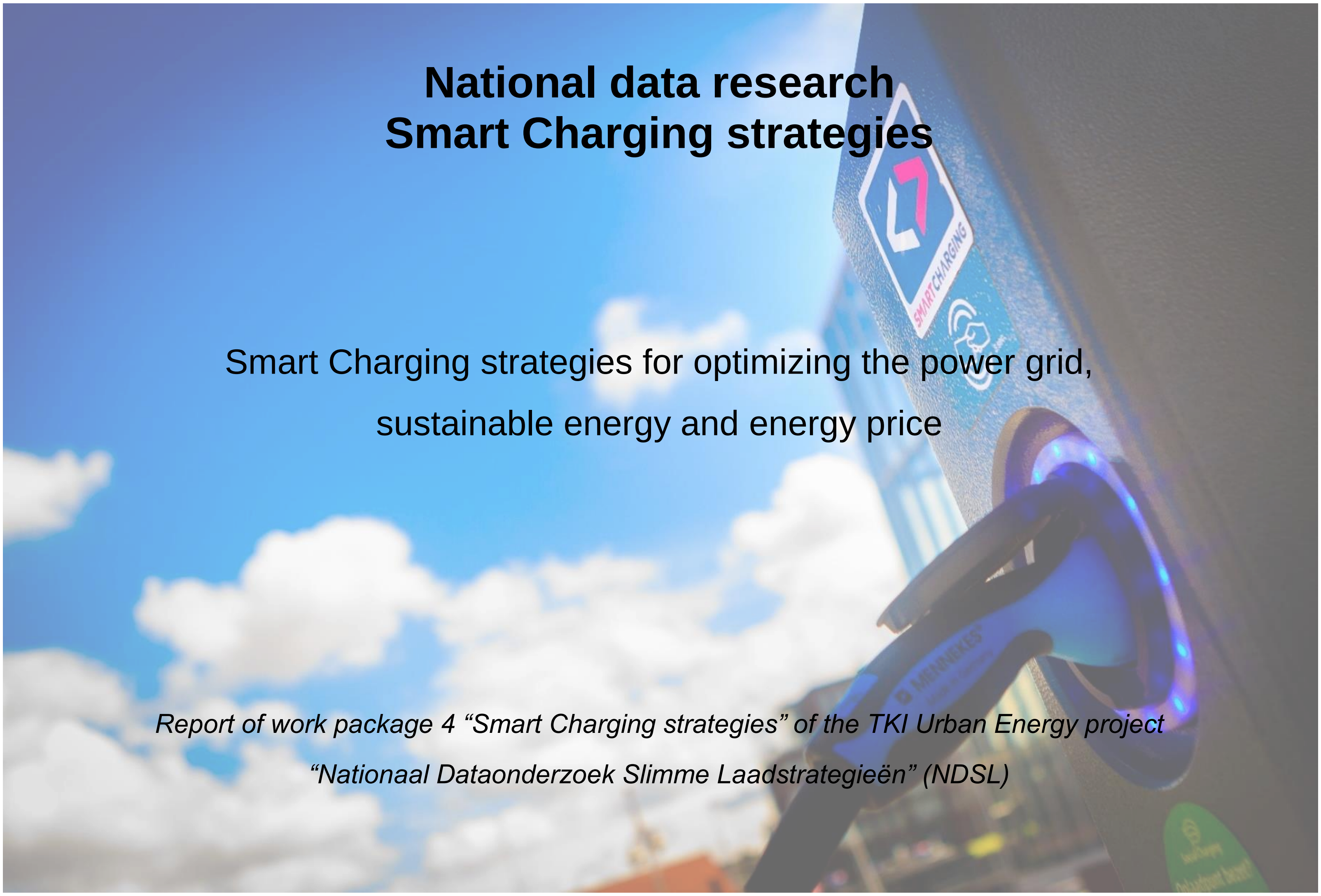


National data research
Smart Charging strategies

Smart Charging strategies for optimizing the power grid, sustainable energy and energy price

Report of work package 4 “Smart Charging strategies” of the TKI Urban Energy project “Nationaal Dataonderzoek Slimme Laadstrategieën” (NDSL)

Authors:

Youssef El Bouhassani (AUAS)
Jeroen Groot (AUAS)
Peter van Bokhoven (ElaadNL)
Ruud Noordijk (ElaadNL)
Nazir Refa (ElaadNL)

Table of contents

Contents

Acknowledgement

The authors highly appreciate the data and input provided by the municipalities of Amsterdam, Rotterdam, The Hague and Utrecht, MRA-Elektrisch, ElaadNL, Living Lab Smart Charging.

We also thank the consortium members for their collaboration and feedback (Municipality of Amsterdam, MRA-Elektrisch, Pitpoint, Enpuls, Amsterdam University of Applied Sciences, ElaadNL, Living Lab Smart Charging).

We kindly acknowledge the financial support of TKI Urban Energy for carrying out this research within the framework of the National Data Research Smart Charging project.

1. Context and background information

In this section the context of the project is described. First the motivation for Smart Charging will be discussed based on previous research on Smart Charging Potential. Next the different types of Smart Charging strategies will be described. This section is concluded by describing the project objectives and the data used for this research.

1.1 Why Smart Charging?

Electric driving is expected to be one of the key drivers in reducing CO2 emissions. And as the number of FEVs increases, the demand for electricity for charging EVs will increase as well. Alongside FEVs, there are also almost 100k PHEVs in the Netherlands with smaller electricity consumptions. Due to (severely) downsized subsidies for PHEVs, sales of PHEVs have dried up and including export of occasions this PHEV base in the Netherlands is shrinking.

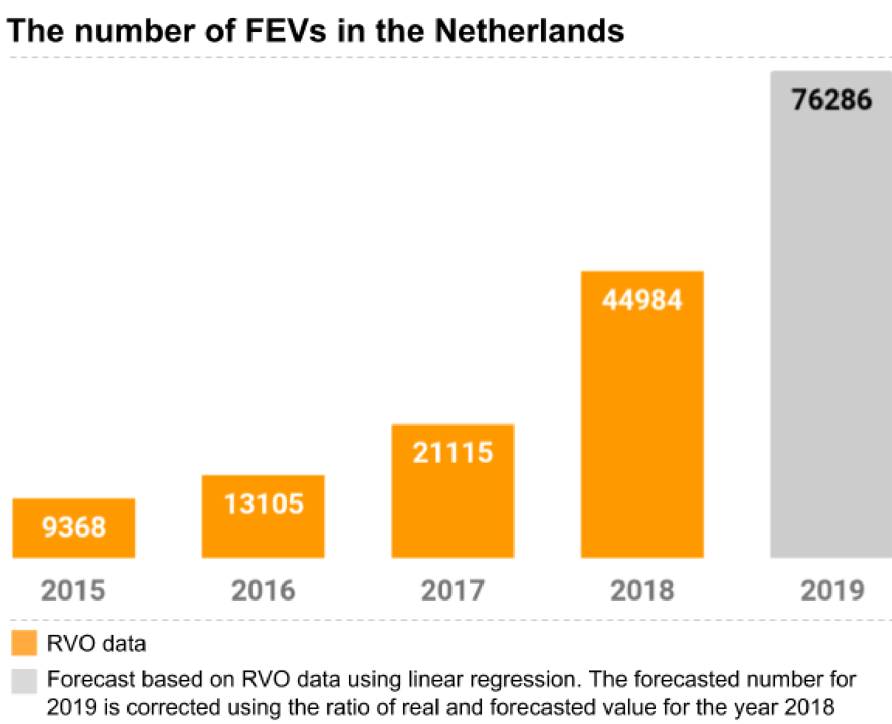


Figure 1: The number of FEVs in the Netherlands[1]

