Rotterdam School of Management, Erasmus University

Master Thesis

Master of Science, Business Information Management

# Charging Ahead - Predicting Optimal Charging Station Locations across Multiple Cities

*Author:*
Jonathan von Rueden
(462656jv)

*Supervisor:*
Micha Kahlen
Dr Rodrigo Belo

June 16, 2017

## Preface

The copyright of the master thesis rests with the author. The author is responsible for its contents. RSM is only responsible for the educational coaching and cannot be held liable for the content.

# Charging Ahead - Predicting Optimal Charging Station Locations across Multiple Cities

Jonathan von Rueden
Rotterdam School of Management, 2017
Thesis, Master of Science

## ABSTRACT

Electric vehicle (EV) adoption and car sharing use have seen strong growth over the last decade due to recent advances in technology and information systems. Yet, the chicken-egg problem of charging station construction vs. EV adoption limits faster expansion. Car sharing of electric vehicles, however, has the potential to help reduce people's risk aversion towards EVs, reduce the overall number of cars in already congested cities, and increase the share of stable demand for operators to incentivize the construction of more charging stations.

This paper aims to address a current research gap in the creation of predictive models using point of interest data to determine optimal charging station locations. The best model is constructed using multiple large data sets of EV car sharing (parking) data, which spans over one year in five large European cities, as well as point of interest information. The results show a high predictive accuracy, especially where information on different types of point of interests is high. Overall the model outperforms any naive model by 24.1% in targeted accuracy and can be used as a decision support system for operators to reduce risk from charging station misplacement. Finally, various areas are identified where information from e.g. charging station operators on the profitability of stations has the potential to further increase the model's performance.

## 1. INTRODUCTION

Recent advances in technology and information systems have significantly improved the offering and market for electric vehicles (EVs) with more than 1 million vehicles on the road worldwide in 2016 [34]. EVs are expected to contribute significantly to climate change mitigation [11] and improvements in urban air quality [7] as well as a decrease in fossil fuel dependency [19]. Mainly for these reasons national governments are working together globally to increase overall EV adoption as in the case of the *Electric Vehicle Initiative* founded by 16 major industry nations with the goal of a "global deployment of 20 million electric cars by 2020" [17].

Nonetheless, electric vehicle adoption does not increase as fast and widely as expected with, for example, Germany recently abandoning its goal of more than one million EVs on the road by 2020 [59]. Various recent research in the field has identified the following problem areas hindering faster adoption: Range Anxiety [25] (the fear of being stranded without any battery charge left), the chicken egg dilemma (of EV charging stations) [28], vehicle purchasing prices [71], overall economic viability [66] and poor consumer informedness [47]. To contribute to potential solutions in the aforementioned problem areas, this paper aims to improve understanding of electric vehicle usage and infrastructure by analyzing EV car sharing data and developing a predictive approach for the placement of charging stations.

The focus on the comparatively narrow car sharing sector with the goal of contributing to rather wide topics like overall electromobility and sustainability is motivated by two main aspects: Car sharing has a signaling effect on potential future EV owners and sharing can often better than owning:

Recent research suggests that car sharing and electromobility form a "circulus virtuosis"[24] by encouraging customers to test the novel vehicles risk-free hence increasing their likeliness to buy an EV and on the other side by providing a baseline demand for charging stations in and around city centers. This interaction is expected to help reduce range anxiety and to increase product awareness on the consumer side while ensuring stable revenue on the provider side. Ultimately, both parties are lowering the impact of the chicken-egg dilemma. It is then more profitable for providers to install charging infrastructure if there are more EVs on the road. Consumers, in turn, are more inclined to purchase battery-only electric vehicles (BEVs) being reassured that the likelihood to search for charging stations for too long or to run out of battery mid-ride is significantly reduced. Finally, Brandt et al. [8] find that users of electric car sharing vehicles adopt similar usage patterns to those of users of internal combustion engine (ICE) vehicles quickly, the positive effect the exposure from car sharing has on people's aversion towards BEVs in general.

The car sharing sector has seen very strong growth rates at 30 - 60% year over year during the last decade [39], [54] and is even at times praised as "the winning game for automotive OEMs [Original equipment manufacturer]" [44]. In fact, car sharing is both, better for social and often individual welfare than owning a car: the average idle time of a privately owned car is estimated at 90 - 95% [4] while car sharing vehicles achieve utilization rates of up to 20 - 25% [1]. Making use of car sharing hence could translate into direct savings for people not using their cars often and can in turn also eliminate cars from the street, which is generally good for social welfare. Mitchell et al. [43] even find that car sharing in an urban setting can eventually reduce idle time so drastically that a mere 15% of the initial fleet of cars would be needed. Replacing private vehicles by shared BEVs and therefore reducing the total number of vehicles in cities would lead to reductions in micropollutants, CO2 emissions and free up tight parking space. And while EVs are going to become more environmentally friendly from year to year, EVs can nowadays still be more polluting than ICEs [31], depending on the source of energy, when and how often they are used. Sharing a vehicle by virtue reduces the emissions per capita enough to make any EV the more sustainable solution now and even more so in the future.

While many large western cities already have (EV) sharing schemes in place, smaller western cities and most cities in Asia, Africa and South America [53] are still heavily underserved. Yet, developing nations are by far the biggest contributors to the increase in urban population for the next decades to come [61] and therefore also those nations with the largest increase in urban vehicle use. This has already become an issue in many cases with pollution and congestion reaching unprecedented levels[48], hampering the economy, damaging the environment and attacking the health of residents. Beltz et al. [5] find that car sharing can have significant effects in abating the impact of these threats. And even though the situation in western mid-sized cities, where BEV car sharing is also rarely in place [50], is not as severe, the number of cars per capita still increases inversely to the size of a city [40]. Hence BEV car sharing in western mid-sized cities can have the power to replace even more privately-owned cars per person and reach more ICE car owners than in larger cities.

Nonetheless, in both cases, introducing BEV car sharing is still subject to the chicken-egg dilemma as described by Gnann et al. [27]. The authors find that without sufficient public charging infrastructure, the desired adoption rates of EVs remain unrealistic and, in turn, operators, receiving no financial benefit, are lacking an incentive to build the needed charging infrastructure. Other scholars like Brooker et al. [10] and Guo et al. [32] confirm this finding. Trying to solve the dilemma, they focus their research on finding the most suitable locations for profitable electric vehicle charging stations. In further studies on charging station placement, Madina et al. [41] and Gopalakrishnan et al. [29] use several different data sources and find a strong relation between charging demand and points of interest (e.g. the location of restaurants, museums, shops) when modeling the problem. Additional research by Wagner et al. [69] applies these insights in a car sharing context, confirming the importance of points of interest (PoI) for parking pattern modeling. Yet the majority of studies use ICE vehicle parking patterns as a proxy for their EV charging models [10] & [14] and none have used real BEV parking behavior in combination with the actual point of interest information to model charging station placement.

With a gap in current research on modeling charging station placement and having access to both real BEV (car sharing) parking behavior and recent point of interest information the following research question was defined:

**How can point of interest information be used to improve decision making in charging station placement?**

In the following, Section 2 reviews related work and literature, Section 3 the data used for modeling and Section 4 the methodology employed. The results are discussed in Section 5 and discussed in Section 6. A conclusion on the results and potential future research is drawn in Section 7.

[1]Car2Go, Amsterdam, 2016

## 2. RELATED WORK AND LITERATURE

With reference to the research question, related work and the necessary literature for building a theoretical model are examined and finally, hypotheses to be tested are derived.

### 2.1 Charging station placement

As the number of BEVs grows every year, so has research on how to optimally meet the charging demand of these vehicles. Since this research has only gained traction over the last decade and information on both charging as well as driving and parking patterns is difficult to acquire, most of the research papers develop theoretical models or use older ICE survey data as proxies.

Lam et al. [38] for example develop an "Electric Vehicle Charging Station Placement Problem" purely based on theoretical assumptions without any data to test their model on. The model developed aims at optimizing location utilization and availability while minimizing station cost. The approach is picked up by Guo et al [32] who develop the model further by validating its assumptions and applying a further theoretical method (fuzzy TOPSIS optimization) to qualitative characteristics of Beijing's charging infrastructure. The model is employed by Sathaye et al. [51] who apply it to real US Census, Department of Transportation trip information from Texas and charging station data. The authors find that information on the density of charging stations, traffic, and points of interest are most important in successfully applying their model. However, the paper concedes that "there remains significant uncertainty regarding the estimation of demand for BEV charging, due to a lack of available information on BEV driver behavior" [51].

Similar to the aforementioned work, Chen et al. [15], Dong [18] and Funke et al. [26] use ICE trajectory data to determine traffic density and derive theoretical charging demand for EVs from the data. The authors also find that points of interest influence the theoretical charging demand and point out the importance of their distance to the destination. Dong finds that "electric [vehicle] miles and trips could be significantly increased by installing public chargers at popular destinations, with a reasonable infrastructure investment". Ultimately all three papers propose that future research should assess the relationship of points of interest and real EV driving data. Along with this proposal, Brooker at al. [10], in a recent study, explore the connection of PoIs with survey and recorded charging data. The study finds that "public charging is most likely to occur at work, shopping and social destinations", all identified by points of interest at the destination and concludes by stating that "understanding the driving patterns of vehicle owners is critical[...]" in determining the need to charge.

The approach of analyzing driving patterns and points of interest is applied in the work of Xi et al. [72] where the authors use ICE driving data to estimate charging station demand for the categories "work", "university" and "shopping". The report finds that all three factors significantly improve utilization modeling, yet state that more should be taken into consideration and that "it is more appropriate to focus on EV arrival and departure times from parking lots since this is when slow charging can be reasonably done".

In a series of papers, Wagner et al. [69], [70] & [68] use free-float (pick up and drop off cars anywhere within the operating area) car sharing data in combination with spatial smoothing and regression analysis. The authors find significant relationships between different types of PoIs and EV charging stations considering the distance of points of interest. When determining the influence of PoIs on the parking behavior of car sharing users, the willingness-to-walk (to different locations), in other words, the distance to close by PoIs, proofed to be statistically significant and was considered with the inversely weighted distance in meters.

### 2.2 Predictive and Spatial modeling

Shmueli et al. [20] argue that in the field of Information Systems (IS) research, where the majority of studies employ explanatory models, "predictive models and testing play an important role in assessing the practical relevance of existing theories, and quantifying the level of predictability of phenomena". Successfully applied predictive models can hence augment the credibility of existing theories. With a general, growing availability of data across industries, countries, and practices, as well as the availability to store and share this data ever more, theoretical models can more often be validated with new data sets hence show their actual value in explaining new observations. Especially in the field of information systems the emergence of "big data" and machine learning technologies supports scholars and companies in gathering, storing and analyzing larger data sets, making it possible to validate theoretical models and to apply them to numerous new observations.

In predictive theory, the interpretability and veracity of a model are of lesser significance: a model that captures all aspects of a problem may be less "predictive" than a less realistic model [56]. Concepts such as transparency, interpretability, and multicollinearity that are firmly established in explanatory research are therefore of secondary importance [20]. Ergo, non-parametric methods like machine learning algorithms are often used in predictive modeling.

In the research area of this paper, very few studies have focused on testing the predictive power of their models, none have developed predictive models for charging station placement in combination with PoI and BEV car sharing data. Arias et al. [2] develop a model for demand prediction that identifies peaks throughout the day and aims at optimizing load. While the authors can show the predictive value of their approach, modeling and predicting the daily fluctuations of EV charging is of no added value to this research. However, Arias et al. confirm that adding PoI data to their predictive models improves the prediction accuracy significantly. Another paper by Wagner et al. uses car sharing data and develops a predictive model for the number of rentals in one city. The authors lay a grid over the city area, aggregate the number of rentals per cell, add the point of interest information and ultimately find "that [their] approach correctly identifies areas with a high car sharing activity and can be easily adapted to other cities." [68].

