Open-data methodology for optimizing the allocations of Charging Stations

Hana Elattar, Ferdinand von Tüllenburg, Sebastian Wöllmann, Javier Valdes
Deggendorf Institute of Technology
Institute of Applied Computer Science
Technologiecampus Freyung, Grafenauer Str. 22, 94078 Freyung, Germany
{hana.elattar, ferdinand.tuellenburg, sebastian.woellmann, javier.valdes}@th-deg.de

Executive Summary

This paper explains the work that started towards creating a tool for optimizing the placement of charging stations. Relying first and foremost on open-source data, the methodology of this paper attempts, as a first step, to forecast the number of additional charging stations needed in the area of study. During the work, we were able to extract the necessary OpenStreetMap data that lead us to estimate, with the help of statistical data, the number of electric vehicles reaching a certain destination during a interval. Overlapping these results with data on the location of charging stations and their classifications, we are able to reach the percentage to which the existing charging stations are fulfilling the demand for electric vehicles and hence the number of additional charging stations required.

Keywords: Infrastructure, Charging, Electric vehicle (EV), Modelling, Research

1 Introduction

In recent years, the distribution of electric vehicles (EVs) has significantly increased. More and more people switch from combustion vehicles (CVs) and replace them, especially with battery-driven electric vehicles (BEVs). Together with the increasing availability of green energy, this development leads to a positive effect on the greenhouse gas balance of our economies. On the other side of the coin, charging infrastructure (CI) is being increasingly requested and utilised. Following the initial spread of publicly available charging stations (CSS) across the countries in order to decrease the hurdles for the shift to more sustainable transportation, now the question arises, at which locations the CI should be strengthened in order to satisfy the needs of the EV users.

In recent years, there have been several research approaches and studies to evaluate the energy demand and market development of electric vehicles including extrapolations into the nearer future either for particular cities [3] or for whole countries [4]. Other studies answered the question about the current utilisation of CI [1]. In order to estimate arrivals at certain locations it has been proposed to analyse the popularity index (as known, e.g., from Google maps), showing the occupation of certain locations by time. From these studies, we know in rather general terms how CI needs to be developed in the upcoming years.

With this information, more targeted investments into CI would be possible taking the actual demands of EV drivers into its focus. On the one hand, it needs to be answered, if visitors of a certain place find sufficient charging possibilities for their EVs within walking distance. On the other hand, it needs to be considered, if it makes sense for EV users to charge their vehicle at a given location, which basically is related to the time, that people typically stay at a given location. The approaches, which analyse traffic
situations at particular locations in order to estimate arrivals do not take into consideration typical stay times.

Our research objective has been to develop an approach to estimate the demand for charging within walking distance of particular locations or buildings. In order to count how many CSs should be available within walking distance of a particular building, we determine the purpose of the building, which gives an indication of typical stay times and the opening hours of a building. Additionally, we consider the size of the building, which gives, together with the building/location purpose, an estimation of the number of people that arrive at the building based on statistical data and building regulations. For the analysis, we limit ourselves to publicly available OpenStreetMap (OSM) data to ensure that the approach is transferable to arbitrary regions and we also consider if buildings have multiple purposes, for instance, if an apartment building has a shop floor. We feel that our approach can be well used as an extension to already existing approaches, which take traffic situations and car arrivals at particular locations into their consideration.

However, an open challenge is to precisely estimate the need for CI around a particular location. This estimation usually depends on several factors including the number of EVs reaching a certain destination at the same time, the state of charge (SOC) of the EVs reaching the destination, the capacity of existing CSs etc. In this paper, we will focus on the first factor while assuming a minimal SOC which will limit the location radius to 300 m around CSs, based on the work by Hummler et al. [13]. With this focus and referencing the two previously mentioned approaches, we come up with a 3-layer approach that considers: Number of EVs (Layer 1) - Time Factor (Layer 2) - The capacity of existing CS (Layer 3). This approach will help us in calculating the fulfilment rate of existing CSs and hence the number of CSs needed in the area. Additionally, based on forecast data on the use of EVs, it will allow us to predict the fulfilment rate in the upcoming years.