[1] Data from: www.rvo.nl/onderwerpen/duurzaam-ondernemen/energie-en-milieu-innovaties/elektrisch-rijden/stand-van-zaken/cijfers

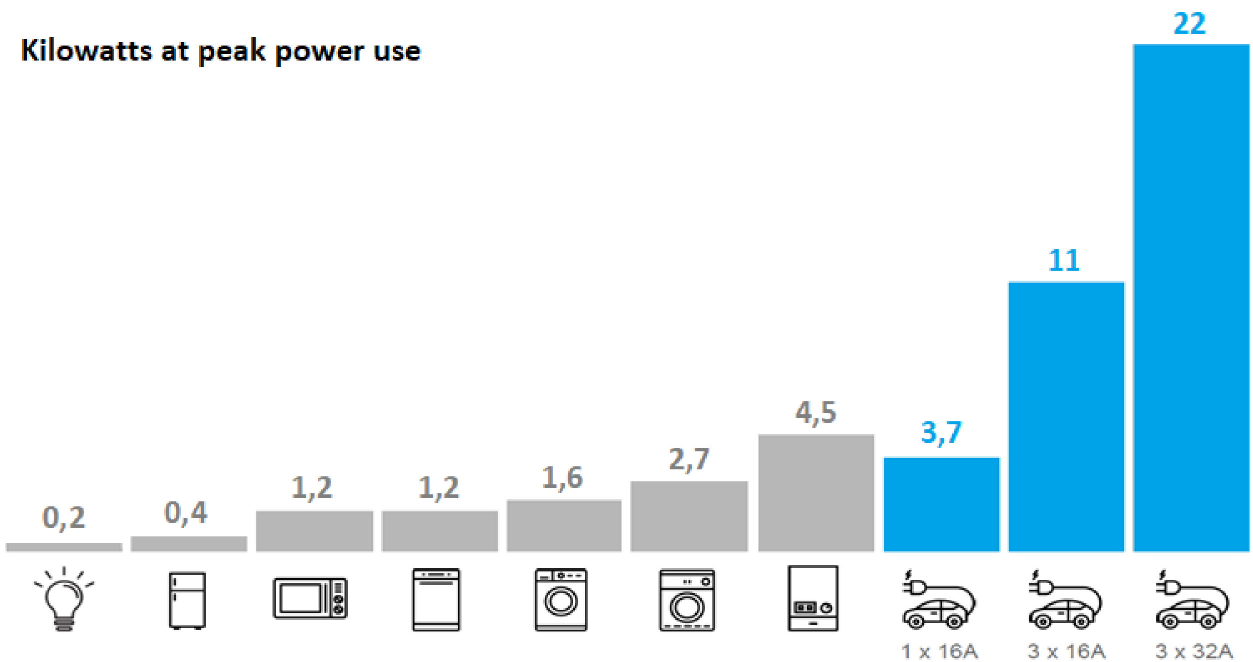


Figure 2: Kilowatts at peak power use of household appliances and EVs

Charging EVs will have a big impact on the grid. Demand for electricity comes from different sources (companies, factories, households, etc.). If we compare charging EVs to typical household electricity demand from different household appliances, you can see that the peak power use of an EV with for example a 3-fase, 16A or 32A charging system is way higher (11 - 22 kW) than all appliances combined.

The demand for electricity required to charge EVs will increase significantly, as the vast majority of the Dutch households currently does not have an EV and is expected to acquire one (or multiple) when the transition unfolds. Given the current charging patterns (which we will zoom in on later in this report), the increased electricity demand will be concentrated in the morning and the evening. This increased demand coincides with the same peak moments for household electricity consumption, especially the evening peak roughly between 17.00h and 20.00h. Moreover, there are other significant developments in the field of the energy transition that have a serious impact on the low-voltage electricity grid. The adaptation of heat pumps (4,5 kW) and solar panels (for example 10 panels of 300 Wp account for 3 KW). The combined developments account for a peak impact that the low-voltage electricity grids are not designed for. In this report we will explain how Smart Charging can help balancing those grid peak loads.

Another important reason to use Smart Charging is to use renewable energy in an optimal way. As figure 3 shows, most electricity generated by wind and solar is generated during the day, when the sun shines and the wind blows the strongest. In a 2030 scenario[2] where an estimated 50% of our total electricity supply comes from sun and wind, balancing supply and demand may become a major challenge. We don't want extra fossil plants running when renewable supply is low, or stopping windmills (curtailment) when supply is high. One way to help solving this problem is using the flexibility in charging EVs through Smart Charging.

[2] https://www.ce.nl/publicaties/1912/power-to-ammonia-energy-and-electricity-prices-scenarios-2020-2023-2030

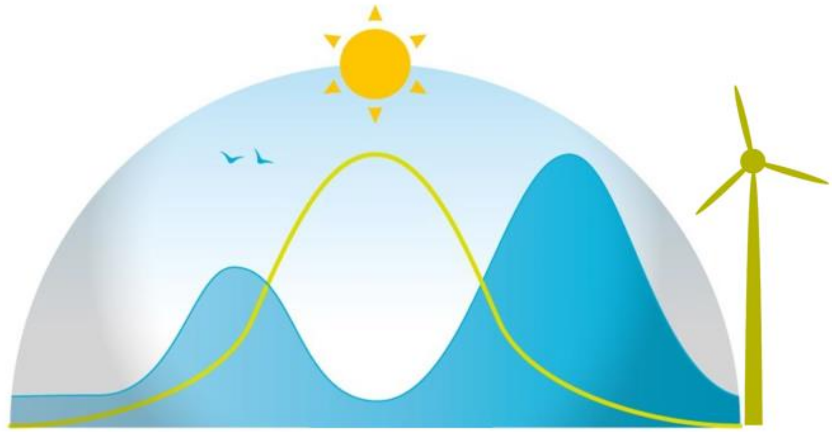


Figure 3: Simplified graph of solar and wind (green) generation and household power use (blue)

A third reason to use Smart Charging is optimization via electricity prices. When renewable power supply is increasing, price fluctuation will also increase because of generation fluctuation. Windmills and solar panels don't generate electricity at the steady rate of a fossil fuel plant. By using Smart Charging combined with flexible energy prices, EV-drivers could be able to charge their car at different moments, enabling lower prices.