In an effort to create predictive models for the change in tree population induced by global warming, Prasad et al. [49] also divide the area under consideration into rasterized cells and compare multiple machine learning algorithms, concluding "superior predictive capability" of Random Forest modeling for spatial data. Oliveira et al. [46] use the ensemble learning method Random Forest to predict wildfires and argue that because the data is highly irregular and clustered spatial autocorrelation might influence the predictive power of their model. Spatial autocorrelation may arise when proximal locations (i.e. adjacent cells) are correlated as in the case of wildfires and also vehicle parking behavior.

Oliveira et al. use adaptive kernel density estimation (also referred to as spatial smoothing) to control for spatial autocorrelation and find that the "[Random Forest] model seems to incorporate much better the effects of spatial autocorrelation".

Yet, arguing that smoothing the data "hides" information in predictive modeling and reduces location accuracy, Kamitani et al. [36] consider the value of adjacent cells when creating a predictive model for the identification of voxels in MRI brain mapping.

Mascaro et al. [42] further add spatial location indicator information to a model predicting the deforestation rate to control for spatial autocorrelation and find that predictions improved by 60% in adjusted $R^2$. Noting that non-parametric models like Random Forests do not expressly incorporate any spatial structure, Chefaoui et al. [13] propose the inclusion of a variable indicating a trend in the geographic space, in other words, an anchor point, to account for spatial structure without reverting to smoothing of any information.

**Research goals**

From the above review of the literature and related work, it is concluded that current research lacks the application of theoretical charging point location prediction models to realistic EV parking data for the creation of robust predictive models. Previous research suggests that prediction problems for charging station placement should be modeled using gridded EV parking data and can be significantly improved when including PoI information. Additionally, the ensemble learning method Random Forests performed well in similar grid-based spatial prediction problems.

It is further to be noted that no research thus far employed the combination of real EV parking data with points of interest to predict the optimal locations for charging point placement. Optimal locations being those that maximize profit for operators and consider charging capacity and availability for car sharing users. Based on the work and findings of related research, this paper will validate whether (1) point of interest information can be used to predict profitable EV charging locations in mid-sized to large cities and (2) the validity of such predictive models holds true across cities.

# 3. DATA

In this section, the data acquisition process, its contents, and transformation are described.

## 3.1 Data acquisition and description

Using a custom-build web scraper application, real BEV parking data from the two major car sharing providers Car2Go and DriveNow was collected. The data was collected over at least 9 months per city (see table 2) and holds the location, timestamp, fuel level and unique id per car. Information on point of interest (PoI) was acquired through Google's public API[2]}. This PoI data holds information on location, type, price and rating of PoIs in each city.

### *3.1.1 Car sharing data*

Both car sharing companies operate a free-float scheme in multiple cities worldwide where users can pick up and drop off vehicles at any public parking spot in a defined area within the city. Customers can rent the cars through an app or RFID card on the spot and pay on demand in-between 29 to 34 Euro cents per minute which include all insurance and fuel cost. Optionally customers can rent the cars for multiple hours at once and pay a chapter fixed rate. For customers to locate the vehicles, both operators offer a mobile and web application to locate all vehicles that are not

[2]https://developers.google.com/places, accessed on June $14^{th}$, 2017

**Table 1: Example of original car sharing data**

| Id | Lon | Lat | FuelLevel(%) | Charging | Timstamp |
|---|---|---|---|---|---|
| WBY1Z210X0V307780 | 55.69196 | 12.57878 | 83 | TRUE | 2016-07-30 13:55:06 |
| WBY1Z210X0V307780 | 55.69196 | 12.57878 | 84 | TRUE | 2016-07-30 14:00:06 |
| WBY1Z21080V308071 | 55.66487 | 12.51916 | 53 | FALSE | 2016-07-30 10:48:16 |
| WBY1Z21080V308071 | 55.66487 | 12.51916 | 53 | FALSE | 2016-07-30 10:53:16 |
| WBY1Z21080V308071 | 55.79005 | 12.59147 | 39 | FALSE | 2016-07-30 12:10:16 |

in use at a given moment. The web scraper application collected this information every five minutes over a period of at least 9 months (`2016-07-31` to `2017-04-30`, about 84 million observations). While cars are rented they do not appear on the mobile or web application and are hence not recorded.

As target cities, Amsterdam, Munich, Stuttgart, Copenhagen, and Berlin were chosen, because of each city's relatively high share of electric vs. ICE vehicles to evaluate the model on. In addition, all cities are in similar climate zones [37] which reduces the variation in battery performance across cities [73]. Amsterdam and Copenhagen are extremely similar in terms of communiting and traffic patterns [21], so are Berlin, Munich, and Stuttgart[22]. Overall the five cities are still similar very similar to each other.

**Table 2: Car sharing data per city**

| City | Operator | Observations | BEVs | Record (months) |
|---|---|---|---|---|
| Amsterdam | Car2Go | 25027666 | 338 | 9 |
| Berlin | Drive Now | 5349875 | 135 | 12 |
| Copenhagen | Drive Now | 17548977 | 399 | 9 |
| Munich | Drive Now | 2144178 | 85 | 9 |
| Stuttgart | Car2Go | 34576399 | 501 | 9 |

#### 3.1.2 PoI data

The PoI data has been collected via the Google development API in May 2017. It holds information on the location, name, rating (on a scale from 1 to 5), price (on a scale from 1 to 4) and type (i.e. a gym holds the attributes *establishment, gym, health*). As the Google API restricts the number of PoIs that can be downloaded at once, for a given coordinate and radius, to 60 per call, a hexagonal grid was used to divide each city's area. To optimize the trade-off between geographic accuracy and computing cost, the distance between the centroids of the hexagons is set to 128 meters where $\approx 0.4\%$ of lesser ranked PoIs[3].\ Charging points are also considered points of interest but were not exported via the Google API. The location information of charging points was extracted from the car sharing data set by identifying all unique locations where *Charging = TRUE* was recorded.

**Table 3: PoI information per city**

| City | No. types | Observations |
|---|---|---|
| Amsterdam | 100 | 110496 |
| Berlin | 101 | 227896 |
| Copenhagen | 105 | 116272 |
| Munich | 107 | 155720 |
| Stuttgart | 106 | 75728 |

### 3.2 Data preparation

To answer the research questions, the charging demand in any given area of each city must be determined to identify whether an area can be considered profitable for charging point operators. Since the data only holds information on car sharing parking behavior and the cars do not charge every time they are parked, one must estimate the potential charging demand per parking instance in kWh and account for the potential charging demand of private vehicles. *Potential charging demand* will, from here on out, be referred to as *charging demand*. The initial raw data must at first be aggregated to parking instances, i.e. a record of the time before and after a car is moved to a new location and is defined as follows:

$$p_i = (id, lon, lat, fuel_s, fuel_e, t_s, t_e, chg) \qquad (1)$$

[3]Ranking determined by Google, see footnote 2

Where $id$ is the unique id of the car, $lon$ & $lat$ the coordinates of the parking instance, $fuel$ the start ($s$) and end ($e$) levels of fuel in percent, $t$ the timestamp at the start and the end of the parking instance and $chg$ whether the car was charging at the point of recording.

Both Car2Go and Drive Now automatically lock down their cars if connected to a charging station and have fuel levels of less than 75%. This ensures that cars are sufficiently charged for the next customer and are not rented before the battery reaches at least 75% again. During the lockdown, the web scraper is not able to record a car's location or activity, as this car is "artificially" rented by the company itself. To control for this loss of information, all parking instances which succeed a lockdown are identified and the fuel level and parking instance start times corrected through mean interpolation. The record for $fuel_s$ at lockdown parking instance $p_t$ is adjusted to $fuel.adj_{s,p_t}$ with:

$$\begin{aligned} fuel.adj_{s,p_t} = fuel_{s,p_t} - \frac{1}{n}\sum_{i=2}^{n}(fuel_e, p_{i-1} - fuel_{s,p_i}) \\ -(fuel_{e,p_{t-1}} - fuel_{s,p_t})^4 \end{aligned} \tag{2}$$

$$fuel.adj_{s,p_t} = \begin{cases} 1, & \text{if } fuel.adj_{s,p_t} < 1 \\ fuel.adj_{s,p_t}, & \text{otherwise} \end{cases} \tag{3}$$

The record for $t_s$ at lockdown parking instance $p_t$ is adjusted to $t.adj_{s,p_t}$ with:

$$\varnothing\delta_{min}{}^5 = \frac{\sum_{j=2}^{j}(t_{e,p_{j-1},chg=1} - t_{s,p_j,chg=1})}{\sum_{j=2}^{j}(fuel_{e,p_{j-1},chg=1} - fuel_{s,p_j,chg=1})} \tag{4}$$

$$\begin{aligned} t.adj_{s,p_t} = t_{s,p_t} - \frac{1}{n}\sum_{i=2}^{n}(t_e, p_{i-1} - t_{s,p_i}) \\ -((fuel_{s,p_t} - fuel.adj_{s,p_t}) \times \varnothing\delta_{min}) \end{aligned} \tag{5}$$

$$t.adj_{s,p_t} = \begin{cases} t_{s,p_{t-1}}, & \text{if } t.adj_{s,p_t} < 0 \\ t.adj_{s,p_t}, & \text{otherwise} \end{cases} \tag{6}$$

With the corrected fuel levels, start and end times the potential charging demand per parking instance can be calculated. To avoid overestimation only the amount of energy used in a car's previous trip ($p_{i-1}$) is accounted for as potential charging demand at the location of the current parking instance ($p_i$). For the sake of simplicity, it is assumed that all cars are put in service with an initial battery load of 100% and hypothetically fully recharge their battery during every parking instance. In such a way, one can safely aggregate real charging demand at the locations of all parking instances and will never account for more potential demand than the amount that is charged in reality. The maximum deviation of total potential charging demand vs. the amount actually charged is $\approx$ 1.5 kWh per car per month in any of the targeted cities due to mean interpolation deviations.

$$\varnothing\delta_{charg} = \frac{\sum_{j=2}^{j}(fuel_{e,p_{j-1},chg=1} - fuel_{s,p_j,chg=1})}{\sum_{j=2}^{j}(t_{e,p_{j-1},chg=1} - t_{s,p_j,chg=1})} \tag{7}$$

$$\sum t_{p_i} = t_{e,p_i} - t_{s,p_i} \tag{8}$$

$$fuel_{t_{p_{i-1},p_i}} = \; \| (fuel_{s,p_i} - fuel_{e,p_{i-1}}) \| \tag{9}$$

$$Q_{d,p_i} = \begin{cases} fuel_{t_{p_{i-1},p_i}}, & \text{if } fuel_{t_{p_{i-1},p_i}} < \\ & \varnothing\delta_{charg} \times \sum t_{p_i} \\ \varnothing\delta_{charg} \times \sum t_{p_i}, & \text{otherwise} \end{cases} \tag{10}$$

With $\varnothing\delta_{charg}$ being the energy (in percent) charged per minute on average, $\sum t_{p_i}$ the duration of the parking instance and $fuel_{t_{p_{i-1},p_i}}$ the amount of battery charge used for the previous trip. $Q_{d,p_i}$, the potential charging demand for parking instance p_{i}, is equal to $fuel_{t_{p_{i-1},p_i}}$ if the duration of the parking instance would permit recharging the car fully, else it is calculated as the duration of the parking instance times the average load charged per minute and left over demand carried on to the next parking instance $p_{i+1}$.

Figure 3.2 describes two sample trajectories of a car starting one rental in cell `A`, stopping it in cell `B` before embarking on its next rental journey to cell `C`. Trajectory 1[6] being the very first trip a given car has ever completed, cell `B` would, in this case, be credited with $Q_{d,p_1}$, cell `C` with $Q_{d,p_2}$.

