneficial for the test performance. The trained model only considers data for CPs in and around Munich, which may not reflect the link between socio-demographic structures and CI usage patterns in other regions. Adding further data of other areas may improve the predictive performance of the model in other regions. Additionally, further attributes concerning the offered charging service agreements at CPs are expected to enhance the performance of the random forest classifier, e.g., by means of tariffs or amount of available offers. Choosing the four largest clusters as a basis for the classification model leads to an exclusion of 25.1% of all AC CPs. Reducing the number of outliers may increase the model accuracy.

Last, we used the trained model to visualize the spatial distribution of predicted usage patterns in and around the center of Munich. This visualization gives insights into the link between local socio-demographic structures and CP usage. However, it also visualizes the link between model accuracy and CP density.

7. Conclusions

With the rise of electric mobility in recent years, the urgent need for reliable information about demand for and usage of CI became apparent. Since there is only a sparse number of publicly available data, researchers have to refrain to data collection techniques to gain an understanding of CI usage. The approach presented here includes such a data gathering, allowing for analysis and further interpretation of such CI. While early research focused on general understanding and adoption of electric vehicles, today, municipalities require more fine-grained data on usage behavior and possible locations for CPs to foster electric mobility in order to achieve emission targets. At the same time, CI operators try to find profitable locations to build and conduct CI, on the one hand, and optimize and encourage usage of existing CI, on the other hand. For a deeper understanding of CI usage, this work gathered and further analyzed data, utilizing the presented two-step approach. In the first step, clusters of four usage patterns were retrieved, allowing for an estimate of usage types. The second step uses a random forest classification based on socio-demographic data to predict the aforementioned usage patterns. Based on these results, it is possible for municipalities, as well as for operators, to estimate the requirements for CI in terms of dimensioning and detailing. Using the presented approach, municipalities have a tool at hand for further development of CI, while operators, on the one hand, can better assess different locations in the cities for further expansion and, on the other hand, adapt their business towards the prevalent usage in the respective area. A third actor to be interested in these results are operators of electric fleets in cities, e.g., carsharing,

ridesharing, or public transport operators, adapting their operational measures or pricing policies towards the optimal usage of public CI, improving usage of such infrastructure. By including the results presented here into the applied pricing policies [31], general fleet utilization can be enhanced, and charging of fleet vehicles can be done in a more economic way [32]. After all, the approach presented here opens the path for further integration of electric mobility and CI towards an optimized utilization and, therefore, less investments in CI needed.

Author Contributions: Conceptualization, C.H. and M.F.; methodology, P.A.F.; software, P.F.; validation, W.M., C.H., and M.F.; formal analysis, P.A.F.; investigation, P.A.F. and W.M.; resources, C.H.; data curation, P.A.F. and W.M.; writing—original draft preparation, P.A.F., W.M., M.F., and C.H.; writing—review and editing, P.A.F., W.M., M.F., and C.H.; visualization, P.A.F.; supervision, K.B.; project administration, C.H. and K.B.; funding acquisition, C.H. and K.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) within the funding initiative “Sofortprogramm Saubere Luft” by the project “München elektrisiert”, grant number: 01MZ18010C .

Data Availability Statement: Both data and software presented in this work are openly available in https://gitlab.lrz.de/philipp.friese/charging_point_paper, (accessed on 19 November 2021), with the exception of the socio-demographic data set, which cannot be shared due to copyright constraints.

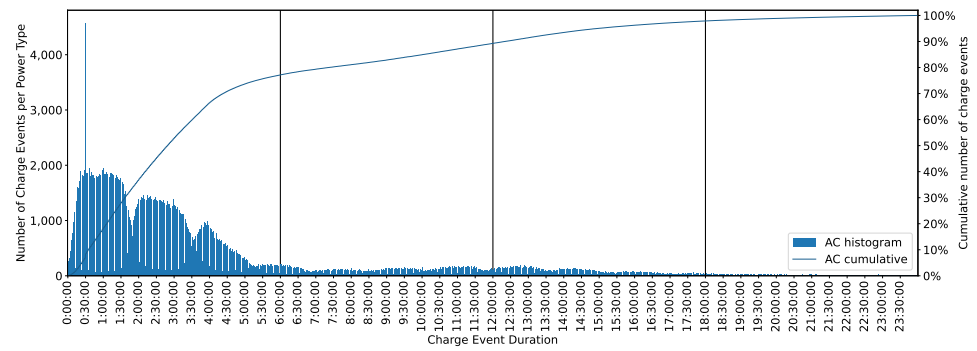
Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