1.2 Types of Smart Charging strategies

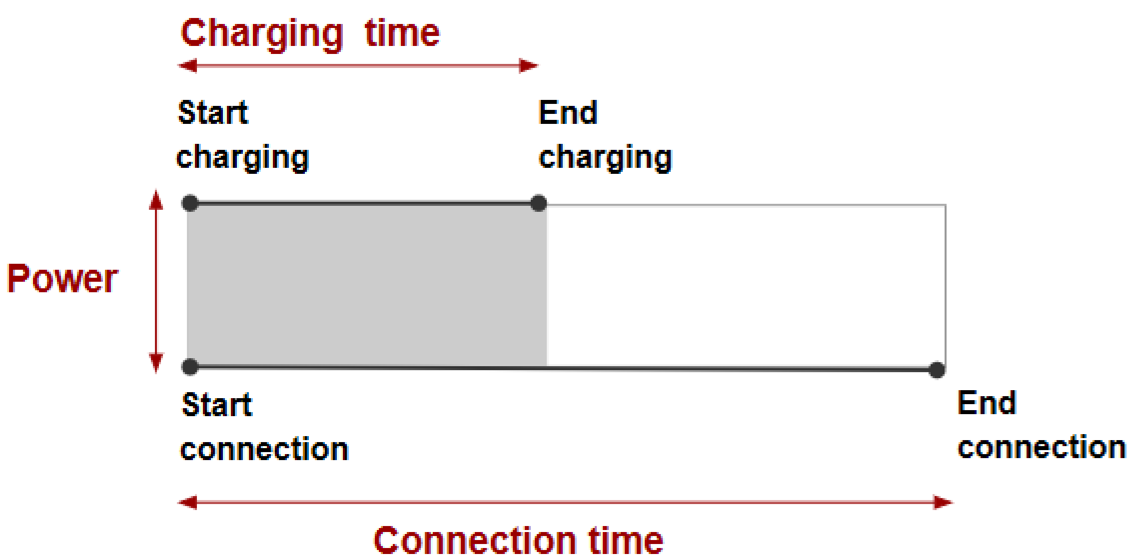


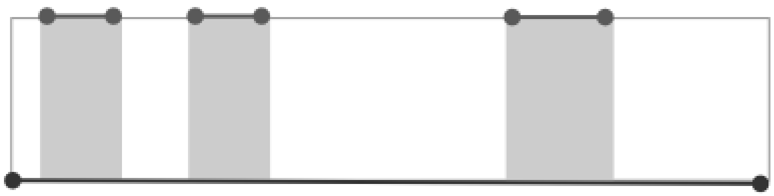
Figure 4: Parameters of a charging session

A charging session is characterized by three important parameters. The first one is the connection time, which is the time interval between starting and ending a connection (to a charging station). Second is the charging time, which is the time an EV is actually charging. Often the connection time is longer than the charging time, which means that the charging session provides flexibility to shift a charging session to a later moment. Third is the power, which corresponds with the height of the rectangle in the figure. The higher the rectangle the higher the power and hence the charging speed. The area of the grey rectangle (Power * Time) is the amount of electricity consumed to charge an EV. This representation can be used to illustrate the different smart charging strategies.

1. **Postpone strategy**

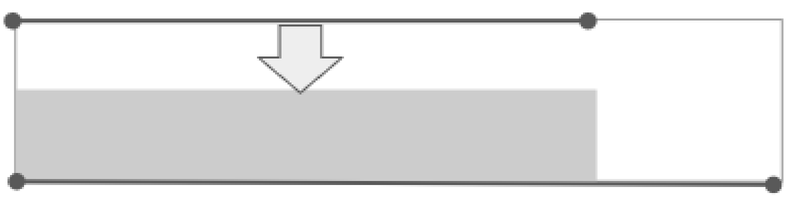
A charging session is shifted completely to start at a later moment. Finding the optimum postpone strategy corresponds with finding the optimal shift of individual sessions. To define what is optimal, we use the performance indicators (see next page)

2. **Cut-and-divide strategy**



A charging session is split up in a set of smaller sessions and distributed within the connection time. Optimizing the cut-and-divide strategy is about finding the optimal number of cuts and intervals between each sub-session.

3. **Slower charging strategy**



Reducing the charging speed means that the charging time is increased (within the connection time) in order to charge an EV to the same level. Optimizing the charging speed is about finding the ideal charging speed such that the overall power demand is within the desired level while at the same time meeting the demands for charging EVs.

4. **Hybrid charging strategy**

In an ideal situation we would like to combine the different strategies and decide what to do given the optimization variables. This approach is dynamic and can take changing variables into account which influence the optimal condition for the individual user. This would require Vehicle-to-Grid (V2G) technology compatibility as well as combining all sorts of strategies together. Due to complexity, V2G is out of scope for this research.

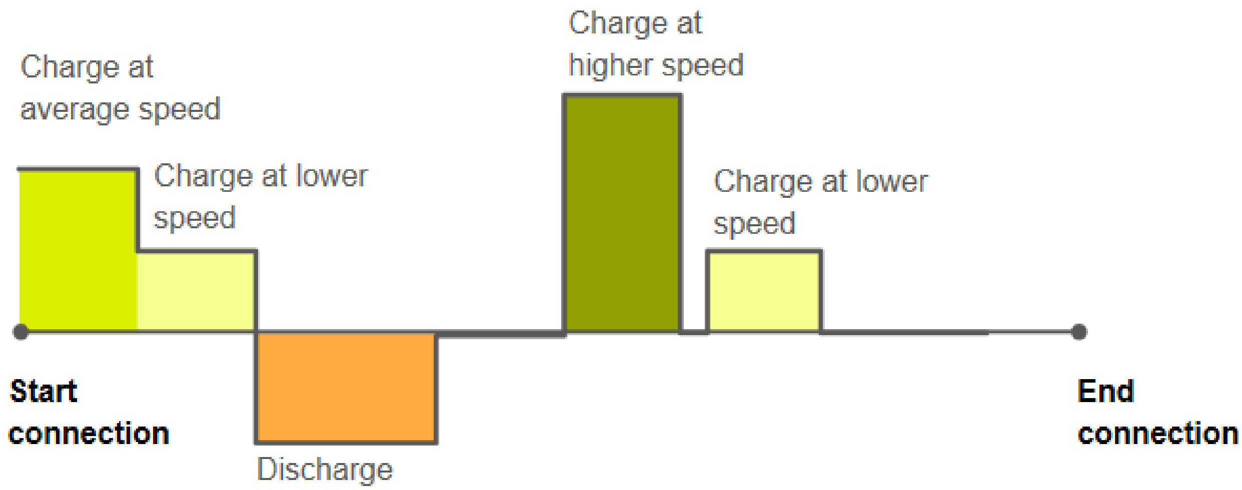


Figure 5: concept of a hybrid charging strategy

For now we focus on the postpone strategy, and we expect to add cut-and-divide and slower charging to our models in 2019.