In the paper, we will explain the general methodology in Section 2, Section 3 will explain in detail the data collected and integrated, and in section 5 we will go through the results of the current model, while expected future results from the development of the model will be elaborated in section 6.

2 Methodology

It is our goal to develop a modelling methodology that can be applied to different spatial contexts and locations, where the case study conducted here is in Germany, the methodology could be applied directly to all cities in Germany, but would require small parameters’ alterations if deployed in a different European city and even more so if deployed outside of Europe. The aim is to define the parameters and framework for an optimal allocation of new CS, by identifying the number of new installations required to fill the traffic electricity demand gap in different areas. The methodology is inspired by the overlapping of the work of Klinkerhardt et al. [2] and that of Hecht et al. [1]. In the first approach, attraction factors of buildings could be used to determine the traffic flow arriving from and to certain points in a city. The approach uses OSM building data as a base and collects the attractiveness factors from different attractiveness tables including those shared by businesses. Whereas the work of Hecht et al. provides a calculation of the usage percentage of CS based on their location (e.g. urban, rural, etc.). In a conclusion, we can identify the hourly occupation rate through the location of a CS. Our methodology proposes an approach where the previous two are combined with daily time intervals and statistical data to identify the required CS needed to fulfil the demand of different areas. Using classified OSM buildings and point of interest (POI) data, we can provide a basic estimation of the number of people visiting a certain building during a chosen time interval. This data is then overlapped with open data on CS which can be categorized based on the study from [1] and therefore have an average percentage of occupancy rate during these same hours. Further information on the methodology development can be found in our GISTAM conference paper [19].

3 Data

The main challenges in traffic and CI modelling is identifying and obtaining the required data and developing modelling methodologies that can be applied at arbitrary regions under investigation (RUIs). This is particularly important as different data sets in varying detail grade are available for different regions. The path for our research includes the model development based on available OSM data including all available parameters for our selected RUI where we find a outstanding high data accuracy. The resulting model is, in turn, of likewise high quality. In future steps we plan to evaluate our model by removing certain parameters or diminishing their accuracy to get insights into the change of the overall model precision (this evaluation is future work and not part of this article). This information can finally be used when applying the model to other RUIs.
3.1 OpenStreetMap Data

If we focus on the work done in Europe and try to clarify what are the current approaches to model traffic or the use of charging stations while relying on open data, we will see that the main advances in this area are carried out through the use of POIs. POIs are used in traffic modelling as they are indicating probable travel destinations of people, be it by foot or transport (public or private). For the identification of POIs, there are currently three main approaches: the use of social sensing [7], digital global navigation satellite system (GNSS) data generated by users, and the use of OSM that contains for this area very detailed and complete data [8, 9]. Along with this source of open data highlights the use of Google Maps Analytics (see [5] for a recent application) and the use of Facebook Analytics that also provides information on location and use of POIs [6].

Our work is inspired by Pagany et al. [9] in which they used OSM information to identify building uses and derive dwell time. To calculate dwell time they used official statistics on time use in different activities published at the national level in conducted case studies in Germany and the Czech Republic. By mapping the activities with information on POIs the authors estimated the approximate amount of dwell time. Hummler et al. [13] are also inspired by the use of distance from a POI to a CS as an element able to explain the attractiveness and usage factor of a CS. They build a probabilistic model that takes this relationship into account, and their results show that there are POIs with greater attractiveness than others. These differences in the powers of attraction are, although derived by other methods, also considered different in other studies for instance by Valdes et al. [8] and Pagany et al. [9].

OSM is one of the largest sources of open-source data and the most-known voluntary map worldwide. The choice for using OSM for our work stems from the flexibility the platform provides in terms of data acquisition, enhancement, and validation. As explained above, we aim to create a methodology which uses transparent and open parameters that can be shifted if needed. Using OSM guarantees the geographical data parameter having the same source and format worldwide. Indeed the quality of OSM data is different based on the country - for the number of volunteers in the area, the local state of technological development, etc - it still provides us with the possibility of personally fixing and enhancing the geographical data in whichever study area. With this focus, it was important to download all existing buildings and POIs in the area of the city of Lindau at Lake Constance in Bavaria, Germany. The data was downloaded via the OSM Application Programming Interface (API) which provides a set of save functions, application programmers can use for building customised applications based on data provided by OSM. This allows for more flexibility and precision during the filtration of the keys and tags assigned in the area. In our interpretation of the data, we focus on two main aspects: the main use and the floor area of the building. We consider these locations (polygons and POIs) as travel destinations.