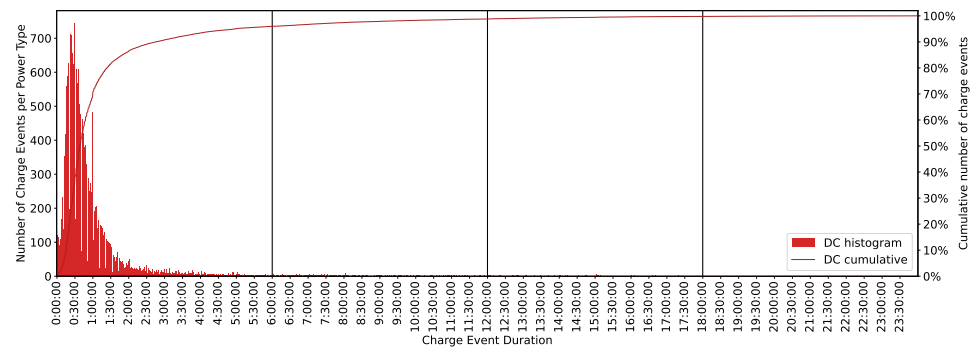
The following abbreviations are used in this manuscript:

CE	Charging Event
CP	Charging Point
CS	Charging Station
EV	Electric Vehicle
EVSE	Electric Vehicle Supply Equipment
EVSE-ID	EVSE Identifier
HH	Household
ICE	Internal Combustion Engine
ISM	Infrastructure Status Map
PEV	Plug-In Electric Vehicle
PHEV	Plug-In Hybrid Electric Vehicle
SOC	State-of-Charge
SWM	Stadtwerke München
TIN	Triangular Irregular Network

Appendix A



(a) Histogram for AC charge points



(b) Histogram for DC charge points

Figure A1. Histogram and cumulative charge interval lengths per power type with bin size 30 s; (a) shows AC charge points, and (b) shows DC charge points. Data from 6 May 2021 to 6 October 2021.