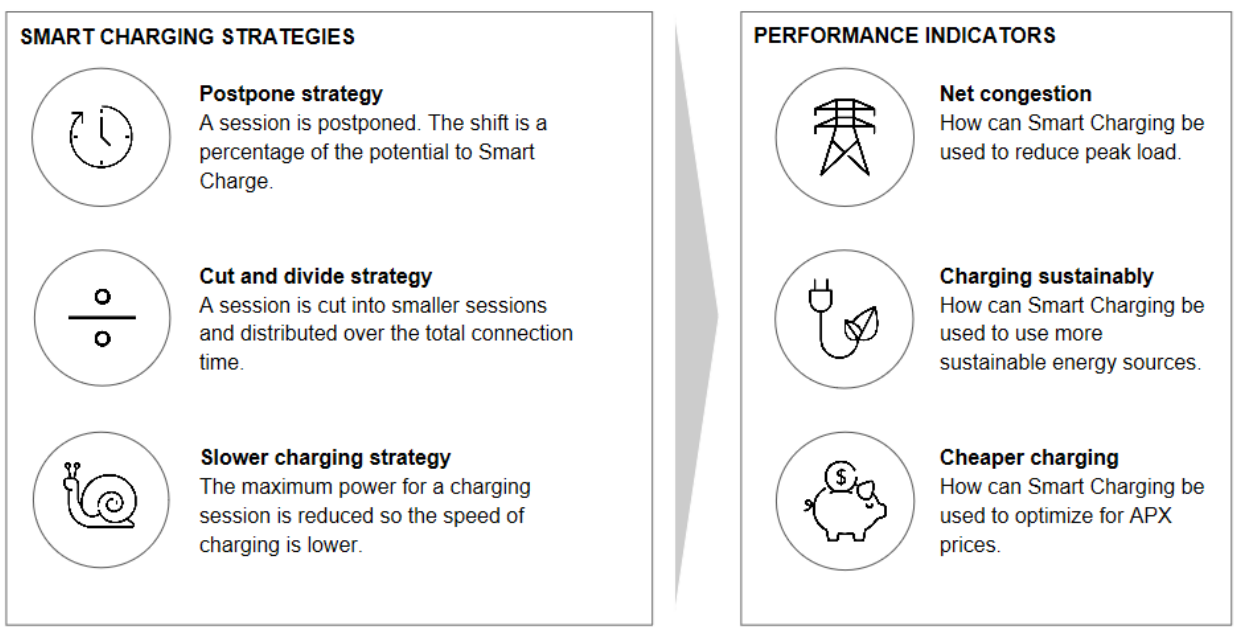


Figure 6: Smart Charging strategies and performance indicators

The goal of this research is to optimize Smart Charging strategies and measure the impact of applying Smart Charging using three types of performance indicators (figure 6). Why these 3 indicators are important is explained in the previous paragraph. (1.1: “Why Smart Charging”). Note that the current optimization goals do not explicitly mention user and user acceptance of smart charging. One way how this is taken along is by considering a overrule function for users not interested in smart charging. In future work optimization strategies could be applied within particular upfront limits, e.g. minimum state of charge or minimum electric mileage that needs to be fulfilled before smart charging can take place.

1.3 About the NDSL project

This report is the result of the research done for work package 4 (Smart Charging strategies) of the National Data Research of Smart Charging Strategies (in Dutch: Nationaal Dataonderzoek Slimme Laadstrategieën; hereafter: NDSL). The goal of the NDSL project is to combine the two main databases with charging sessions on public charging stations in The Netherlands and analyse this data to explore the potential of Smart Charging strategies, analyse interregional charging behavior, develop protocols for researchers to use this data and develop charging strategies for fleet-owners. The project consortium consists of Amsterdam University of Applied Sciences (project secretary), ElaadNL, Living Lab Smart Charging, Enpuls, Pitpoint, Municipality of Amsterdam, Deudekom and MRA-Elektrisch.

1.4 Description of the dataset

The dataset used in this project provides around 90% of all the public charging data in almost the entire country. There are some areas where charging data is missing (the black parts in figure 7), but in general the dataset offers more than enough relevant data for our research.

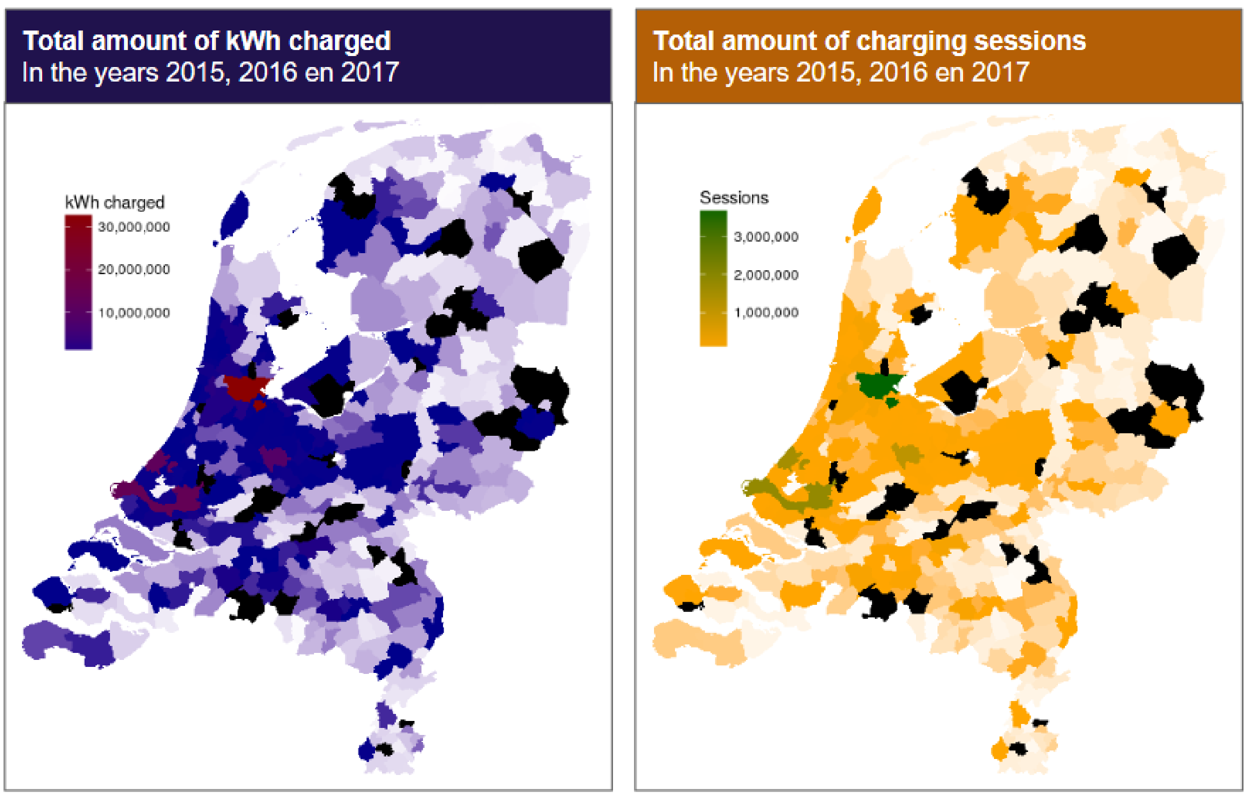


Figure 7: Amount of kWh charged and total amount of charging sessions based on the used dataset

As shown in figure 7, the charging density in the “Randstad” (Amsterdam, Rotterdam, The Hague, Utrecht and surrounding areas) is significantly higher than in the rest of the Netherlands. This is mainly because it is a densely populated area, and because of that a relatively bigger part of all public charging infrastructure and EVs is located and charged in that area.

12 million	101 million	118 million	20 million	83 thousand	7 thousand
sessions	kWh charged	hours connected	hours charging	active RFIDs	charging stations

The dataset spans three years:

YEAR	SESSIONS	TOT. KWH	TOT. CT	TOT. LT	RFIDs	TOT. CP
2015	2,703,359	23,906,101	25,239,971	4,984,990	48,143	4,429
2016	4,411,538	35,555,324	41,676,230	7,452,944	70,437	6,028
2017	5,030,872	41,753,752	51,313,039	8,319,250	83,171	6,850
TOTAAL	12,145,769	101,215,177	118,229,240	20,757,184	83,171	6,850

Explanation
TOT. KWH = The amount of kWh charged in a given year.
TOT. CT = The amount of hours connected in a given year.
TOT. LT = The amount of hours charging in a given year.
TOT. CP = The amount of charging stations used in a given year.