3.1.1 Framework for OSM data classification

These are problems already highlighted in both Valdes et al. [8] and Klinkhardt et al. [2], that lie in how a given POI is represented in OSM by users. For example, a supermarket can be entered as a point or a polygon. If it is a point, it can stand alone or in combination with other POIs on or outside the same building. In the case of a polygon, it can be a single polygon or be composed of several polygons representing the same POI, but it may or may not be labelled as a single POI. The API combined with our data pipeline allows us the freedom of classifying the data based on our needs. This allowed us to focus in this research on publicly used areas as destination points and overlook residential buildings until the next development phase. It also gives us the freedom of checking and re-working the tags if the spatial context demands it.

The steps of our pipeline for OSM downloading, processing, aggregation and classification are depicted in Figure 1. Our approach is founded on PyQGIS, an API for the open-source Geographic Information System (GIS) application QGIS 1. Written in the Python programming language, the pipeline includes multiple independent steps that can be executed as needed. The different steps are:

- Via a Command line interface (CLI), the pipeline depends on two input parameters: the RUI, which should be downloaded, and a save path on the file system, where the results should be saved to.
- Next, the module Downloader tries to get the OSM-ID of the given area and divides it in smaller areas, so the API calls do not time out while downloading.
- Then, the actual data is downloaded via the Overpass API2, which guarantees a current and raw data set.
- Next, the raw .osm file is converted to an Esri-type Shapefile. A shapefile is vectorised standard data format that can be digested by GIS applications.

1https://www.qgis.org
2https://overpass-api.de/
input
OSM area name
and save path via
CLI
Downloader
Divide in subareas
and download via
Overpass API,
convert .osm to .shp files
LayerMerger
Merge .shp files of
subregions into one
file
LayerCleaner
>Delete empty fields,
delete duplicated
geometries and fix
corrupted geometries
output
Shapefile
TrafficChargingStationAnalyzer Calculates usage of CS
FieldCalculator Calculate custom fields with our methodology

Figure 1: Seven steps of the data processing pipeline towards providing a sound Shapefile for further analysis.

• Going forward, the LayerMerger and LayerCleaner modules first combine all subareas into one file again and then delete duplicated or corrupted geometries. Also, fields which only include null values (meaning 'no data'), get deleted.

• The last step is handled by the module FieldCalculator, which implements the methodology described below this item list and calculates a measure, travel purpose and attractiveness factor for each building.

• The result is then saved in the given save directory as in Esri-type Shapefile format.

The most critical phase and where our main contribution lies are the FieldCalculator. In this phase, the tags associated with the OSM data are extracted. These tags are highly varying and are not always in the same order. Because of this heterogeneity, previous studies have focused on OSM data that have already been structured around feature classes (f-classes). In machine learning, a feature is a specific property of an object and feature classes are a collection of features describing sets of similar objects. This in our case, translates into an "fclass" being the main use of the building (other than residential).

In the case of Valdes et al. [8], f-classes are derived from the Grofabrik database and aggregated, in this study we use a similar methodology, but with defining the f-classes ourselves based on the OSM feature tags from the OSM API. We do this in order to apprehend details that pre-defined f-classes from other sources do not necessarily take into account and that are specific to traffic models. This leads us to generate a mapping of tags into f-classes that are specific to the traffic model we are developing and that could again be modified if the spatial context requires (i.e. adding new fclasses or removing country-specific ones).

An additional problem has to do with data coverage. In this direction, there is enough empirical evidence about the good coverage of OSM data in Europe that no validation has been done.