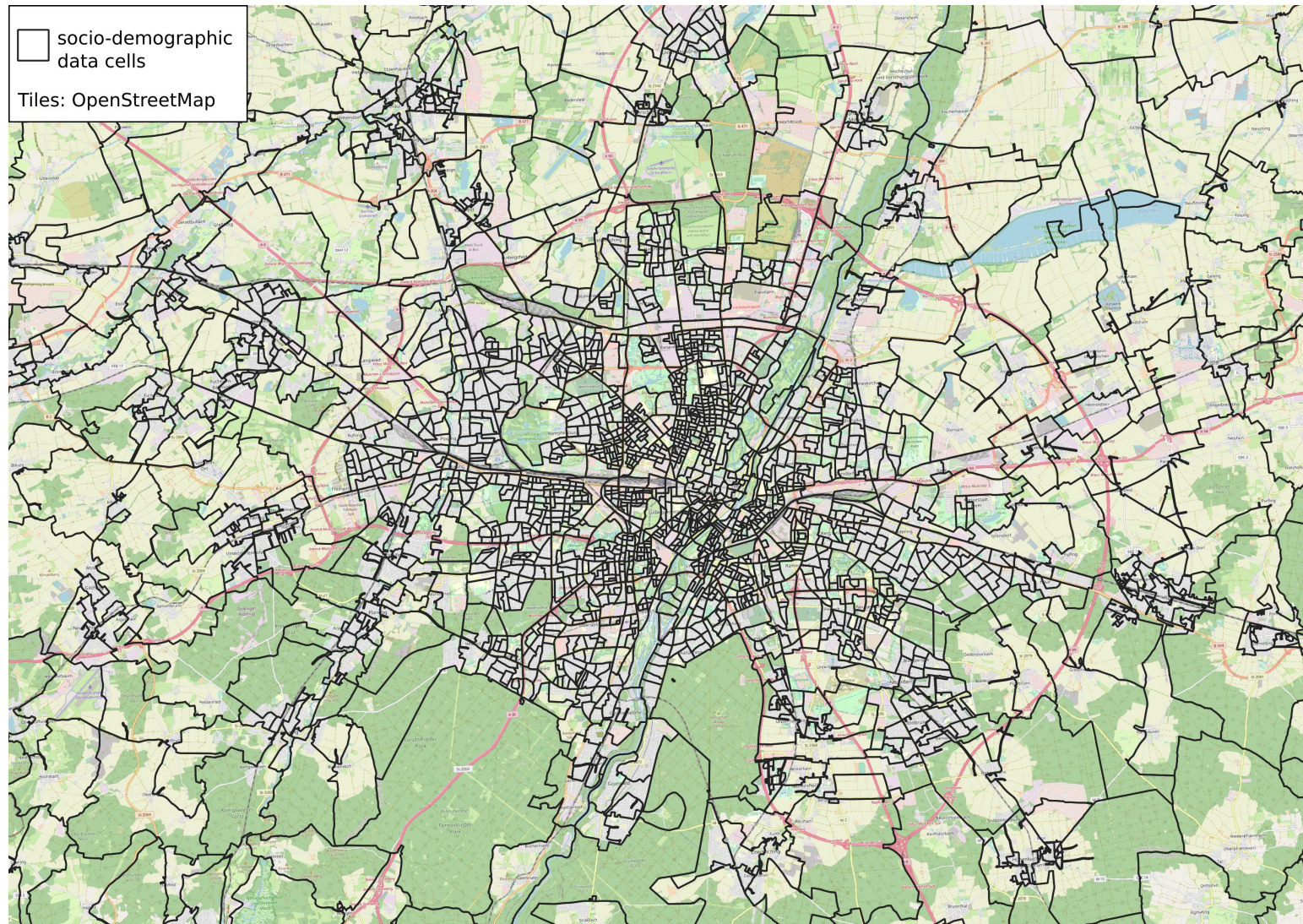


Figure A2. Geographic cells of socio-demographic data set. Background tile credit: OpenStreetMap contributors.