Figure 8: dataset key figures (2015-2017)

The dataset of charging sessions used in this research is for the years 2015, 2016 and 2017. Using data that spans multiple years provides us with the opportunity to research trends through time.

In addition to the charging sessions data, we need other data sources to assess the impact of smart charging. The three indicators we use to measure the impact of Smart Charging are net congestion, renewable energy usage and energy price. Net congestion is taken into account by utilizing grid data from the grid operators. In 2019 we will implement real time monitored grid data into the model. The sustainable energy data is acquired from the ENTSOE[3] platform. Finally, the electricity prices are taken into account by using the Amsterdam Power Exchange (APX) prices.

[3] transparency.entsoe.eu

2. Modelling approach

2.1 Model pipeline description and motivation

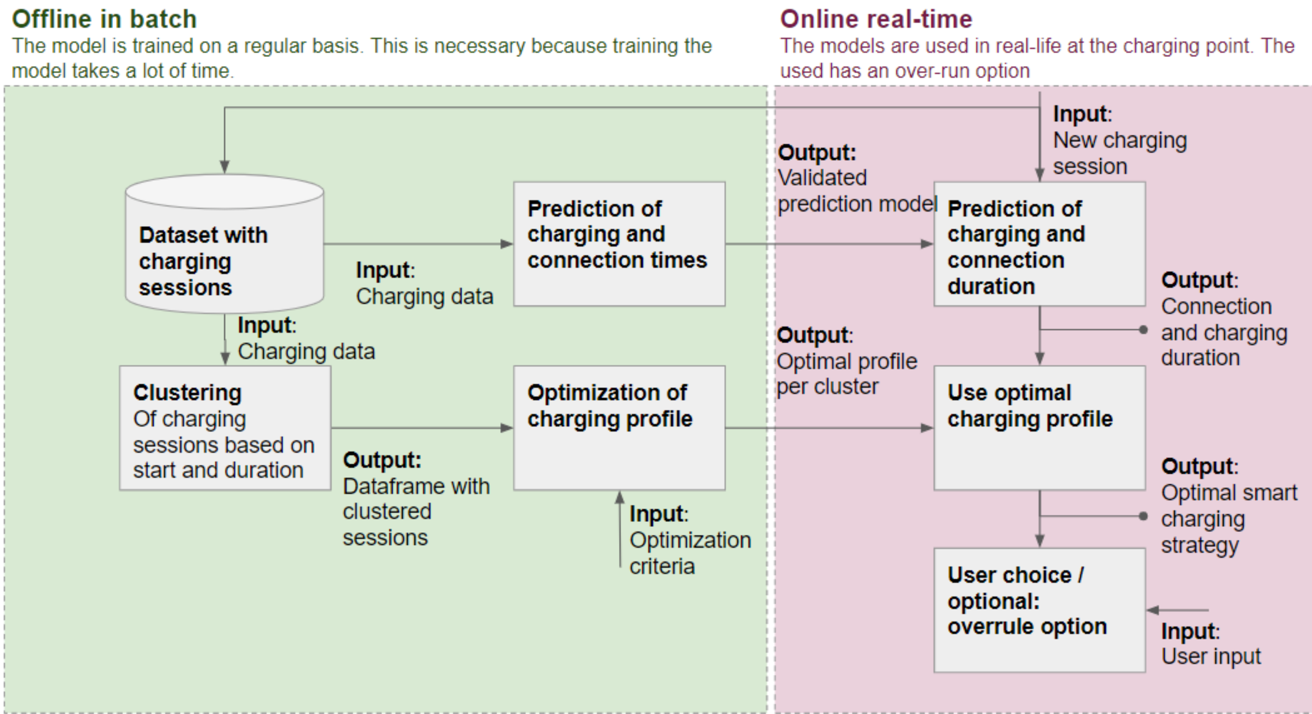


Figure 9: Model pipeline of the used model

Developing Smart Charging strategies consists of a number of steps. In this project we have defined a model pipeline as shown in this picture. A complete models consists ideally of two parts. An offline model which is trained in batch. Historical data is used to train a model on a regular basis. This model is refined online during the day as new charging sessions take place. In this project we focus on developing the offline model part which consist of three main steps.

Firstly, it is necessary to be able to predict the charging requirements of a new session. When a new charging sessions starts, we want to know what the connection time and the expected charging time. Based on these two variables we know what the potential is for smart charging.

Secondly, it is important to find a cluster in the charging sessions. Not all charging sessions are the same. Depending on location, time and usage, a charging session will have different characteristics. Previous research[4] has shown that clusters of charging sessions can be found on the basis of two variables: the time at which a charging session starts and the connection duration. Based on these two variables, we can label the different clusters.
Clustering of charging sessions is important for the optimization step that follows. Each cluster of sessions can be considered as a homogeneous group that has the same characteristics and optimum for a given smart charging strategy. Of course there are variations within each group. At a later stage, the clusters can be refined and define different optima for the sub-groups. For the time being this is outside of the scope of this research.

[4] In this blogpost, a methodology is proposed to quantify the potential of Smart Charging. http://www.idolaad.nl/gedeelde-content/blogs/youssef-el-bouhassani/2018/pinpointing-the-smart-charging-potential.html

Thirdly, it is important to determine what the best potential smart charging strategies per cluster are. This is done based on historical charging data and an optimization criterion. Optimization criteria for reducing grid load and using more sustainable energy can be different and therefore we use different optimization cost functions. Optimization will result in the optimal charging strategies per cluster. The moment a new session is assigned to a cluster, we know immediately what the optimum is for this charging session.

Finally, for practical application it may be desirable to give the user the possibility to overrule Smart Charging. Even though we only use the available flexibility of the users charging session in our model, it is possible that in unexpected situations the proposed Smart Charging optimization doesn't fit the users needs at that specific moment. In our model it is possible to build in an overrule option for the user (for example by using a mobile phone app or a different type of charging card), and execute the charging session without Smart Charging.

2.2 Prediction of connection times and charging times

As mentioned, it is important to be able to predict the charging requirements of a new charging session. When the session starts we want to give an accurate prediction of the connection time and charging time. To predict the connection time of a charging session we tried different methods and models of increasing complexity:

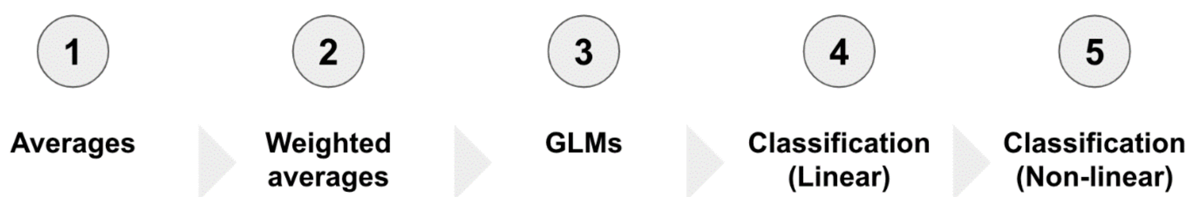


Figure 10: methods that can predict connection and charging times

2.2.1. Averages

We started with the method of calculating averages (method 1, Figure 10) from past sessions in our database.