[4] Averages always calculated without lockdown instances and extreme values

[5] Average time (in minutes) per percent of charge loaded

[6] The exact trajectory of a car is not known, figure only for illustration

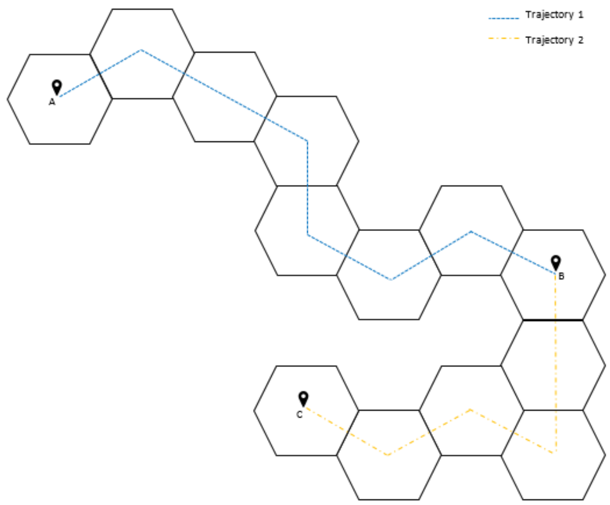


**Figure 1: Charging demand accounting**
Note: Sample trajectory of one car in a given city for illustration. The delta in fuel level from first tip (A to B) will be recorded as potential charging demand at the desitnation (B).

To aggregate demand in a meaningful way, the city map is overlayed with a grid, splitting the area into thousands of cells as proposed by Wagner et al. [70] and Ahn et al. [1]. This transformation primarily helps to understand the charging demand of a certain area, rather than just at a single parking spot. Most cities have parking rules by which only EVs can park on EV charging point lots and must leave when charged [3]. It is therefore assumed that reducing the hypothetically available parking spaces (when charging) from all parking spots (where potential charging demand is recorded) in one cell to only two per charging station, would cover the parking demand. Similar to Wagner et al. a cell size of 200 meters × 200 meters is chosen, to design a fine-granular grid, however, using a hexagonal instead of a regular square (fishnet) grid.

Mapping the potential charging demand to the hexagonal grid was done for a multitude of reasons as described by Birch et al. [6]: First and foremost hexagons lower the sampling bias from edge effects and are more compact, having a lower shape index (the $\frac{perimeter^2}{area}$ ratio) than squares. This allows hexagonal grids to model patterns more accurately and to tessellate into a continuous grid (rather than aggregating the data in circles which have perfect shape indices but must overlap to cover the entire data set).

Furthermore, all neighbors of a hexagonal grid cell have the same distance to its centroid. This is of great importance when considering the willingness-to-walk (WTW) of car sharing users in the modeling efforts. Van der Groot [65] linearly weighs the importance of points of interest for the parking decision of drivers when considering their WTW. With a maximum radius of a 40-minute walking distance, a point of interest would be 50% as important for the parking decision if it is 20 minutes away by foot compared to a parking location with no walking distance to a specific point of interest. As a grid is employed for the aggregation of charging demand, weighing every point of interest individually based on their distance is not possible, but using different thresholds for the willingness-to-walk as proposed by Untermann et al. [62] is. Taking advantage of the equal centroid to centroid distance of neighboring hexagons one can construct multiple "circles of influence" for the willingness-to-walk around each hexagon. The thresholds in willingness-to-walk are loosely based on Untermann et al.'s suggestion of 100, 200 and 500 meters with the actual values being 107, 214 and ≈ 399 meters to accommodate a hexagonal side length of 214 meters which produces an area of 40,000 $m^2$, similar to Wagner et al.'s 200 × 200-meter grid.

Figure 2 shows the advantage of a hexagonal grid vs. a fishnet grid where all willingness-to-walk distances are equal for the first ring and almost equal to the second. On the contrary, the fishnet grid has much larger differences in distance for each ring. While Van der Groot et al. and Wagner et al. weigh the different distances, linearly and with a Gaussian kernel density estimation respectively, this paper's approach refrain from imposing a weight on the importance of the distance of certain PoIs as the machine learning method *Random Forest* will do so when training the model [60].

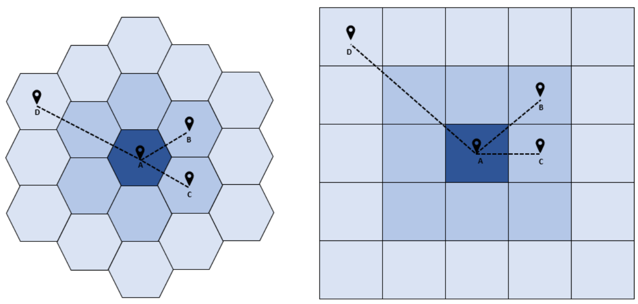


**Figure 2: Hexagonal grid vs. fishnet grid**

Note: In a hexagonal grid, $\overline{\mathbb{AB}}$ and $\overline{\mathbb{AC}}$ have the same distance, in a fishnet grid point $\mathbb{B}$ will be further away. For the grid granularity used in this paper, $\overline{\mathbb{AC}}$ in the fishnet grid would be ≈ 88 meters further away than $\overline{\mathbb{AB}}$.

To add a spatial component to the data model, as discussed in the literature review, the distance between an anchor point (the city center) and all cells are calculated. The distance was computed using the Haversine or great-circle distance which accounts for the earth's spherical surface between two points [64].

To account for seasonality, all observations over all full nine months are taken into consideration and aggregated per cell and month. The data sets for Amsterdam and Stuttgart hold observations of more than a year at the time of this writing, therefore all months over the span of one full year (`2016-03-31` to `2017-04-30`) were considered. The months `June` and `July` had to be eliminated as they only held observations of a factor of 10 or less than all other months across the entire data set. This extreme fluctuation cannot be considered a change in seasonal demand as both `May` and `August` hold as many observations as the all other regular months. To calculate the potential charging demand per cell per year the average demand per cell and months is derived and multiplied by twelve to obtain the yearly demand per cell $Q_{d,yr}(c_{id})$:

$$Q_{d,yr}(c_{id}) = 12 \times \frac{1}{m} \sum_{i=1}^{m} (Q_{d,m}(c_{id})) \tag{11}$$

With `m` being all recorded months except `June` and `July` and $(Q_{d,m}(c_{id})$ the aggregate demand per cell and month. The yearly aggregate is then multiplied by the kWh value of one percent of battery capacity [7] to transform the charging demand from battery percentage points to kWh.

## 3.3 Profitability

As the potential profitability of one or more charging stations in a given cell needs to be assessed, the further potential charging demand from private and company EVs must also be taken into consideration. The ratio of total kWh sold to Car2Go over the total amount of kWh sold in Amsterdam in 2016 [16] is used for adjustments. The resulting `14.5%` [8] is used to inflate the charging demand to amounts that account for car sharing and private demand per cell. The ratio is used to adjust demand across all cities assuming that the share in the other four cities is equal and that the additional demand at all stations used by car sharing BEVs can be covered with this ratio.

The cost for the installation and maintenance of two-outlet charging points is based on a report by NPE, a German EV charging infrastructure initiative [45] and confirmed by two of its members [9]. Charging stations are expected to be operable for 10 years on average [52], yearly cost, ignoring depreciation and interest, is therefore calculated over a time horizon of 10 years. The following cost estimates are assumed to hold true across all three countries (Germany, Netherlands, and Denmark) the data was collected for.

**Table 4: Cost per charging station per year**

| Expenditure | Cost in Euro |
|---|---|
| Installation (CAPEX) | 1,000 |
| Maintenance (OPEX) | 800 |
| Total cost | 1800 |

The revenue per kWh sold for charging station providers, however, varies substantially per country and depends on whether a provider is also an energy producer or needs to buy the energy for retail prices. For the analysis, a conservative scenario where the charging station provider has to buy the energy at a retail price level was chosen. The wholesale energy prices per county are also listed to illustrate the additional margin energy producing companies could reap when offering their energy at charging stations. All prices are in Euros and per kWh:

**Table 5: Energy prices for station providers**

| Country | Production cost | Retail price [10] | Sales price (at station) |
|---|---|---|---|
| Denmark | 0.053 [11] | 0.39 | 0.71 [12] |
| Germany | 0.056 [13] | 0.29 | 0.54 [14] |
| Netherlands | 0.044 [15] | 0.19 | 0.33 [16] |

With the above price and cost listings one can calculate the number of stations that would be profitable per cell given the calculated demand, cost, and potential revenue:

$$Q_{d,st}(c_{id}) = \left\lfloor \frac{Q_{d,yr}(c_{id}) \times margin \times \dfrac{1}{0.145}}{-(cost) + profit} \right\rfloor \tag{12}$$

[7] 0.188 kWh/percent for DriveNow, 0.176 kWh/percent for Car2Go

[8] $\frac{559507kWh}{3858673kWh}$, charging demand Car2Go over the total in Amsterdam 2016, calculated from acquired data and report by the city of Amsterdam[@Charg2016]

[9] Fabian Deipenbrock (DB Energie GmbH) on May $13^{th}$, 2017, Johannes Hauck (Hager Electro GmbH Co. KG) on May $19th$, 2017

[10] Eurostat, for year 2015

[11] Levitt 2016, for year 2015

[12] EONDenmark 2017, retrieved on June 14th, 2017

[13] Stromreport 2017, for year 2015

[14] Goingelectric 2017, retrieved on June 14th, 2017

[15] Planbureau 2016, for year 2015

[16] NuonEnergy 2017, retrieved on June 14th, 2017

With $Q_{d,st}(c_{id})$ being the number of stations expected to be profitable per cell, *margin* the sales price minus the purchase price of energy per kWh, cost the yearly cost per station and profit the targeted profit per station (200 Euros in profit per year and station assumed).

As EV charging infrastructure is still largely subsidized [55] and the method may not have captured all potential demand, the calculated number of the profitable charging station is always lower than the number of charging stations accessed by the car sharing vehicles as can be seen in the table below [6]. Using real EV parking data to identify potentially profitable locations guarantees that no-parking zones are not considered at all. As no parking data was recorded in no-parking zones in the first place, they cannot be identified as profitable locations in the calculation.

**Table 6: Charging stations per city**

| City | Existing stations | Profitable stations |
|---|---|---|
| Amsterdam | 1303 | 351 |
| Berlin | 462 | 149 |
| Copenhagen | 419 | 348 |
| Munich | 154 | 77 |
| Stuttgart | 669 | 494 |

The last step in preparing the data for analysis and model application is to count all points of interest per cell ($c_{id}$) and for the two rings of adjacent cells representing the willingness-to-walk features. Even though up to 107 unique types of points of interest were recorded, only 97 unique types of points of interest which [10] matched across all cities were used for modeling. The point of interest count per cell was marked with just the PoI type's name, i.e. `shopping_mall`, the first willingness-to-walk ring holding the sum of all 6 adjacent cells with `shopping_mall_r1` and the second ring holding the sum of all 12 next-adjacent cells with `shopping_mall_r2`.

## 3.4 Descriptive statistics

Previous research in the field has repeatedly stressed the strong influence of points of interest on parking and charging demand. One could, therefore, be led to the assumption that the density of points of interest would itself already influence the parking and charging demand and expect an increasing relationship of PoI density and charging demand per cell. However, looking at Figure 3, where extreme values from cells with e.g. airports have been removed, one can see that the trend is at most of the moderately inclining nature.