3.2 Google Maps statistics

Attractiveness factor of buildings or destination attractiveness signifies the factor relating to the number of trips arriving at said destinations. When investigating the attractiveness factors of different buildings and POIs, several sources have been identified. In the work of Klinkerhardt et al. [2] they used official attractiveness factor tables for public places in Australia shared by the governments. Another methodology they used was to identify the number of visitors to places such as restaurants, shopping malls, etc. by simply checking the official websites in some cases or initiating contact in others. As explained in the previous sections, our methodology’s purpose is to remain open and adaptable. For this reason, a shift from direct attractiveness factors to a calculation of visitor rate- and with which number of vehicles- was decided. Building on the concept of attractiveness factor, we consider in this paper the visits statistics attached to public spaces and POIs in Google Maps. Taking the city of Lindau as an example, and focusing as a first step in our research on restaurants, cafes, and hotels falling in a 400 m radius from public CS, we can view the percentage of visitors arriving at these locations during different hours of the day; with 100% being the maximum number of visitors that the place could receive. As presented in the graph in Figure 2, we can view the occupancy rate of a fast food restaurant in Lindau throughout the week with an hourly interval. However the disadvantage is that this statistical data does not exist for all public spaces, it does give an indicator of the patterns within the city. From the research of Sparks et al. [6],
we get an understanding of how temporal signatures for different POIs differ based on the spatial context [6]. These patterns are affected by the cultures inhabiting the city, which religion is dominating, and the different times of the year (i.e. seasons, periods of cultural or religious significance - as is the case for the Christmas season - etc.). In our current research, we focus merely on testing our methodology in the span of a normal time period while focusing on weekdays which are usually more stable and predictable than weekends.

Figure 2: Average Occupancy percentage of a McDonald’s, Lindau at Lake Constance throughout the week

3.3 Charging Stations Data

Data on charging stations are widely available, with virtually all public or semi-public charging stations offering this data to their users in real-time. However, historical data on their use are not so readily available, in fact, many of the data are not shared by the researchers who have collected them, nor do statistical offices collect these data in a systematic way. Calero et al. [11] and Firese et al. [10] give us an idea of the difficulty that this represents within a research project since there are no automated means to do this in a systematic way. Hecht et al. [1] have made part of the data public for Germany and have been used intensively for different applications related to the use of CS. Mortimer et al. [12] recently used the data to verify POI attraction models, and their results confirm only that the existence or not of POIs is not related to the actual installation of CS.

In this research, we focus on two types of information concerning the charging stations in Lindau: geographical and statistical. That geographical information refers to the location of charging stations; the statistical information refers to the statistical data on the usage rate of charging stations. The latter is based on the study by Hecht et al. [1], classifying the CSs based on location and assigning usage rates to each class. Considering our area of study, we will focus on the data for urban-class CSs.

From OSM and other public sources, we acquire locations and other parameters such as charging power, accessibility, and opening hours of the charging stations currently available in the RUI. However, this data does not contain information on their typical utilisation which can be described mainly by the typical number of charging events per day. As a starting point, we use the classification as proposed by Hecht et al. [1], who classify charging stations according to the location and the nominal charging power. Location classes are “industrial”, “urban”, “suburban”, and “uninhabited”, while classes for the charging power are shown in Table 1. The authors provide utilisation profiles for each of these classes by weekday. For our analysis, we use the profiles of “urban”-classed charging stations with 22 kW which is CS class 3 in the provided table. To improve our modelling results we use a correcting factor according to the building classification and the typical stay time.

4 Methodology Application

In this section the results of a method applied in the case of the study for the city of Lindau at Lake Constance, Bavaria, Germany is provided. We aim with this application to test our methodology and try out the different parameters of the method. The choice of Lindau (at Lake Constance) was based on the work with our partners in the CrossChargePoint project, who have calculated the energy demand of EV in the city and which we could use for validation at a later step of development. Another supportive
Table 1: Overview of the five CS charging power classes as defined by [1]