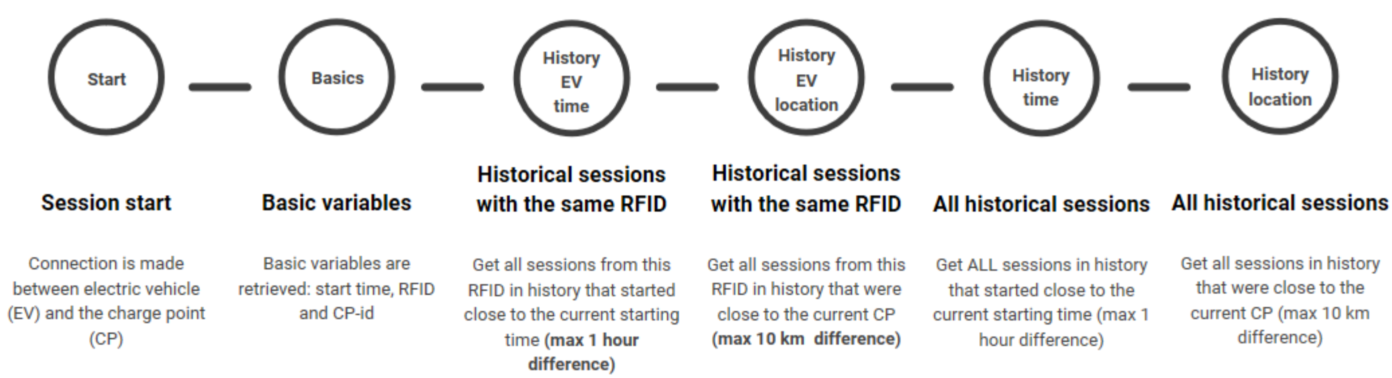


Figure 11: used steps for the "Averages" method

As a session starts we have only 3 input variables: the RFID, the starting time of the session and the charging point. Next to that we have a database with several million sessions. We believe that historical charging sessions will tell us something about charging sessions that will happen in the future.

We want to calculate averages on several axes. The first one is the time axis. We assume that people go to work and come back from work at more or less the same time. So, we calculate the average transaction time from historical sessions with the same RFID (from the same user) around the start time. The max difference in minutes is set to 60, this means that historical sessions where the difference in minutes to the current starting time is greater than 60 are not taken into account. We could do this again with the distance to the current charge point. We have the coordinates from charge points and can calculate the distance to the current charge point. If the difference is greater than 10km, it is not taken into account. Next to the averages from the same RFID, we could also check what all the other chargers have done in the past around the current starting time and charge point. If we calculate all these averages and add them together, we have the new transaction time for the current session. The results (R-squared 0.15 – 0.22, average error 2.1 – 4.8 hours) are not satisfying for real life implementation. This approach is too general and certain patterns are averaged out, which is something we don't want.

2.2.2. Weighted averages

In the previous method we treated every historical session as equally important, which is limited because sessions from multiple years ago are unlikely to tell us much about tomorrow and a session from yesterday is probably way more representative for the future. This is why we added weights to sessions to provide that level of importance. The weights can be added in 3 ways: time, distance and history. As shown below:

	What	Why	How
	Time	If we have a new session at 15.00, the historical sessions closest to 15.00 are more important than historical sessions that are far off.	$\begin{cases} 1 - \frac{abs(\Delta t)}{60}, if\ abs(\Delta t) < 60 \\ 0,\ if\ abs(\Delta t) \geq 60 \end{cases}$
	Distance	If we have a new session at a certain charge point, historical sessions close to that charge point are more relevant than sessions that are not	$\begin{cases} \frac{(\Delta d - 10\,000)^4}{10\,0000^4}, if\ \Delta d < 10\,000 \\ 0,\ if\ \Delta d \geq 10\,000 \end{cases}$
-1Y	History	Recent sessions are more relevant than sessions from a few years back. The charging behaviour changes over time.	$\begin{cases} 1 - \frac{\Delta ht}{100}, if\ \Delta ht < 100 \\ 0,\ if\ \Delta ht \geq 100 \end{cases}$

Figure 12: different weight functions to determine the importance of historical charging sessions

For "Time": within the 60 minutes that we set as boundary, we assumed that a session that has a difference in starting time of 1 minute is more important to take into account than a sessions that has a difference of 59. The same principle is used for the distance weight. If somebody charges every day at the same charge point and once a month at a chargepoint 8 km away, that outlier is weighed as less important. As stated earlier, sessions from 5 years ago are weighed as less important than sessions from yesterday. The way we applied these weights is to use a formula that adds a weight to all our sessions (within the boundaries)

The results are now slightly better (R-square of 0.25-0.31, average error 1.6 – 4.1 hours) but still not satisfying. The averaging approach seems to be too general, even with weights added to it. Taking averages does not always distinguish different charging behaviors, and the correlation between variables is not taken into account.

2.2.3. Generalized Linear Models (GLMs)

To model charging behavior we will start with regression. Regression analysis is a set of statistical processes for estimating the relationships among variables. Ordinary linear regression would predict the the transaction time as a linear combination of a set of our variables. However, this implies that a constant change in a variable (feature) leads to a constant change in the transaction time (i.e. a linear-response model). However, this is possibly not the case when we try to model charging behavior. Therefore we use Generalized Linear Models (GLMs), they allow response variables that have arbitrary distributions[5]. The following image shows a visualization of the difference between the ordinary and generalized linear model.

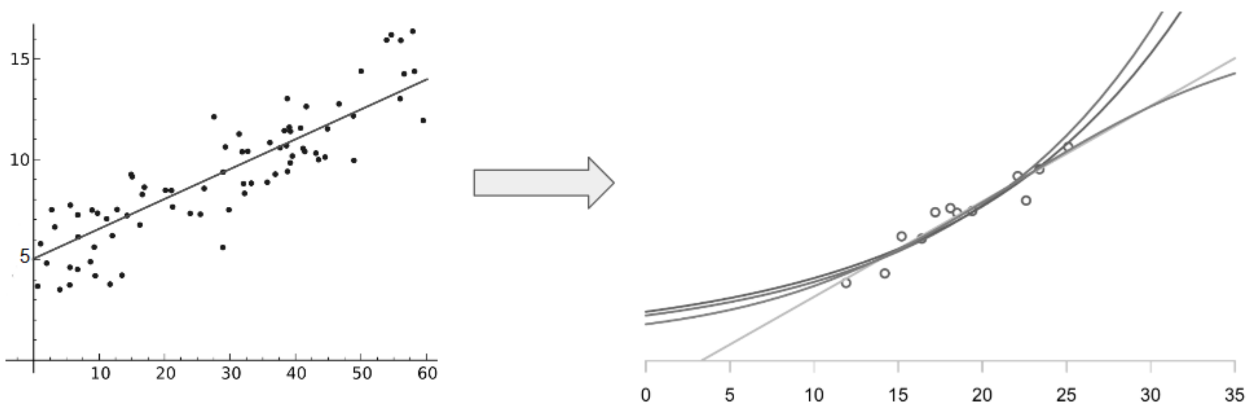


Figure 13: the difference between linear models and generalised linear models

On the left you see a linear regression model in place (basic example). Although this is just an example, the x-axis could be the minutes of a hour and the y-axis could be the transaction time. The dots in this image then represent a transaction time of a session from the past with a certain starting minute. We try to draw a line through these points such that it is the closest to all the points (i.e. minimize the squared error of all the points). If we now have a new data point, we fill in the values for the x-axis (starting minute) and get the new transaction time (value on y-axis). In an ordinary linear model our squared error is quite high because there is no good fit (we can only draw straight lines here). Because GLMs allows for more distributions we can fit a better line to the points (we can draw curved lines now). The outcome of using this in our model resulted in an R-square of 0.49 and an average error of 1.6-3.6 hours. Still, the prediction of a transaction time minute by minute seems very hard and is maybe not even necessary. Rather than predicting accurate connection times, we decided to use classes instead.