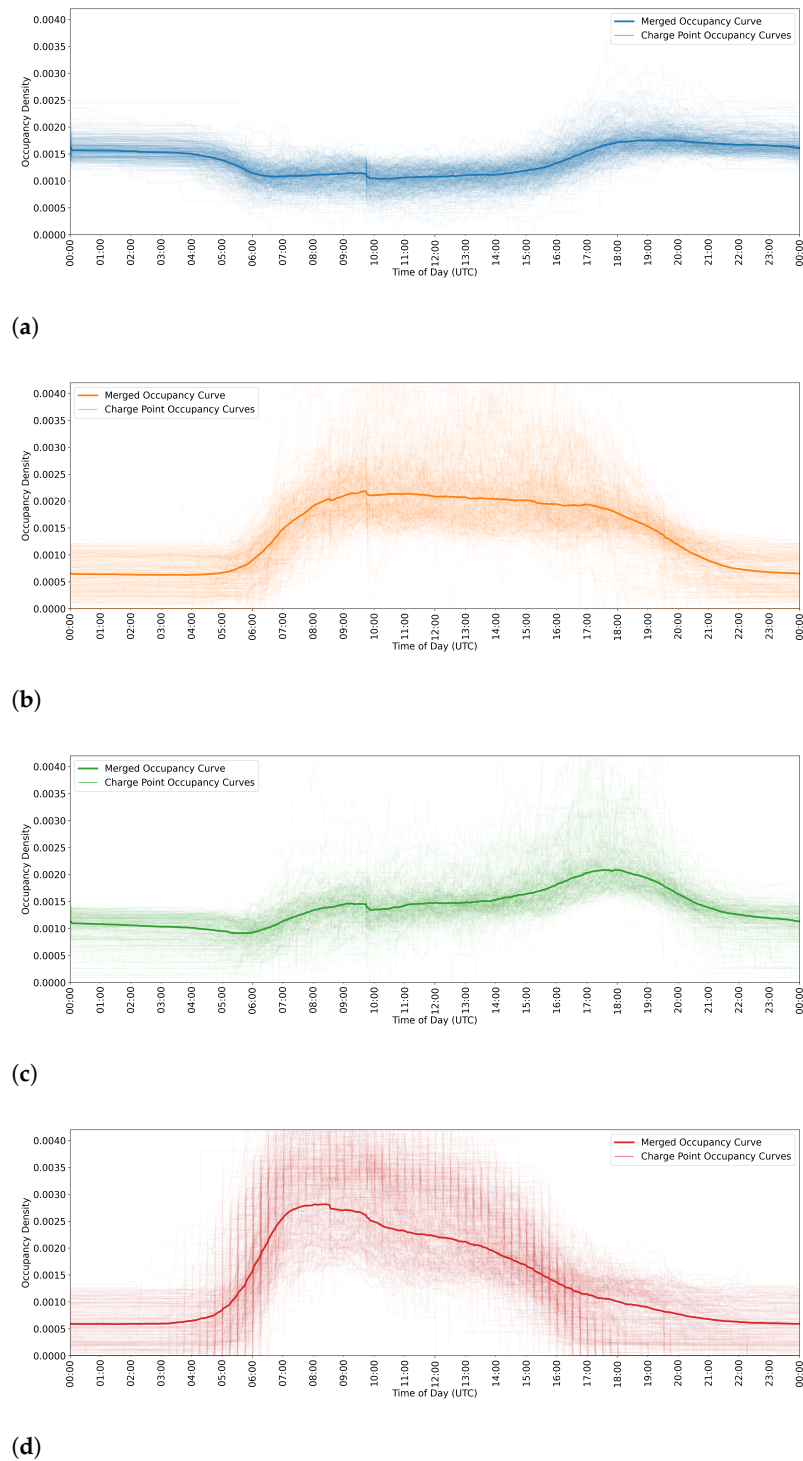


Figure A3. Occupation clusters with merged occupancy curves. Occupancy curves have a resolution of two minutes. Similarity threshold is set to 0.8. Data from 6 May 2021 to 6 October 2021. (a) Cluster 1 containing 445 charge points (23.3%), 92,924 charge events (29.9%); (b) Cluster 2 containing 312 charge points (16.3%), 64,961 charge events (20.9%); (c) Cluster 3 containing 238 charge points (12.5%), 47,897 charge events (15.4%); (d) Cluster 4 containing 435 charge points (22.8%), 36,909 charge events (11.9%).