<table>
<thead>
<tr>
<th>CS class</th>
<th>Charging power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$P \leq 4, \text{kW}$</td>
</tr>
<tr>
<td>2</td>
<td>$4, \text{kW} &lt; P \leq 12, \text{kW}$</td>
</tr>
<tr>
<td>3</td>
<td>$12, \text{kW} &lt; P \leq 25, \text{kW}$</td>
</tr>
<tr>
<td>4</td>
<td>$25, \text{kW} &lt; P \leq 100, \text{kW}$</td>
</tr>
<tr>
<td>5</td>
<td>$P &gt; 100, \text{kW}$</td>
</tr>
</tbody>
</table>

Figure 3: Lindau’s buildings classification based on the OSM classification framework
The document was the report from the project of expansion of Lindau Park, that helped give a better overview of the city and the methods it uses for calculating both building capacities and vehicles usage [14]. In the first step, the number of visitors is estimated before the CI utilisation is evaluated. The area of Lindau (at Lake Constance) is 33 sq.km, divided between the island and the mainland. As shown in Fig 3, we are focusing in this work on the old town area (the island) and a slit of the mainland area which is more central and occupied than the rest of the mainland. We made a point of including the highest number of existing CS with commercial surrounding in our area, to compare results. The overall sequence of application of the methodology is as viewed in Figure 4.

4.1 Calculating the number of EV

Estimating the number of EV is a difficult task since there is no open-data on the actual use of the spaces, nor is there any systematic data collection on the use of private spaces. In addition, there is no open-data on daily traffic inside the city to estimate the number of visitors to different locations. To analyze the number of visitors it is possible to carry out different techniques. The first one is to measure traffic usage directly or by using surveys. However, these results are not generalisable, so we discard them for our analysis. A possible approximation can be based on the maximum and minimum number of visitors associated with a location, for example by using square meters and maximum capacity rules. As explained in section 3, we were able to obtain both the buildings floor areas and the average number of visitors per hour. In the first step, OSM data on buildings and POIs are categorized with the help of the pipeline explained in section 3.1 above to eliminate and regroup the exiting tags attributed to each building. A result of this step can be viewed on Fig 3. This allows us then to have a specified list of usages distributed over the buildings. Following back to the Google Maps usage statistics, we are able to identify the maximum number of visitors ($V_{max}$) based on the maximum building capacity ($Bc$) identified for each building use. This information was easily extracted from architectural books (e.g. [16]); and attributed to our usage lists in the pipeline. We arrive with this to the maximum number of visitors that can be multiplied by the percentages in the Google statistics ($V_{avg}$ % / h) to get an average number of visitors for each hour as shown in the equation below. Theory for traffic modelling we can regroup each of the OSM buildings into different travel purposes that would grant them specific attractiveness factors. Using QGIS to calculate the building area we could have an estimation of the maximum number of visitors per day and hence during the span of our studied interval.

\[ V_{max} = Bc \times \text{area} \]  
\[ V = (V_{avg} \%) \times V_{max} \]
For the estimation of the number of e-vehicles (BEV and PHEV) in Lindau we take general statistical values for Germany into consideration. Official statistics indicate a market share of 2.6% of BEV (1.3%) and PHEV (1.3%) of the total number of vehicles in 2022 [18]. For the development of the market share until 2030 several studies have been conducted, most of them aiming at a total market share of BEV and PHEV between one-fourth [17] and one-third [15]. For 2025, Borscheid [17] assumes a stock share of about 11%. Thus, for our study, we assume a stock share of BEV and PHEV vehicles of 11% for 2025. For 2030, we assume 30% stock share. Assuming all visitors reach the destination in means of private vehicles, we can multiply the stocks percentages with our number of visitors \( V \) to obtain the number of EV reaching the destination.

### 4.2 Overlapping Charging stations

![Figure 5: Adding a buffer of 300m around existing CS in the city](image)

In this step, we start to add on our data-set the layer of CS, to put them in question. As explained by Hummler et al., there is usually a maximum distance a visitor of a POI is willing to walk between where they would charge the vehicle and their destination POI [13]. This distance depends on mostly on the type of POI which in this study translates into the building use. We therefore have created a buffer around exiting CS to limit the number of surrounding buildings, and with it the number of potential CS users. Based on the observation that the CS in Lindau are usually surrounded by at least one restaurant, we set up the buffer of 300m as recommended by Hummler et al [13]. Both of the CS chosen for the study share the utilisation profile of urban-class CS identified in section 3, with a usability rate of 20% of their working times.