2.2.4. Classification (linear)

Linear classification was applied next; first with a relatively simple starting point. We took 2 classes: greater or smaller than 6 hours of connection time. We could use 4 or 5 hours here, but

[5] https://en.wikipedia.org/wiki/Generalized_linear_model

the assumption is that with 6 hours we have a reasonable smart charging potential to work with in our optimization. The average charging time is about 4-5 hours, so there is not much potential there to optimize.

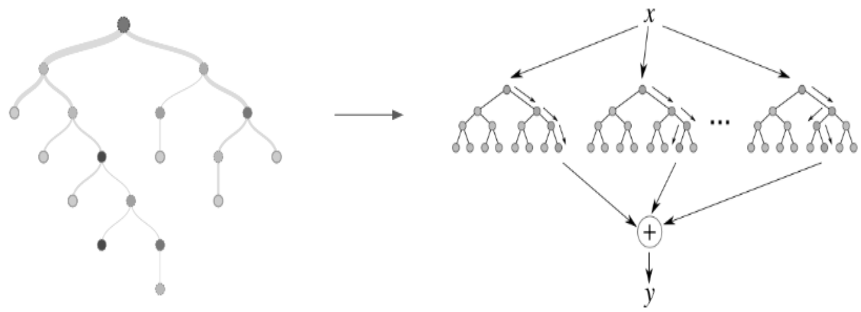


Figure 14a: Visualization of the difference between decision tree and a random forest.

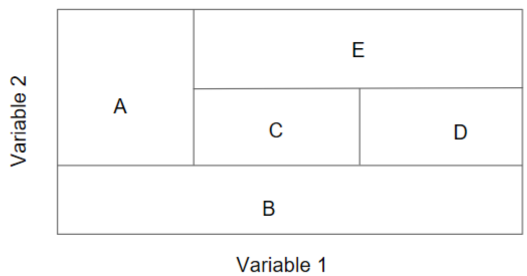


Figure 14b: Simplified visualization of the idea of predicting classes from variables

For this we used a common and easy to implement random forest. A random forest can describe non-linear relations in data using multiple predictors (decision trees). Combining these multiple trees can lead to more stable and precise predictions[6].

If we compare this to the regression algorithm in 2.2.3, we do not draw one line but draw multiple blocks instead and check in which block (class) a certain session will be. However, this is still linear in the way of having straight lines in our graph. In reality this will not be this straight forward, so we want to have a more flexible approach. Result of this method is a 87% accuracy in predicting 2 classes (longer or shorter than 6 hours of connection time)

2.2.5. Classification (non-linear)

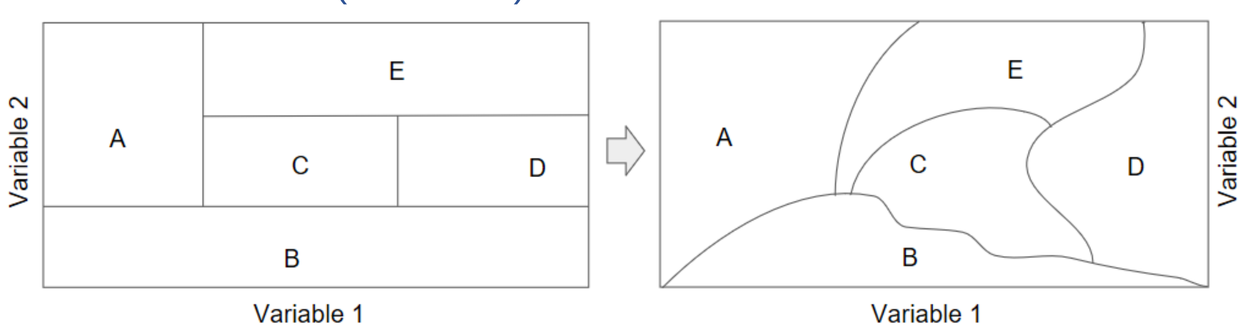


Figure 15: Simplified visualization of the difference between linear and non-linear prediction of classes

[6] https://en.wikipedia.org/wiki/Random_forest

For the flexible approach we used Keras neural networks. Our Keras model is implemented in R, has 3 Dense layers with 64, 64 and 8 units and Relu activation functions. Between these Dense layers we have dropout layers with a rate set to 0.2. Our last layer is an output layer with the number of units equal to the number of classes that we want to predict: 2-5 (we started with 2, ended with 5 classes). The function used for the last layer is softmax. The loss function is categorical crossentropy, next to that we used different optimizers but the best one was SGD. More info on Keras can be found on the Keras website[7]. The output provides in different classes and we have more flexibility in our model to draw complex lines in our graph. The result is 78-82% accuracy in a 5 class prediction. The set of features can be found in the appendix.

2.2.6. Results of prediction methods

	Averages	Weighted averages	GLMs	Classification (Linear)	Classification (Non-linear)
Why?	The average connection time of historical charging sessions should be an indication of the average connection time now.	Some charging sessions are more significant than other. Charging sessions should be weighed higher if they are significantly more important	The average connection time is a poor measure. Connection time is influenced by different factors which should be modeled.	GLMs attempt to predict the exact connection time. Another approach is to predict the time window (class) to which a charging sessions belongs.	Linear classifiers can not take complex non-linear boundaries into account.
How?	Calculate average transaction times based on historical sessions related to RFID, time and charge point	Within the maximum boundaries rate sessions based on different formulas	Fit a multivariable regression model to calculate the transaction times in the future	Use a flexible, easy and widely used classification model Random Forests to predict 2 time windows	Use neural networks to encapture non-linear correlations to predict 2+ time windows
Result	**Average error:** 2.1 - 4.8 hours	**Average error**: 1.8 - 3.5 hours	**R-squared:** 0.49 **Average error:** 2.6 hours	2 classes: >6 and <6 hours **Accuracy** 78%	4 classes: ≤6, <12, <16, ≥16 **Accuracy** 78% - 82%

Figure 16: explanation and results of used methods

As a wrap-up the results are shown in the image above. As stated, the first two methods produced some poor results and show relatively high errors. Also the regression model showed an high error because we wanted to predict the connection time from minute to minute. The classification results were much better, due to the decision to start with only two classes. Two classes is not something you want to end up with, but it can be a great start to get some initial results and a feeling if classification will work in the first place. After that we began with increasing the number of classes, with the same level of accuracy. This went reasonably well and we managed to get an accuracy with an average of 80% with both the random forest and Keras neural networks. Even though the results show a significant result, it’s still too premature to implement these results into a real life setting as the classes are too broad and we require an 90% or higher accuracy in predicting.

[7] https://keras.io

2.3 Clustering charging sessions

The charging sessions can be clustered based on two questions

1. What time does a charging session start?
2. How long does a charging session last?

These two questions correspond with the x and y axis respectively in the figures below.