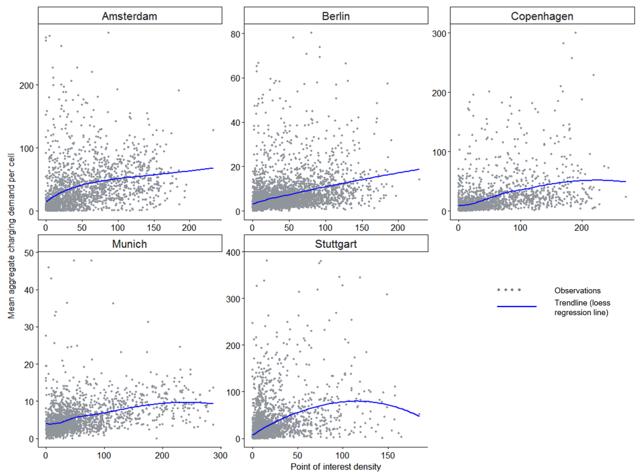


**Figure 3: Point of interest density on mean cell-aggregate charging demand (in kWh)**

Note: The number of points of interest observed (x-axis) vs. the mean aggregate charging demand per cell (y-axis). A loess (LOcal regrESSion) line was fitted onto the scatterplot to illustrate that a growing number of point of interests has often a bell-shaped relation to the charging demand. At very high PoI count numbers, charging demand often drops again, therefore the PoI density is not expected to be a sole predictor of charging demand.

It is likely that this bell shape is caused by parking restrictions and bans in the densest parts of city centers and indicates that charging demand can be influenced by the number of PoIs up to a certain extent, after which the type of PoI might play a bigger role.

A simple linear regression allows for the first impression on features which have a strong influence on the charging station locations that have been identified as profitable. Table [11] in Appendix I shows all statistically significant ($p<0.01$) features of a linear regression[17] with the entire feature space on the count of profitable charging stations per cell. One can see that, as expected and consistent with previous findings, the *density of points of interest*, the existence of *charging points* and the *distance to the city center* have great influence on the target variable.

PoI types such as *bar*, *university* and *car rental* are also positively relevant and significant. Given that all three types typically have very different surroundings (bars would be in more dense areas whereas universities have more space around them and car rental branches are often close to major roads and away from the city center), one can expect the

[17] Fixed effects (seasonality, charging stations and neighborhood) included

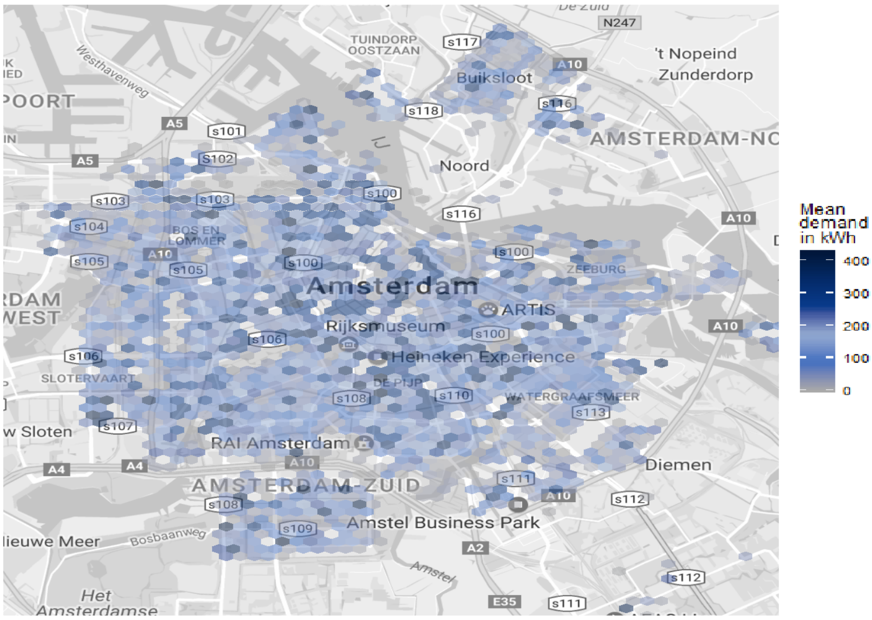


**Figure 4: Charging demand per hexagonal cell in Amsterdam**

Note: Each cell holds the aggregate mean monthly charging demand. While the city center holds more cells with high charging demand, no clear pattern or can be observed. This underlines the observation in Figure 3 that PoI density, which is higher in the city center, does not forcibly correlate with a higher charging demand.

specific type of PoI to play an important role in the predictive model. Aside from an interesting overview of the data one should interpret the results of this regression with caution as the spatial autocorrelation cannot be fully ruled out with only a spatial *anchor point* (distance to city center).

Confirming this finding, Figure 4 shows that the geographic distribution of potential charging demand is clearly skewed towards the city center (the darker the cell, the more demand), yet not entirely concentrated there but also widely spread throughout the business area of Car2Go in Amsterdam. Figure 13, 14, \ref{fig:muc_dmdand 16 in Appendix II show similar patterns for the remaining four cities[18].

Nonetheless, parking patterns and hence charging demand vary substantially per city. Figure 5 shows the scaled distribution of charging demand across the five cities. Munich and Berlin, having very similar demand patterns, both have many *low-demand* cells, i.e. cells where cars either do not park very often (remote areas) or do not stay long enough to generate demand (city center and train stations). In ascending order, Stuttgart, Copenhagen, and Amsterdam tend to have more cells with a high average charging demand.

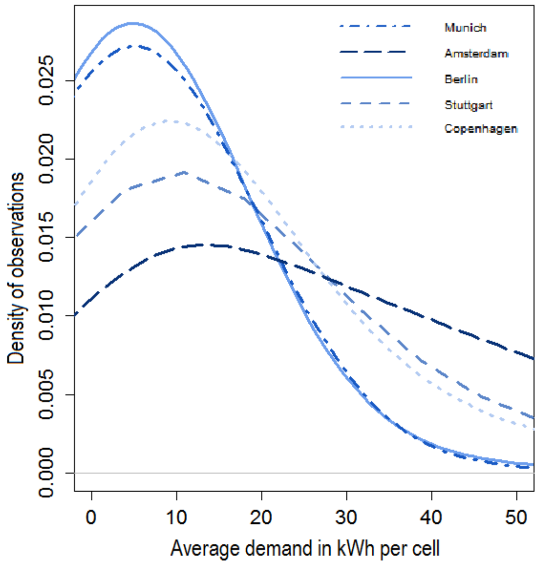


**Figure 5: Charging demand density across all cities**

Note: The density of the average charging demand per cell per month. All cities have different average demand distributions due to the city of the city, hence the number of cells in the grid, due to the number of shared vehicles and the city's road infrastructure.

[18]Hexagons on map are not representative of their actual size (see Data preparation)

This heterogeneous distribution can partly be explained by varied road infrastructure and partly by the fact that the latter three cities have a smaller business area and therefore fewer cells which lead to more demand per cell on average. Ultimately, this underlines the need for a differentiated calculation of the target variable (the optimal number of charging stations per location) which was introduced by varying revenue and private vehicle-shares of total demand per city.

When aggregating charging demand per cell a potential pitfall is not to consider the number of simultaneously parked vehicles per cell. The assumption is made that in an optimal scenario a given cell would hold a constant parking demand (or less) over 24 hours as illustrated in Figure 6.

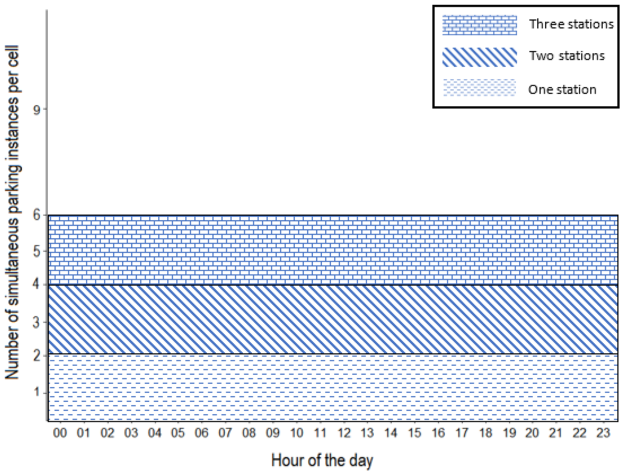


**Figure 6: Optimal density of simultaneous parking instances per hour and cell**

Note: The calculation of the number of potentially profitable stations per cell assumes a uniform distribution as displayed in the graph. This assumption of an optimal, uniform distribution infers that there would never be more cars than charging outlets (2x per station) in a cell at a time.

As each charging station can host a maximum of two cars at a time, a constant occupation without peaks through parking instance overlaps would be ideal to aggregate the charging demand without accidentally over accounting for situations where a cell has a capacity for two cars (one charging station) but three cars park in the cell at the time. To confirm this assumption, the actual distribution of simultaneous parking instances for the top 10% of charging demand creating cells was studied.

As can be seen in Figure 7, the assumption holds true for the majority of observations in the top 10% cells. Only a mere 2% of these 10% top ten cell observations hold more than two simultaneous parking instances in a given hour and date. However even these peaks are, in most cases, considered as

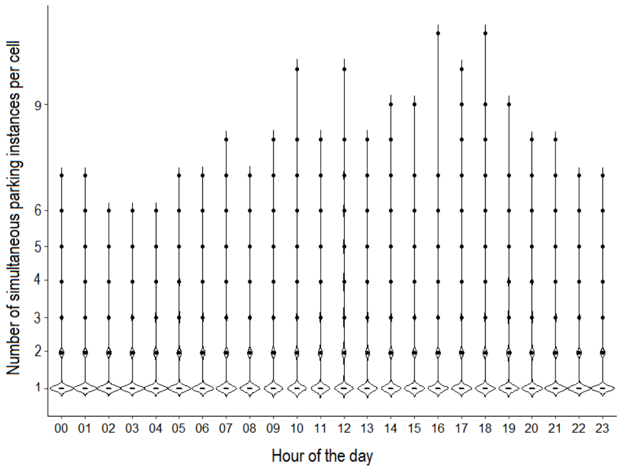


**Figure 7: Actual density of simultaneous parking instances per hour of top 10% cells**

Note: The cells with the top 10% most charging demand mostly have only one or two cars parked in it. The violin plot's width indicates the number of observations per timestamp in any cell of the top 10% cells. This finding affirms the assumption of a uniform distribution 6 of parked parks.

41% of the top 10% of cells hold more than one (calculated to be profitable) charging station.

## 3.5 Predictive model

Incorporating the previously collected ideas and modeling assumptions (see Related work and literature), a predictive model is created using the *Random Forest* algorithm which was developed by Breiman et al. [9] and its implementation in the `R` package `randomForest` [23]. The implementation requires a model to execute its algorithm on and needs only a few parameters such as `ntree` (number of trees to generate), `mtry` (number of variables to be sampled at each node split), `classwt` (weights for outcomes in classification problems) and `cutoff` (confidence level for class prediction). The model is optimized with these parameters measured against the F1 score which will be further elaborated upon. The final model is built with the following feature space:

$$\begin{aligned} Q_{d,yr}(c_{id}) \sim \sum poi_1 + ... \quad & ... + \sum poi_j + \\ & \sum poi.adj.r1_1 + ... + \sum poi.adj.r1_j + \\ & \sum poi.adj.r2_1 + ... + \sum poi.adj.r2_j + \\ & \frac{1}{n}\sum_{i=1}^{n} rating_{j,cid} + \\ & \frac{1}{m}\sum_{i=1}^{m} price_{j,cid} + \\ & d(cid, citycenter) \end{aligned} \tag{13}$$

With $Q_{d,yr}(c_{id})$ being the calculated optimal number of charging points in cell *cid*, poi_{j} being one observation of the point of interest of type *j*. $poi.adj.r1_j$ and $poi.adj.r2_j$ one observation of PoI type $j$ in the first and second willingness-to-walk ring respectively, the distance to the city center $d(cid, citycenter)$ and finally the average price and rating per cell.

To validate the results and confirm the predictive power of the model, a 5-fold cross validation is applied when calculating results and performance measures. As one of the goals of this work is to validate whether the model can be successfully applied to new cities or areas where no charging infrastructure has yet been placed, each fold splits the data set into 4 observed cities against one. For example, the model is trained on Copenhagen, Berlin, Munich, and Amsterdam and is then applied to Stuttgart. For all five combinations, the performance measures and predictions were recorded.