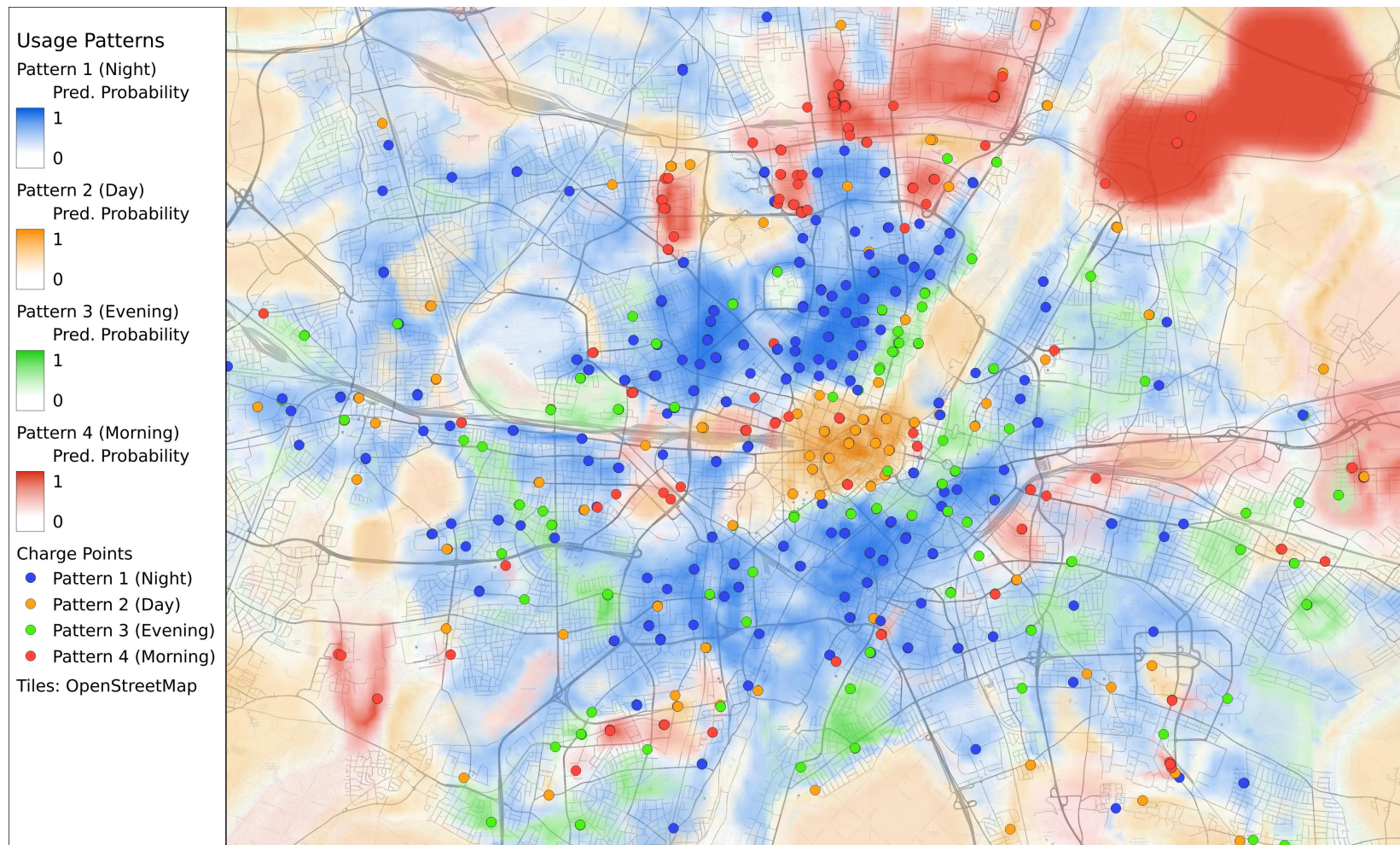


Figure A4. Spatial distribution of predicted usage patterns with true usage pattern per charge points. Based on interpolation of grid with 100 m spacing. Probabilities of all four clusters at all sample points have been interpolated using TIN interpolation. Color intensity indicates prediction probability. Charge point marking indicates true usage pattern. Map Projection: EPSG:25832. Background tile credit: OpenStreetMap contributors