### 5 Results

In theory, such a methodology can be applied to the whole city of Lindau at Lake Constance or any other city. We chose, however, to compile first calculations on the buildings with the largest \( Bc \) and which simultaneously have a CS in its surrounding 300m radius. This filtration allowed us to conduct a first test on the microscopic scale through comparing the utilisation rate of the CS with the demand produced by the number of EV reaching the destination in a 3-hourly interval. Our results are then defined as rough estimations of how many charging stations are needed additionally in the neighbouring areas of the public buildings- namely the ones containing already one neighbouring public CS. The results are, as presented in Table 2 below, the estimation of the fulfillment rate of exiting CS now and in the year
2030 according to the EV stocks share forecast. It is easy then to identify the gap in demand fulfillment and provide a better indicator to where new CS could be needed.

Table 2: Percentage of the fulfillment of charging demand

<table>
<thead>
<tr>
<th></th>
<th>Current percentage</th>
<th>Forecast for the year 2030</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building A</td>
<td>20 %</td>
<td>8 %</td>
</tr>
<tr>
<td>Building B</td>
<td>50 %</td>
<td>20 %</td>
</tr>
</tbody>
</table>

6 Conclusion and Future Work

The work presented here has several contributions to the state of the art. First of all, it is developed using raw data from OSM instead of data that have already been classified into certain classes, and may not adequately represent the needs of a transport model. The OSM data for Germany has a high level of completeness and due to its current spread to other countries and widespread use, we expect other regions to reach similar completeness figures in the coming years. However, the tagging and mapping work has been carried out only for the case study region and therefore tags are used in a special and intensive way in this region and their use may not be uniform in the OSM data.

On the other hand, the way in which the OSM data are completed is not the same over all dimensions. In the case of tags associated with the number of floors, which is directly related to our calculation of the number of visitors, the lack of this data affects our calculations. An additional problem is an area associated with the square meters of a specific activity. This information is very difficult to obtain, using the current way of collecting OSM data. However, even if these data are not available, it is possible that models of attraction and visitor number estimates are associated with a range of uncertainty. These levels of uncertainty may be associated with parameters that follow a specific distribution for specific contexts or for special classes.

The motivation behind this work stems from both the movement towards more sustainable vehicles which subsequently results in a rise in the use of EVs, and the movement towards enhancing and providing more open-source data. This methodology, therefore, provides the groundwork for how planning CS infrastructure could be carried on using open-source data but also provides an overview of the challenges that researchers would face in that attempt. In future works, this methodology could be enhanced through the acquisition of more case-specific data such as CS profiles, demographic data for driver behaviours, more specific data on building uses, vehicle behaviour, etc. As open-source data would fall short in accuracy, the results acquired through this work don’t represent, therefore, a real-life estimation but rather showcases the falling short of CS infrastructure. It is hence important to keep track of the rate of accuracy of the results and to provide more validation methodologies in the future.

Acknowledgments

This work was created as part of the ERA-Net Smart Energy Systems project CrossChargePoint (CCP), funded by the European Union’s Horizon 2020 research and innovation program under grant agreement no. 775970 and the German Ministry for Economic Affairs and Climate Action.

References

[1] Hecht, Christopher and Das, Saurav and Bussar, Christian and Sauer, Dirk Uwe Representative, Empirical, Real-World Charging Station Usage Characteristics and Data in Germany. In 2020 eTransportation, pages 100079.


[16] Neufert, Ernst and Neufert, Peter and Kister, Johannes Architects’ data, 2012,


Presenter Biography

Hana Elattar is a researcher in the Deggendorf Institute of Technology in Germany. Holding a master of science degree in international cooperation urban planning from the University of Grenoble-Alpes - France, she is currently preparing to start her PhD in the field of energy spatial planning. Her work with DIT includes working on ERA-Net’s CrossChargePoint project with as focus on planning CS infrastructure. She contributes to the project as a researcher and co-team coordinator. She has experience with working on projects in topics of Smart Cities and participatory planning.