To further test the added predictive value of the modeling assumptions information was added iteratively to the model and measure its change in performance. The following models have been considered:

**Table 7: Iterative model creation**

| Model name | PoI density | All PoIs | WTW r1 | WTW r2 | Distance center |
|---|---|---|---|---|---|
| Model 1 | x | | | | |
| Model 2 | x | x | | | |
| Model 3 | x | x | x | | |
| Model 4 | x | x | x | x | |
| Model 5 | x | x | x | x | x |

Testing different ranges for the willingness-to-walk, hence cell sizes, is not pursued as the hexagonal cell size applied in this research is conform with previous research and changes would likely eliminate information from the data. An increase in the WTW ring size could cause loss of information because actual charging stations in the observed cities are on average $\approx$ 302 meters[19] away from each other.

Assuming that charging station operators have placed the stations with consideration of anticipated demand and knowing that actual charging points have an influence on potential charging demand [11], the proposed grid granularity captures most of the interaction effect two charging stations have on each other and the potential charging demand.

[19] As observed in the five target cities

Increasing the hexagon size would group more stations together and reduce information for the predictive model.

**Performance measures**

In this predictive classification problem, accuracy, or the ratio of correctly classified cases to the total number of cases, is used as a baseline measure to evaluate model performance. While it is an easily interpretable measure it should be taken with a grain of salt in classification problems and may not always be the ideal measure for a definitive evaluation. The accuracy paradox states that "predictive models with a given level of accuracy may have greater predictive power than models with higher accuracy"[63], since the predictive power depends on the target of the classification. For example, in a binary classification problem where the classes are highly unevenly distributed (99% of observations are zero, only 1% is one) and the target is one, an accuracy of 99% looks promising at first but is not any better than a naive model. Nonetheless, accuracy remains an important measure in this case since it is almost as important to identify cells with zero charging stations as it is to identify the ones with charging stations to avoid poor placement of stations.

According to the accuracy paradox, it may, at times, be advisable to choose a model with a lesser accuracy if it has a stronger predictive power with regards to the target. For this reason, other measures often used in classification problems are the *true positive rate* (TPR) and *true negative rate* (TNR), also referred to as sensitivity and specificity respectively.

$$TPR = \frac{TP}{(TP + FN)} \qquad TNR = \frac{TN}{(FP + TN)} \tag{14}$$

The rates are calculated using a confusion matrix which compares predicted with actual values (see Figure 8). Both rates are typically visualized on a receiving operator curve (ROC), showing the trade-off between `TPR` and `(1-FNR)` [57].

However, evaluating ROC as a performance measure for skewed data sets in machine learning classification problems, Jeni et al. [35] find that "while ROC was unaffected by skew, precision-recall curves suggest that ROC may mask poor performance." and suggest that using the F-score (also F1) improves balanced measurement. A second advantage of the F1-measure is that it can report on the performance of a multinomial classification (more than two classes, i.e. binary classification) problem by weighting and summing up the performance measure per class (in this case whether to

| | | **prediction outcome** | | |
|---|---|---|---|---|
| | | **p** | **n** | **total** |
| **actual value** | **p′** | True positive | False negative | P′ |
| | **n′** | False positive | True negative | N′ |
| | **total** | P | N | |

**Figure 8: Confusion matrix**

place zero, one, two, three, ... charging stations) [30]. The F1-score is a weighted sum of the predictive precision and recall, where $\beta$ is the relative importance of precision compared to recall.

$$PR = \frac{TP}{(TP + FP)} \qquad RC = TPR = \frac{TP}{(TP + FN)} \tag{15}$$

$$F = \frac{(\beta^2 + 1)RCPC}{\beta^2 PC + RC} \tag{16}$$

To account for skewed classes in the data, Zhang et al. [74] and Sokolova et al. [58] propose a micro and macro weighted F-score where the micro weighted score is biased towards the most frequent class (i.e. zero charging stations per cell) and the macro weighted score towards the least frequent class (i.e. more than three charging stations per cell).

$$\begin{aligned} PR_{micro} &= \frac{\sum_{i=1}^{|C|} TP_i}{\sum_{i=1}^{|C|} TP_i + FP_i} \\ RC_{micro} &= \frac{\sum_{i=1}^{|C|} TP_i}{\sum_{i=1}^{|C|} TP_i + FN_i} \end{aligned} \tag{17}$$

$$\begin{aligned} PR_{macro} &= \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{TP_i}{TP_i + FP_i} \\ RC_{macro} &= \frac{1}{|C|} \sum_{i=1}^{|C|} \frac{TP_i}{TP_i + FN_i} \end{aligned} \tag{18}$$

where `C` describes the class.

The models are evaluated with the F1 macro & micro weighted score and overall accuracy as calculated in the `R` package `mldr` by Charte et al.[12].

## 3.6 Modeling assumptions

To adjust the potential charging demand to include further potential charging demand from private vehicles, the amount of kWh charged into Car2Go vehicles in Amsterdam in 2016 is calculated and put in relation to the total amount charged by all vehicles in Amsterdam in 2016. The resulting number, adjusted for fleet size per city, is used to estimate the total charging demand in the remaining four analyzed cities assuming that all cities have sufficiently similar commuting and infrastructure usage patterns. A second assumption is made by increasing the demand linearly for every cell. While the patterns for simultaneous parking instances per cell do not diverge significantly from the optimal distribution assumption, rare double counting may arise (Figure 6 and 7).

Whether a location is suitable for the profitable placement of a charging station is calculated with a pessimistic scenario in which a charging station operator must pay full retail price for the energy he sells through his station. Additionally, charging station locations are deemed profitable if sufficient demand to cover the cost of one station is present. The profit margin per kWh sold considered in the calculation merely assumes one car per station while the station type on which the cost calculation is based has the capacity of providing energy for two cars at the same time. In simultaneous parking instance cases, at least two cars can, therefore, be served which reduces the impact of the parking duration distribution assumption.

Once placed, it is assumed that a charging station will be used whenever a car is parked in the station's cell. Both Car2Go and DriveNow already have incentive systems in place where customers are rewarded (if grossly negligent even punished) for charging the cars. Additionally, this assumption is based on the fact that all five cities in focus have regulations which bar non-EVs to park on charging station parking spots. A free parking spot within an 110-meter radius is certainly also an incentive for both car sharing and private users.

## 4. RESULTS

The final model performs well above any naive model, confirming that the assumptions made throughout this paper indeed add value and that using point of interest information significantly improves the predictive results. Table [8] shows the detailed results, where the best performing measure across all models is marked with an asterisk.

The model shows consistent improvement in performance when more information is added. The macro F-measure (weight on the importance of predicting $> 0$ charging stations) is highest in model 5, indicating that a model with all information available is by far best at predicting charging station locations while it is slightly less strong than model 2 at predicting locations where no charging station should be placed. The micro F-score, weighing the importance of predicting *no charging station*, is highest in model 2 making it the most favorable for avoiding the mispositioning of charging stations, but poorly suitable for predicting profitable locations. Model 3 ranks highest in overall accuracy, but, just as model 2, fails to identify suitable locations well.

**Table 8: Model performance results**

| City | Naive Model | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|---|
| Macro F1 | 26.2% | 23.1% | 39.3% | 44.9% | 46.4% | 50.3%* |
| Micro F1 | 92.9% | 85.9% | 93.4%* | 91.3% | 91.3% | 91.3% |
| Accuracy | 93.5% | 92.4% | 93.6% | 94.7%* | 94.3% | 94.6% |

Model 5 is best at identifying profitable locations and thus the best overall model. Consequently, the following analysis is solely based on the results of Model 5 with Figure 9 showing the scaled importance of predictors when training the model on the cities of Amsterdam. Further similar graphs displaying the scaled variable importance of the remaining four training iterations can be found in Appendix II (Figure 17, 18, \ref{fig:imp_minus_mucand 20).

The `%IncMSE` in Figure 9 displays the increase in the mean squared error of predictions (estimation based on an out-of-bag cross validation) if variable `i` is permutated, i.e. its values shuffled randomly, compared to a model with a non-shuffled `i`. It therefore shows how important the variable is to reduce uncertainty in the model. The `IncNodePurity` indicator lists features which, when used for a split in a decision tree, reduced the impurity of a given node the most on average across all trees in the Random Forest.

The graph in Figure 9, obtained from a model trained on the observations of Copenhagen, Berlin, Stuttgart and Munich lists, as the 5 most important variables to decrease the MSE, point of interest observations in willingness-to-walk ranges 1 and 2. All, except for `atm_adj_r1`, are points of interest such as `dentist` or `insurance_agency` where one would typically spend an extended amount of time and can hence accumulate plenty of potential charging demand. But also observations such as `grocery_or_supermarket` or `lawyer` within the very same cell significantly contribute to the predictive power of the model. While still an important feature, `point_of_interest`, or the density of PoIs, is much less a decisive factor than other types of PoIs. It is important to note that the specific types of points of interest are not proxy for density. Looking at e.g. three of the most important indicators such as `insurance_agency`, `liquor_store` and `hair_care` only 0.2% of all PoI observations are labled `liquor_store` and 0.5% `insurance_agency`, while almost 1% are labeled `hair_care`, it becomes apparent that not the number but the type of point of interest plays an important role.

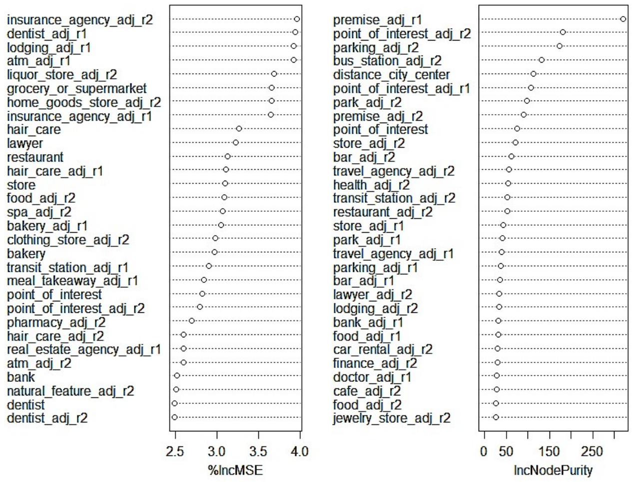


**Figure 9: Variable importance for model without Amsterdam**

Note: The variable importance shown in this plot mainly lists PoI types such as restaurants, lawyer's office or dentist, where one is likely to spend an extended amount of time. The notation of *adj.r*1 or *adj.r*2 indicates the willingness-to-walk of people who just finished their trip in a given cell. For example, the number of insurance agencies which are two hexagonal cells away (max. 642-meter walking distance) can have a great influence on the potential charging demand of a cell.

While ensemble learning methods such as Random Forests, even with dozens of trees created, correct for the habit of decision trees to overfit data [33], overfitting will eventually occur and the model will perform worse with increasing size when validated on test data. For the data set at hand, the optimal number of trees is 98 as can be seen in Figure 10 where the red dotted line represents the optimal tree size for the model. The optimal number of variables to choose from at each split (`mtry`) is 3 for the model and data in use. Both measures were derived and tested in a five-fold cross-validation to ensure robustness. Random Forests can further be optimized by manipulating the parameters `classwt` and `cutoff`, however, in this case, the tweaking of the parameters did not significantly improve results as a higher `cutoff` in most cases increases the *recall* of a prediction and the F-measures already account largely for differences in class weight. Finally, the graph also shows that training the model without `Berlin` greatly increases the out-of-bag error rate, likely due to Berlins large and rich set of informative points of interest.