References

1. Kessler, L.; Bogenberger, K. Mobility patterns and charging behavior of BMW i3 customers. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, Rio de Janeiro, Brazil, 1–4 November 2016; pp. 1994–1999. [\[CrossRef\]](#)
2. Krause, J.; Ladwig, S.; Schwalm, M. Better (Not) charge in your garage! perceived benefits and optimal positioning of public fast charging infrastructure for electrical vehicles from user’s perspective. *Adv. Intell. Syst. Comput.* **2018**, *600*, 261–269. [\[CrossRef\]](#)
3. Philipsen, R.; Brell, T.; Brost, W.; Eickels, T.; Ziefle, M. Running on empty—Users’ charging behavior of electric vehicles versus traditional refueling. *Transp. Res. Part Traffic Psychol. Behav.* **2018**, *59*, 475–492. [\[CrossRef\]](#)
4. van den Hoed, R.; Helmus, J.R.; de Vries, R.; Bardok, D. Data analysis on the public charge infrastructure in the city of Amsterdam. In Proceedings of the 2013 World Electric Vehicle Symposium and Exhibition (EVS27), Barcelona, Spain, 17–20 November 2013; pp. 1–10. [\[CrossRef\]](#)
5. Wolbertus, R.; van den Hoed, R.; Maase, S. Benchmarking Charging Infrastructure Utilization. *World Electr. Veh. J.* **2016**, *8*, 754–771. [\[CrossRef\]](#)
6. Wolbertus, R.; Van den Hoed, R. Managing parking pressure concerns related to charging stations for electric vehicles: Data analysis on the case of daytime charging in The Hague. In Proceedings of the European Battery, Hybrid & Fuel Cell Electric Vehicle Congress, Geneva, Switzerland, 14–16 March 2017.
7. Wolbertus, R.; van den Hoed, R. Charging station hogging: A data-driven analysis. In Proceedings of the Electric Vehicle Symposium EVS30, Stuttgart, Germany, 9–11 October 2017.
8. Wolbertus, R.; Kroesen, M.; van den Hoed, R.; Chorus, C. Fully charged: An empirical study into the factors that influence connection times at EV-charging stations. *Energy Policy* **2018**, *123*, 1–7. [\[CrossRef\]](#)
9. Gerritsma, M.K.; AlSkaif, T.A.; Fidler, H.A.; van Sark, W.G. Flexibility of Electric Vehicle Demand: Analysis of Measured Charging Data and Simulation for the Future. *World Electr. Veh. J.* **2019**, *10*, 14. [\[CrossRef\]](#)
10. Almaghrebi, A.; Shom, S.; Al Juheshi, F.; James, K.; Alahmad, M. Analysis of User Charging Behavior at Public Charging Stations. In Proceedings of the 2019 IEEE Transportation Electrification Conference and Expo (ITEC), Detroit, MI, USA, 19–21 June 2019; pp. 1–6. [\[CrossRef\]](#)
11. Olk, C.; Trunschke, M.; Bussar, C.; Sauer, D.U. Empirical Study of Electric Vehicle Charging Infrastructure Usage in Ireland. In Proceedings of the 2019 4th IEEE Workshop on the Electronic Grid (eGRID), Xiamen, China, 11–14 November 2019; pp. 1–8. [\[CrossRef\]](#)
12. Almaghrebi, A.; Juheshi, F.A.; Nekl, J.; James, K.; Alahmad, M. Analysis of Energy Consumption at Public Charging Stations, a Nebraska Case Study. In Proceedings of the 2020 IEEE Transportation Electrification Conference & Expo (ITEC), Chicago, IL, USA, 23–26 June 2020; pp. 1–6. [\[CrossRef\]](#)
13. Almaghrebi, A.; Aljuheshi, F.; Rifaie, M.; James, K.; Alahmad, M. Data-Driven Charging Demand Prediction at Public Charging Stations Using Supervised Machine Learning Regression Methods. *Energies* **2020**, *13*, 4231. [\[CrossRef\]](#)
14. van der Kam, M.; van Sark, W.; Alkemade, F. Multiple roads ahead: How charging behavior can guide charging infrastructure roll-out policy. *Transp. Res. Part Transp. Environ.* **2020**, *85*, 102452. [\[CrossRef\]](#)
15. Hecht, C.; Das, S.; Bussar, C.; Sauer, D.U. Representative, empirical, real-world charging station usage characteristics and data in Germany. *eTransportation* **2020**, *6*, 100079. [\[CrossRef\]](#)
16. Fischer, M.; Hardt, C.; Michalk, W.; Bogenberger, K. Charging or Idling: Method for quantifying the Charging and the Idle Time of public Charging Stations. In *TRB 101st Annual Meeting Compendium of Papers*; Transportation Research Board: Washington, DC, USA, 2022.
17. Anderson, J.E.; Böttcher, N.; Kuhnimhof, T. An Approach To Determine Charging Infrastructure for One Million Electric Vehicles in Germany. In *TRB 95th Annual Meeting Compendium of Papers*; Transportation Research Board: Washington, DC, USA, 2016; pp. 1–22.
18. Adenaw, L.; Lienkamp, M. Multi-Criteria, Co-Evolutionary Charging Behavior: An Agent-Based Simulation of Urban Electromobility. *World Electr. Veh. J.* **2021**, *12*, 18. [\[CrossRef\]](#)
19. Kessler, L.; Bogenberger, K. Dynamic traffic information for electric vehicles as a basis for energy-efficient routing. *Transp. Res. Procedia* **2019**, *37*, 457–464. [\[CrossRef\]](#)
20. Fischer, M.; Elias, J.; Bogenberger, K.; Schramm, S. Quantification of transmitted energy and power for system services by battery electric vehicles based on real mobility and charging profiles. In Proceedings of the CIRED Berlin 2020 Workshop How to Implement Flexibility in the Distribution System? Online Conference, 22–23 September 2020. [\[CrossRef\]](#)
21. Fischer, M.; Bogenberger, K.; Elias, J.; Schramm, S. Evaluation of charging concepts for high density urban areas based on real mobility and charging profiles of BEV. In Proceedings of the CIRED Geneva 2021 Conference, Online Conference, 20–23 September 2021.
22. Chakraborty, D.; Bunch, D.S.; Lee, J.H.; Tal, G. Demand drivers for charging infrastructure-charging behavior of plug-in electric vehicle commuters. *Transp. Res. Part D Transp. Environ.* **2019**, *76*, 255–272. [\[CrossRef\]](#)
23. Andrenacci, N.; Ragona, R.; Valenti, G. A demand-side approach to the optimal deployment of electric vehicle charging stations in metropolitan areas. *Appl. Energy* **2016**, *182*, 39–46. [\[CrossRef\]](#)

24. Betz, J.; Walther, L.; Lienkamp, M. Analysis of the charging infrastructure for battery electric vehicles in commercial companies. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 1643–1649. [[CrossRef](#)]
25. Morrissey, P.; Weldon, P.; O'Mahony, M. Future standard and fast charging infrastructure planning: An analysis of electric vehicle charging behaviour. *Energy Policy* **2016**, *89*, 257–270. [[CrossRef](#)]
26. Flammini, M.G.; Prettico, G.; Julea, A.; Fulli, G.; Mazza, A.; Chicco, G. Statistical characterisation of the real transaction data gathered from electric vehicle charging stations. *Electr. Power Syst. Res.* **2019**, *166*, 136–150. [[CrossRef](#)]
27. Open Charge Alliance. *Open Charge Point Protocol*, 2.0.1 ed.; Open Charge Alliance: Arnhem, The Netherlands, 2020.
28. Maimon, O.; Rokach, L. (Eds.). *Data Mining and Knowledge Discovery Handbook*; Springer: New York, NY, USA, 2010. [[CrossRef](#)]
29. Herrmann, S.; Schulte, F.; Voß, S. Increasing Acceptance of Free-Floating Car Sharing Systems Using Smart Relocation Strategies: A Survey Based Study of car2go Hamburg. In *Computational Logistics*; González-Ramírez, R.G., Schulte, F., Voß, S., Ceroni Díaz, J.A., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; Volume 8760, pp. 151–162. doi:10.1007/978-3-319-11421-7_10. [[CrossRef](#)]
30. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
31. Hardt, C.; Bogenberger, K. Dynamic Pricing in Free-Floating Carsharing Systems—A Model Predictive Control Approach. In *TRB 100th Annual Meeting Compendium of Papers*; Transportation Research Board: Washington, DC, USA, 2021; p. 19.
32. Dandl, F.; Fehn, F.; Bogenberger, K.; Busch, F. Pre-Day Scheduling of Charging Processes in Mobility-on-Demand Systems Considering Electricity Price and Vehicle Utilization Forecasts. In Proceedings of the 2020 Forum on Integrated and Sustainable Transportation Systems (FISTS), Delft, The Netherlands, 3–5 November 2020; pp. 127–134.