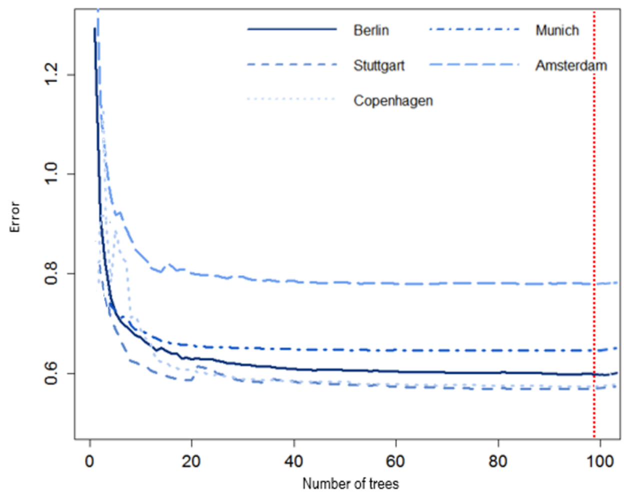


**Figure 10: Optimal number of Random Forest trees**

Note: The number of trees built in the Random Forest and the reduction in out-of-bag error shows that for this paper's data and model, an optimal forest size of 98 trees should be created.

### Practical evaluation

In a new city where, in theory, the charging station operator has no knowledge of the city except for the point of interest constellation, the model can serve as a decision support system for the placement of charging stations. Table [9] illustrates a sample calculation for a new city with different scenarios of prediction accuracy achieved by the model.

Scenario 2 simulates a situation where a charging point operator would have built 100 stations with a predictive model but using the model he can identify 30 out of 50 profitable stations which can cover 60% of the demand. As the operator has no way of knowing which station out of the predicted 50 will be a *true positive* he will build all of them. Since in all 5 cities 80% of stations deemed profitable are within at least 300 meters[20] of an actual station, 80% of the falsely predicted stations can replace a "regular" station out of the initial 100. However, 20% of the *false positives* may be at risk of being too far away and receiving too little charging demand. Hence a penalty of 4 stations or 20% on the falsely predicted stations is calculated. The operator can cover 60% of demand with 30 stations, 20% with the "regular" 20 stations of falsely predicted stations, has to build 4 more stations to cover the potentially missed demand (penalty) and lastly build 20 more "regular" stations to reach 100% demand coverage, saving 26 stations overall.

**Table 9: Example of station savings calculation**

| | Scenario 1 | Scenario 2 | Scenario 3 |
|---|---|---|---|
| Actual | 100 | 100 | 100 |
| Calculated | 50 | 50 | 50 |
| Correctly Predicted | 50 | 30 | 15 |
| Demand covered | 100% | 60% | 30% |
| Penalty (20%) | 0 | 4 | 7 |
| Station need | 50 | 74 | 92 |
| Total Savings | -50 | -26 | -8 |

Calculating the potential savings using a predictive model[21] and approach described above, it is concluded that in the case of hypothetically building charging stations, operators could save up to 11% in fixed cost investment per city. In detail, the hypothetical savings for operators amounted to 82,000 Euros in Amsterdam, 16,000 Euros in Copenhagen, 53,000 Euros in Berlin, 20,000 Euros in Stuttgart and 11,000 Euros in Munich in fixed cost per year.

Figure 11 shows the results of the final prediction in Amsterdam, the calculated stations and the actual stations on a map. The blue triangles represent the predicted stations, the red squares the actual stations and the green diamond shapes the locations where the calculation identified the potential demand for profitable stations. It is apparent that the model seems to predict more stations in the city center where the point of interest information density is much higher than in the outskirts.

[20]The average distance between actual stations is 315 meters, hence about 80% of potentially profitable stations are as

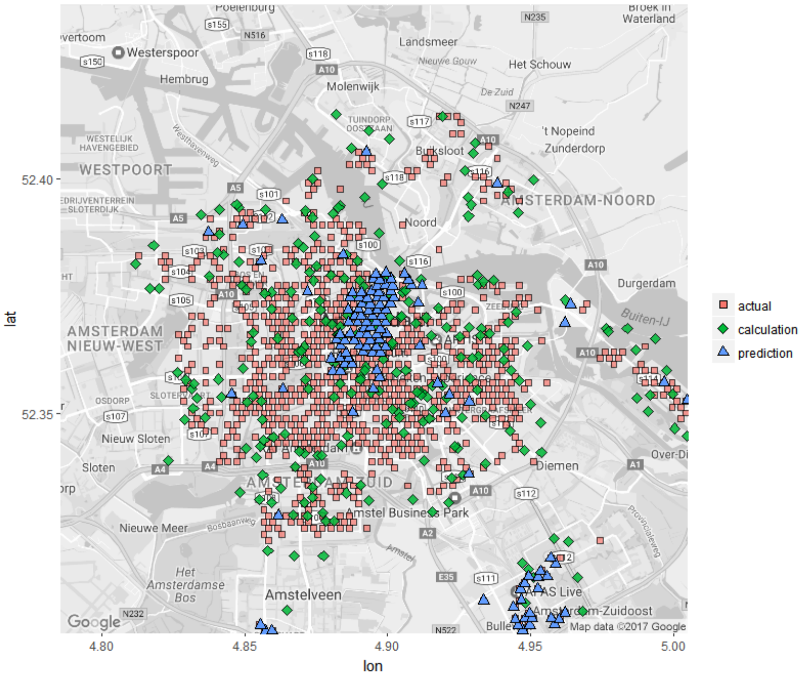

**Figure 11: Actual, calculated and predicted charging stations in Amsterdam**

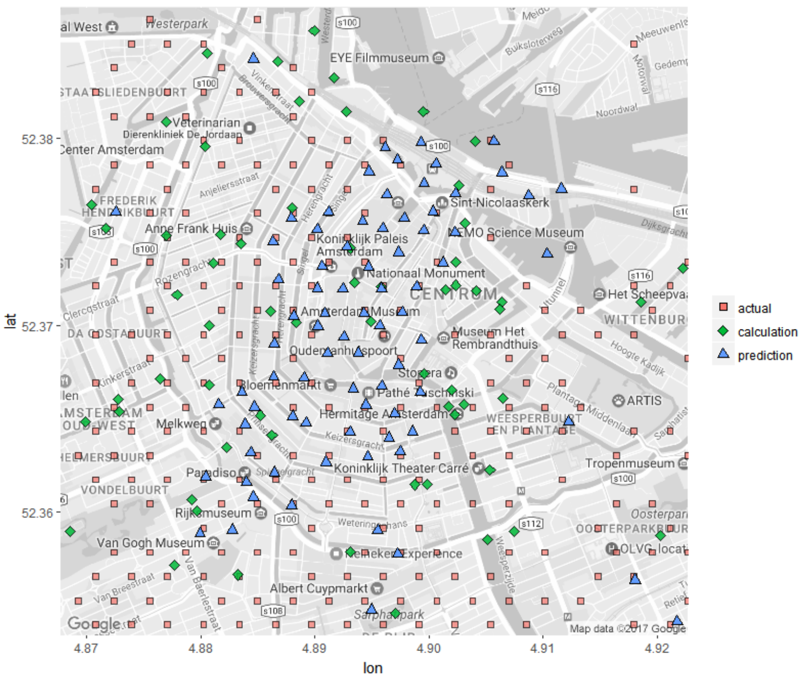

**Figure 12: Zoom in: Actual, calculated and predicted charging stations in Amsterdam**

Figure12 [22] confirms this finding and illustrates the accuracy with which the model predicts charging locations in the city center. Nonetheless, the model also matches numerous calculated charging station locations in the outer parameters of the city. Surprisingly the model also aims to place a few charging stations where no calculated, only actual charging stations are located. Figure 12 confirms this finding when looking at a close up of the city center where various predicted points are closer to actual points than to calculated, optimal locations.

Since the model was trained in four different cities, some profitable, calculated charging station locations may have

reachable as regular stations if misclassified.

[21] Best macro F-measure performing model 5

[22] Points on map repelled from each other with $geom\_jitter()$ to avoid overplotting but which causes the points to be more ordered, hence not fully accurately displayed.

similar characteristics to the actual stations in Amsterdam, hence the model identifies those as optimal while the training locations were not characterized as such initially. It also underlines the need for further research and collection of data when defining optimal (profitable) stations, as it seems that the current calculation has missed some.

## 5. DISCUSSION

The results of the final predictive model clearly show that point of interest information can be used to predict charging locations and can be applied across cities. While the results show a variation in performance for different cities, they are consistent in displaying the ability of the model to detect optimal charging locations in a new city. The predicted locations matched both actual charging station locations and locations which were identified as profitable. The "headstrong" behavior of the Random Forest model shows that the learning method can ingest information on optimal locations and even make up for the shortcomings of the method for calculating optimal charging station locations. Successive research on the topic should, therefore, aim to obtain more information from charging station operators on the profitability of stations to improve the model's understanding and correctness about the target variable. As a significant improvement in accuracy was observed when adding spatial components to the model, it is concluded that extending this information in future models will likely have further improving effects.

The approach and model, however, can already be used by charging station operators, who are looking to expand their portfolio in current cities or to build charging stations in completely new cities, as a business decision support system. Operators can identify interesting locations easily and fast without having to acquire potential location information exclusively through costly surveys, studies, and interviews. The model can help in this case with a pre-selection of potentially interesting locations before further scoping and evaluation.

Especially when expanding to new cities, operators need to be careful about the predictions of the model. As discussed above[16], different weights in importance can be assigned to either the positive identification of optimal locations or the positive identification of non-optimal stations to avoid mispositioning charging points and missing out on charging demand. For this purpose, a model can be optimized by maximizing either the F-micro or macro measure and adjusting $\beta$ in the F-score calculation to rebalance *precision* and *recall*. While the model should not be used as a sole decision-making tool it can still reduce uncertainty and risk when entering new markets.

Another clear advantage of this modeling approach is the transferability across cities. Unlike previous attempts focusing on the prediction of charging demand in kWh, this approach first identifies profitable locations per city, hence scales the strongly varying amounts of kWh demand and only then applies a predictive model. The modeling approach has the potential to improve much further with more information on which stations are confirmedly profitable and personal EV usage patterns.

Carsharing operators can, as they grow, use the model to predict charging station location tailored to the need of their customers. The car-sharing companies could, if a sufficient level of one's own demand is reached, start constructing own stations to decrease cost or partner with a station provider to offer exclusive car sharing charging stations with an ensured profitability for station operators. Overall reducing risk and increasing informedness in investment decision making for car-sharing station placement should reduce the impact of the chicken-egg dilemma and increase EV and car sharing. Finally, car sharing operators could use their real-time data and point of interest information to improve predictions on hourly demand and make use of this information to distribute charging demand across the city for more flexibility and efficiency.

## 6. CONCLUSION AND FUTURE WORK

This study finds significant gains in targeted accuracy for predicting optimal locations of charging stations when adding multiple layers of point of interest information. The approach can be used by charging station operators as a decision support system for expanding the charging station network by reducing risk and speeding up the discovery process. The study found that, for five European cities, potential savings of up to 11% in fixed cost can be achieved with overall prediction accuracies of about 95%. Future research can extend and improve the modeling approach by collecting and adding information on confirmed profitability from charging station operators, private EV parking patterns and further sources of PoI information. When fast charging is introduced for the car sharing companies in focus of this research, further analysis will be needed to create and train models that can predict both fast and slow charging station locations. For potentially even more refined and robust results, deep learning algorithms could be applied to this modeling approach. As such algorithms have proven to be very successful in image and pattern recognition [67], defining the distance between every cell in the city to have an "image-like" representation of the data on a map, could enable deep learning techniques to improve the model further.

## 7. APPENDIX I, TABLES

**Table 10: Points of interest variables used for all cities**

*doctor, car repair, electronics store, church, beauty salon, cafe, art gallery, bakery, book store, electrician, bus station, clothing store, accounting, bar, bicycle store, atm, dentist, car dealer, cemetery, convenience store, department store, car rental, courthouse, embassy, casino, car wash, bowling alley, airport, campground, amusement park, aquarium, point of interest, general contractor, food, health, hair care, plumber, locksmith, lodging, park, funeral home, lawyer, finance, painter, gym, furniture store, jewelry store, city hall, moving company, florist, museum, insurance agency, fire station, local government office, hospital, mosque, night club, parking, library, place of worship, gas station, pet store, laundry, hardware store, liquor store, bank, natural feature, real estate agency, school, store, travel agency, physiotherapist, grocery or supermarket, meal delivery, veterinary care, shopping mall, pharmacy, meal takeaway, university, roofing contractor, police, premise, shoe store, stadium, subpremise, storage, subway station, spa, train station, zoo, home goods store, transit station, restaurant, movie theater, hindu temple*

**Table 11: Most significant influencers of profitable charging stations (linear regression, across all cities)**

| | *Dependent variable:* |
|---|---|
| | profitable_many |
| poi density | $0.001^{***}$ |
| | (0.0004) |
| car_rental | $0.108^{***}$ |
| | (0.024) |
| charging_point | $0.422^{***}$ |
| | (0.012) |
| parking | $0.154^{***}$ |
| | (0.034) |
| university | $0.055^{***}$ |
| | (0.015) |
| bank_adj_r1 | $0.025^{***}$ |
| | (0.008) |
| natural_feature_adj_r1 | $0.294^{***}$ |
| | (0.055) |
| premise_adj_r1 | $0.044^{***}$ |
| | (0.016) |
| bar_adj_r2 | $0.008^{***}$ |
| | (0.003) |
| hair_care_adj_r2 | $0.007^{***}$ |
| | (0.003) |
| parking_adj_r2 | $0.056^{***}$ |
| | (0.010) |
| subway_station_adj_r2 | $0.038^{***}$ |
| | (0.012) |
| distance_city_center | $0.00001^{***}$ |
| | (0.00000) |
| Observations | 14,809 |
| $R^2$ | 0.143 |
| Adjusted $R^2$ | 0.126 |
| Residual Std. Error | 0.704 (df = 14519) |
| F Statistic | $8.412^{***}$ (df = 289; 14519) |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

**Table 12: Table of notation**

| Variable | Description | Unit |
|---|---|---|
| $p_i$ | Parking instance | record |
| id | Unique id of the car | tag |
| lon | Longitude | 0° 00' 0.036" DMS |
| lat | Latitude | 0° 00' 0.036" DMS |
| $fuel_s$ | Fuel level at start $(s)$ | % of battery |
| $fuel_e$ | Fuel level at end $(e)$ of trip | % of battery |
| $t_s$ | Start time $(s)$ of trip | timestamp |
| $t_e$ | End time $(e)$ of trip | timestamp |
| $fuel.adj_{s,p_t}$ | Adjusted fuel level when in lockdown | % of battery |
| $\varnothing\delta_{min}$ | Average time (in minutes) per percent of charge loaded | $\frac{min}{\%fuel}$ |
| $t.adj_{s,p_t}$ | Adjusted time stamp | timestamp |
| $\varnothing\delta_{charg}$ | Energy charged per minute on average | $\frac{\%fuel}{min}$ |
| $fuel_{t_{p_{i-1}},p_i}$ | Battery charge used for the previous trip | % of battery |
| $Q_{d,p_i}$ | Potential charging demand | % of battery |
| $Q_{d,m}(c_{id}$ | Aggregate demand per cell and month | $\frac{\%ofbattery}{month}$ |
| $Q_{d,yr}(c_{id}$ | Aggregate demand per cell and year | $\frac{\%ofbattery}{year}$ |
| $Q_{d,st}(c_{id})$ | Number of stations expected to be profitable per cell | Charging stations |
| $margin$ | Revenue for operator from sale minus purchase price of energy | $\frac{€}{100}$ |
| $profit$ | Profit from charging stataion for operator | € |
| $cost$ | Cost of charging station placement and operations | € |
| $d$ | Distance to city centering | Meters |

# 8. APPENDIX II, FIGURES

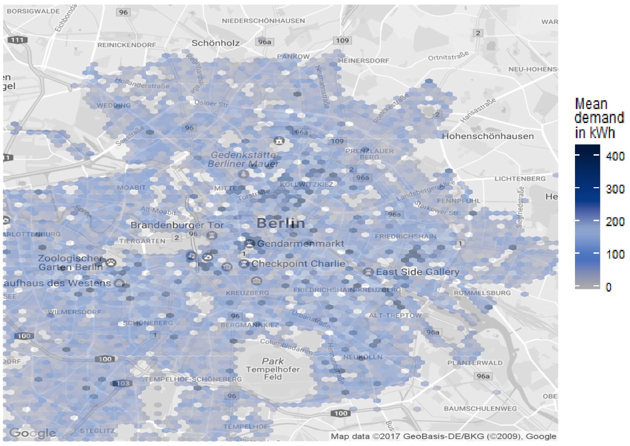


Figure 13: Relative charging demand per hexagonal cell in Berlin

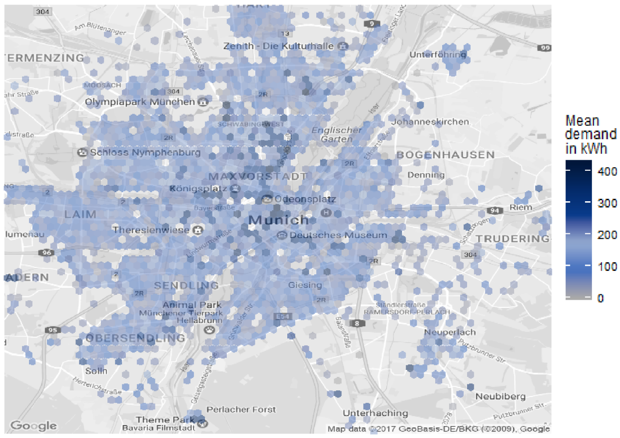


Figure 15: Relative charging demand per hexagonal cell in Munich

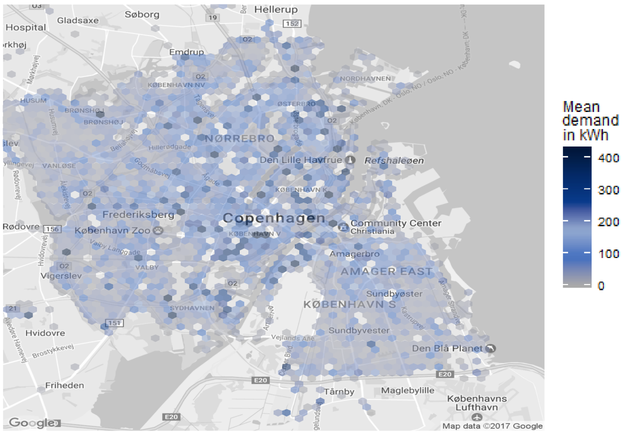


Figure 14: Relative charging demand per hexagonal cell in Copenhagen

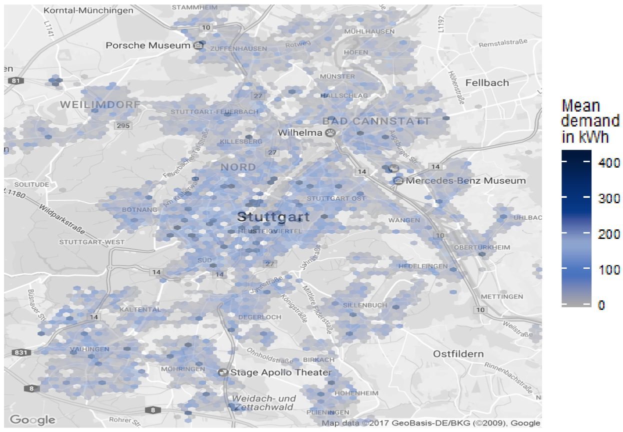


Figure 16: Relative charging demand per hexagonal cell in Stuttgart

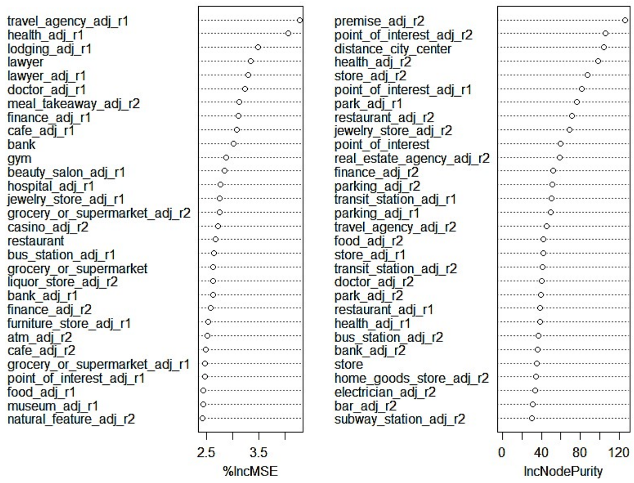


**Figure 17: Variable importance for model without Berlin**

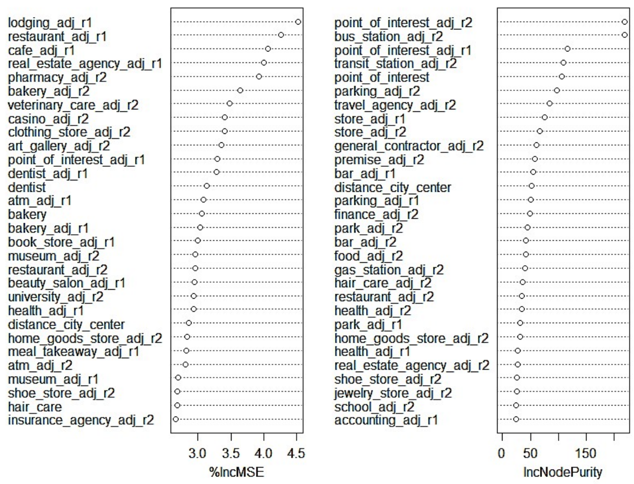


**Figure 19: Variable importance for model without Munich**

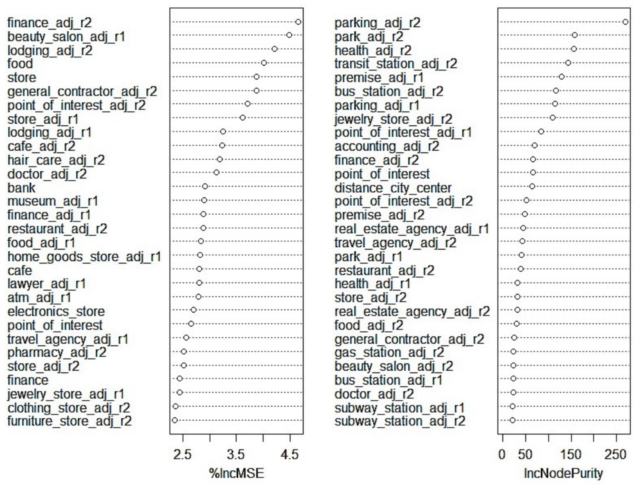


**Figure 18: Variable importance for model without Copenhagen**

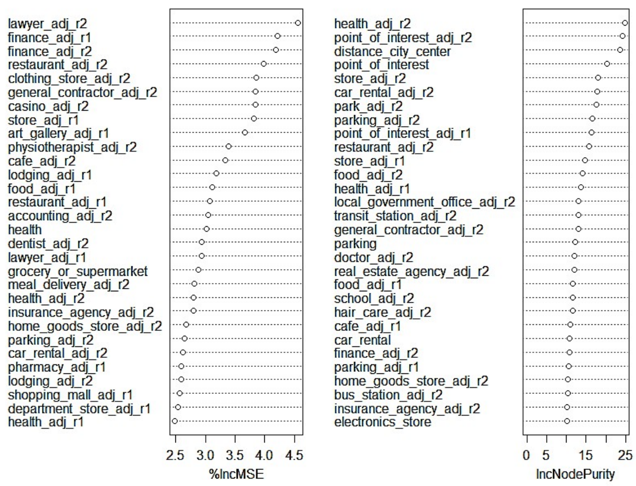


**Figure 20: Variable importance for model without Stuttgart**

## References


[1] Ahn, Y. and Yeo, H. 2015. An analytical planning model to estimate the optimal density of charging stations for electric vehicles. *PLoS ONE.* 10, 11 (2015), 1–26.

[2] Arias, M.B. and Bae, S. 2016. Electric vehicle charging demand forecasting model based on big data technologies. *Applied Energy.* 183, (2016), 327–339.

[3] Bakker, S. 2013. Standardization of EV Recharging Infrastructures. *E-mobility NSR.* December (2013).

[4] Bates, J. and Leibling, D. 2012. Spaced Out Perspectives on parking policy. *RAC Foundation.* July (2012), 132.

[5] Belz, A. et al. 2016. 2016 Car Sharing: A Feasibility Study in Hong Kong. *WPI.* 44-CXP-C16, (2016).

[6] Birch, C.P.D. et al. 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling.* 206, 3-4 (2007), 347–359.

[7] Brady, J. and O'Mahony, M. 2011. Travel to work in Dublin. The potential impacts of electric vehicles on climate change and urban air quality.

[8] Brandt, T. and Neumann, D. 2016. Sharing is Caring – Understanding the Relationship Between the Sharing Economy and Sustainable Mobility. *Icis.* (2016), 1–9.

[9] Breiman, L. 2001. Random Forests. *Machine Learning.* 45, 1 (2001), 5–32.

[10] Brooker, R.P. and Qin, N. 2015. Identification of potential locations of electric vehicle supply equipment. *Journal of Power Sources.* 299, (2015), 76–84.

[11] Chapman, L. 2007. Transport and climate change: a review. *Journal of Transport Geography.* 15, 5 (2007), 354–367.

[12] Charte, F. and Charte, D. 2015. Working with Multilabel Datasets in R: The mldr Package. *The R Journal.* 7, 2 (2015), 149–162.

[13] Chefaoui, R.M. and Lobo, J.M. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling.* 210, 4 (2008), 478–486.

[14] Chen, T.D. et al. 2013. The electric vehicle charging station location problem: a parking-based assignment method for Seattle. *Transportation Research Record.* (2013).

[15] Chen, T.D. et al. 2013. The Electric Vehicle Charging Station Location Problem: A Parking-Based Assignment Method for Seattle. *Proceedings of the 92nd Annual Meeting of the Transportation Research Board in Washington DC, January 2013, and forthcoming in Transportation Research Record.* (2013), 22.

[16] City of Amsterdam 2016. *Public charge update Amsterdam.*

[17] Clean Energy Ministerial 2009. Electric Vehicles Initiative (EVI).

[18] Dong, J. et al. 2014. Charging infrastructure planning for promoting battery electric vehicles: An activity-based approach using multiday travel data. *Transportation Research Part C: Emerging Technologies.* 38, (2014), 44–55.

[19] Egbue, O. and Long, S. 2012. Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. *Energy Policy.* 48, (2012), 717–729.

[20] Esearch, S.Y.R. and Koppius, O.R. 2011. Predictive Analytics in Information Systems Research. *MIS Quarterly.* 35, 3 (2011), 553–572.

[21] Eurostat European cities - spatial dimension - Statistics Explained.

[22] Eurostat Eurostat - Transport in cities and greater cities.

[23] Fortran original by Leo Breiman and Adele Cutler, R. port by and Andy Liaw and Matthew Wiener 2015. Package 'randomForest'. CRAN.

[24] Fournier, G. et al. 2015. Car sharing with electric vehicles: a contribution to sustainable mobility? *Interdisciplinary Management Research.* 11, (2015), 955–975.

[25] Franke, T. and Krems, J.F. 2013. EV RANGE PREFERENCES 1 What drives range preferences in electric vehicle users? 30, (2013), 56–62.

[26] Funke, S. et al. 2015. Placement of loading stations for electric vehicles: No detours necessary! *Journal of Artificial Intelligence Research.* 53, (2015), 633–658.

[27] Gnann, T. et al. 2015. How to address the chicken-egg-problem of electric vehicles ? Introducing an interaction market diffusion model for EVs and charging infrastructure. *Eceee 2015.* June (2015).

[28] Gnann, T. et al. 2015. How to address the chicken-egg-

problem of electric vehicles ? Introducing an interaction market diffusion model for EVs and charging infrastructure. *Eceee 2015*. (2015).

[29] Gopalakrishnan, R. et al. 2016. Demand Prediction and Placement Optimization for Electric Vehicle Charging Stations. *International Joint Conference on Artificial Intelligence*. 2, (2016), 3117–2123.

[30] Goutte, C. and Gaussier, E. 2005. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. *Proceedings of the 27th european conference on advances in information retrieval research*. Springer-Verlag. 345–359.

[31] Graff Zivin, J.S. et al. 2014. Spatial and Temporal Heterogeneity of Marginal Emissions: Implications for Electric Cars and Other Electricity-Shifting Policies. *Journal of Economic Behavior & Organization*. 107, (2014), 248–268.

[32] Guo, S. and Zhao, H. 2015. Optimal site selection of electric vehicle charging station by using fuzzy TOPSIS based on sustainability perspective. *Applied Energy*. 158, (2015), 390–402.

[33] Hastie, T. et al. 2009. The Elements of Statistical Learning. *Elements*. 1, (2009), 337–387.

[34] International Energy Agency 2016. *Global EV Outlook 2016 Electric Vehicles Initiative*. IEA/OECD.

[35] Jeni, L.A. et al. 2013. Facing Imbalanced Data Recommendations for the Use of Performance Metrics. *2013 humaine association conference on affective computing and intelligent interaction* (Sep. 2013), 245–251.

[36] Kamitani, Y. and Sawahata, Y. 2010. Spatial smoothing hurts localization but not information: Pitfalls for brain mappers. *NeuroImage*. 49, 3 (2010), 1949–1952.

[37] Kottek, M. et al. 2006. World map of the Köppen-Geiger climate classification updated. *Meteorolog. Zeitschrift*. 15, 3 (2006), 259–263.

[38] Lam, A.Y.S. et al. 2014. Electric vehicle charging station placement: Formulation, complexity, and solutions. *IEEE Transactions on Smart Grid*. 5, 6 (2014), 2846–2856.

[39] Le Vine, S. et al. 2014. Full-text. *Centre for Transport Studies, Imperial College London*. 22, (2014).

[40] Lehmann, S. 2014. *Low carbon cities : transforming urban systems*.

[41] Madina, C. et al. 2016. Methodology for assessing electric vehicle charging infrastructure business models. *Energy Policy*. 89, (2016), 284–293.

[42] Mascaro, J. et al. 2014. A tale of two “Forests”: Random Forest machine learning aids tropical Forest carbon mapping. *PLoS ONE*. 9, 1 (2014), 12–16.

[43] Mitchell, W.J. et al. 2013. Reinventing the Automobile: Personal Urban Mobility for the 21st Century. *The MIT Press, Cambridge, Massachusetts London, England*. (2013).

[44] Nash, A. 2012. Mobility of the Future. *McKinsey & Company*. (2012), 1–24.

[45] Nationale Plattform Elektromobilitaet 2015. Ladeinfrastruktur für Elektrofahrzeuge in Deutschland - Statusbericht und Handlungsempfehlungen 2015. *Gemeinsame, Geschaeftsstelle Elektromobilitaet der Bundesregierung*. (2015), 36.

[46] Oliveira, S. et al. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*. 275, (2012), 117–129.

[47] Pedersen, J.S. et al. 2012. Bringing the electric vehicle to the mass market a review of barriers , facilitators and policy interventions. *Working Paper*. (2012), 1–77.

[48] Pojani, D. and Stead, D. 2015. Sustainable Urban Transport in the Developing World: Beyond Megacities. *Sustainability*. 7, 6 (2015), 7784–7805.

[49] Prasad, A.M. et al. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*. 9, 2 (2006), 181–199.

[50] Rupprecht Consult Forschung Beratung GmbH 2013. Prospects for electric car sharing and use in small and medium cities and towns. *ELMOS*. (2013).

[51] Sathaye, N. and Kelley, S. 2013. An approach for the optimal planning of electric vehicle infrastructure for highway corridors. *Transportation Research Part E: Logistics and Transportation Review*. 59, (2013), 15–33.

[52] Schroeder, A. and Traber, T. 2012. The economics of fast charging infrastructure for electric vehicles. *Energy*

*Policy.* 43, (2012), 136–144.

[53] Shaheen, S. and Cohen, A. 2008. Growth in Worldwide Carsharing: An International Comparison. *Transportation Research Record: Journal of the Transportation Research Board.* 1992, May (2008), 81–89.

[54] Shaheen, S. and Cohen, A. 2016. Innovative Mobility Car Sharing Outlook Worldwide Carsharing Growth Continues. *Transportation sustainability research center - University of California, Berkeley.* (2016).

[55] Sierzchula, W. et al. 2014. The influence of financial incentives and other socio-economic factors on electric vehicle adoption. *Energy Policy.* 68, (2014), 183–194.

[56] Sober, E. 2002. Instrumentalism, Parsimony, and the Akaike Framework. *Philosophy of Science.* 69, S3 (2002), S112–S123.

[57] Sokolova, M. and Lapalme, G. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing and Management.* 45, (2009), 427–437.

[58] Sokolova, M. and Lapalme, G. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management.* 45, 4 (2009), 427–437.

[59] Tagesschau.de 2017. Autogipfel in Stuttgart: Zu wenig, zu langsam, zu spät | tagesschau.de.

[60] Tastle, W.J. 2013. *Data Mining Applications Using Artificial Adaptive Systems.* Springer New York.

[61] United Nations 2014. Department of Economic and Social Affairs. *United Nations.* ST, (2014), 32.

[62] Untermann, R.K. and Lewicki, L. 1984. *Accommodating the pedestrian : adapting towns and neighborhoods for walking and bicycling.* Van Nostrand Reinhold.

[63] Valverde-Albacete, F.J. and Peláez-Moreno, C. 2014. 100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox. *PLoS ONE.* 9, 1 (2014).

[64] Van Brummelen, G. 2013. *Heavenly mathematics : the forgotten art of spherical trigonometry.* Princeton University Press.

[65] Van Der Goot, D. 1982. A model to describe the choice of parking places. *Transportation Research Part A: General.* 16, 2 (Mar. 1982), 109–115.

[66] Vannieuwenborg, F. et al. 2014. Deploying charging infrastructure for electric vehicles; viability analyses for municipal and private car parking facility operators. *European Journal of Transport and Infrastructure Research.* 4, 14 (2014), 425–448.

[67] Vincenzi, S. et al. 2011. Application of a Random Forest algorithm to predict spatial distribution of the potential yield of Ruditapes philippinarum in the Venice lagoon, Italy. *Ecological Modelling.* 222, 8 (2011), 1471–1478.

[68] Wagner, S. et al. 2016. In free float: Developing Business Analytics support for carsharing providers. *Omega (United Kingdom).* 59, (2016), 4–14.

[69] Wagner, S. et al. 2014. In Free-Float: How Decision Analytics Paves the Way for the Carsharing Revolution. *ICIS 2014 Proceedings.* (2014), 1–17.

[70] Wagner, S. et al. 2015. Data Analytics for Location-Based Services: Enabling User-Based Relocation of Carsharing Vehicles. *Business & Information Systems Engineering.* 3, 5 (2015), 279–287.

[71] Wallis, N. et al. 2013. Electric vehicles: Improving consumer information to encourage adoption. (2013), 1–14.

[72] Xi, X. et al. 2013. Simulation-optimization model for location of a public electric vehicle charging infrastructure. *Transportation Research Part D: Transport and Environment.* 22, (2013), 60–69.

[73] Yuksel, T. et al. 2016. Effect of regional grid mix, driving patterns and climate on the comparative carbon footprint of gasoline and plug-in electric vehicles in the United States. *Environmental Research Letters.* 11, 4 (2016), 044007.

[74] Zhang, M.-L. and Zhou, Z.-H. 2014. A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering.* 26, 8 (Aug. 2014), 1819–1837.