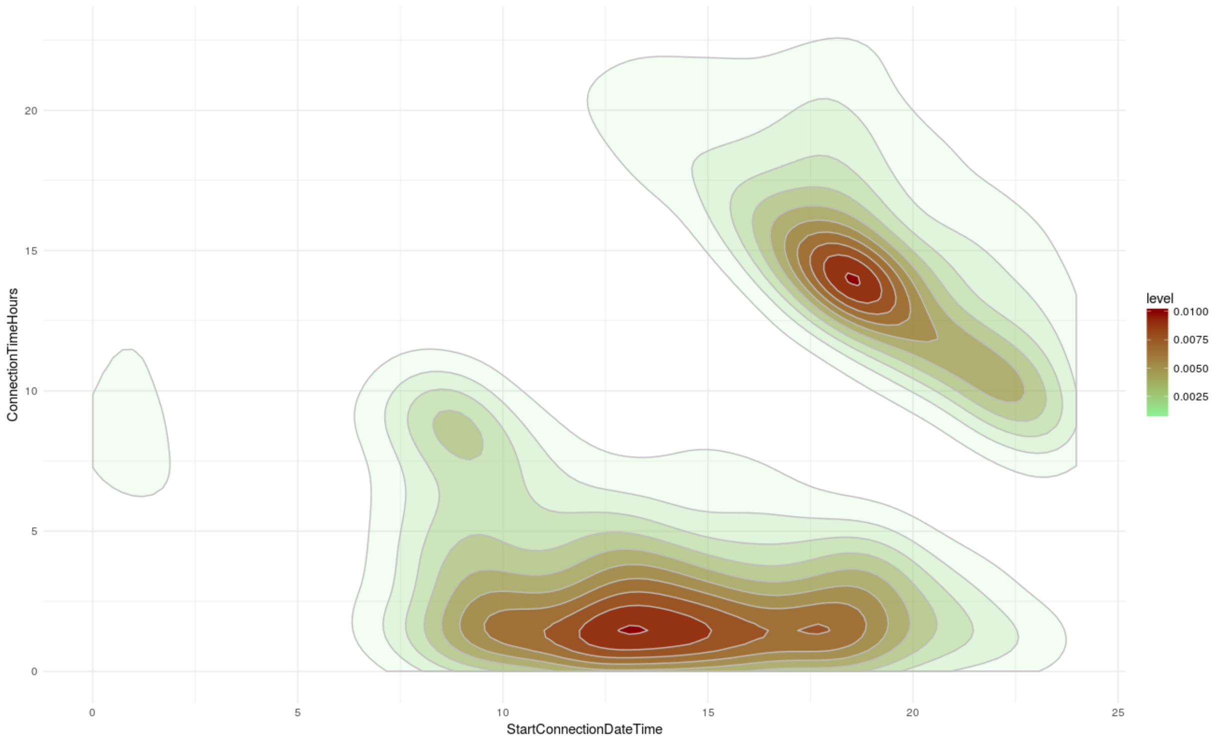


Figure 17: daily density plot for charging sessions in Amsterdam (2015-2017)

The levels in this density plot (or height map) are an indication of the number of charging sessions. Figure 17 shows only Amsterdam to simply illustrate the density differences. The distribution of charging sessions might be different over time. To get an indication of how the distribution of charging sessions depends on temporal variations we have a look at the same height map from different temporal perspectives. In figure 17 shown above, we see a typical density heightmap for one full day. We can already distinguish three main clusters.

The first one on the top right is the “evening sessions” cluster. This group mainly consists of people who come home approximately between 16.00h-20.00h, and are on average connected for 12-16 hours (until the next morning).

The second one in the middle is the “morning sessions” cluster. These are the typical working people who arrive at their offices between 7.30h-9.30h and who stay connected for an average 8-10 hours.

The third cluster is a more mixed group with more variable charging behavior, we call them the “sessions throughout the day” cluster. They charge during the day and are recognizable by their short (typically <4 hours) connection times. This group consists of a divers set of users that visit locations such as city centres, restaurants, offices and other Points of Interest for a short period of time.

An individual user can have charging sessions in multiple clusters.

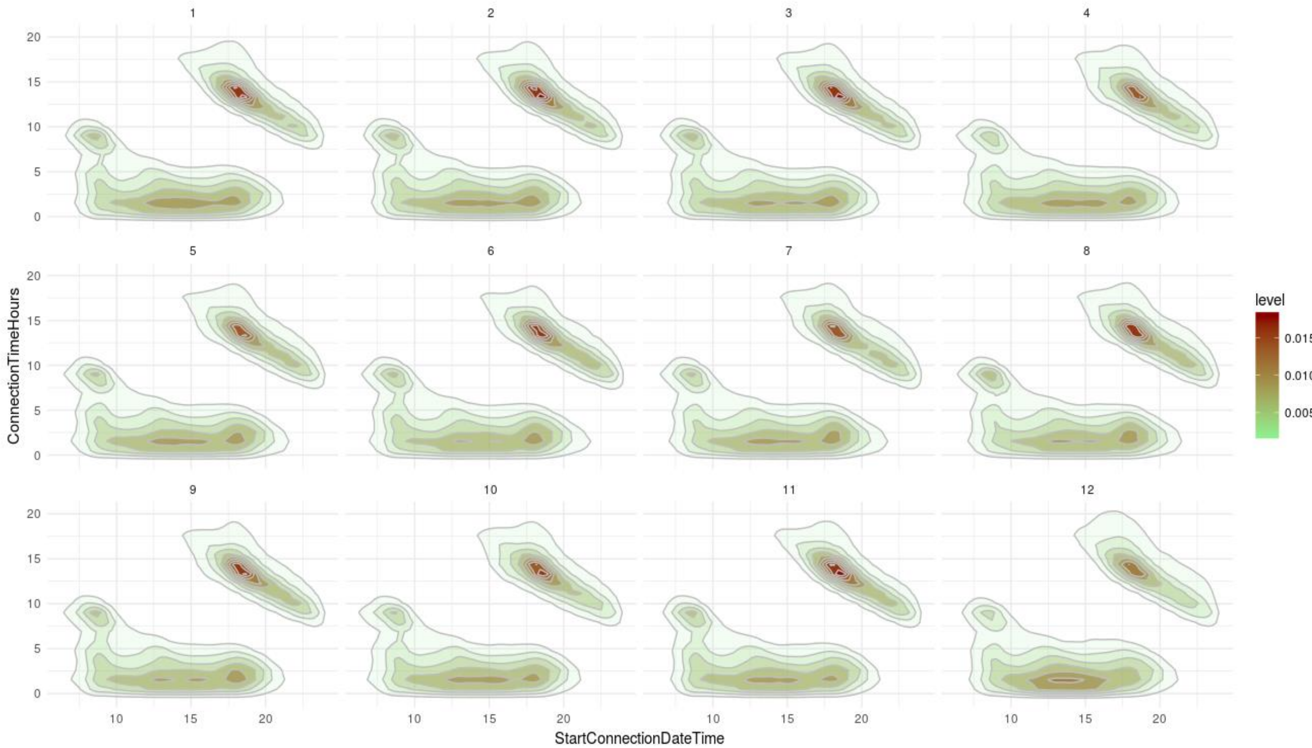


Figure 18: monthly density plot for Amsterdam (1=January, 12= December, 2015-2017)

Looking at the density plots for Amsterdam over the years for each month, we see little difference in the distribution of sessions. We recognize the same clusters shown in the previous plot for all the sessions. This indicates that there is barely any seasonality in distribution of start time and connection times.

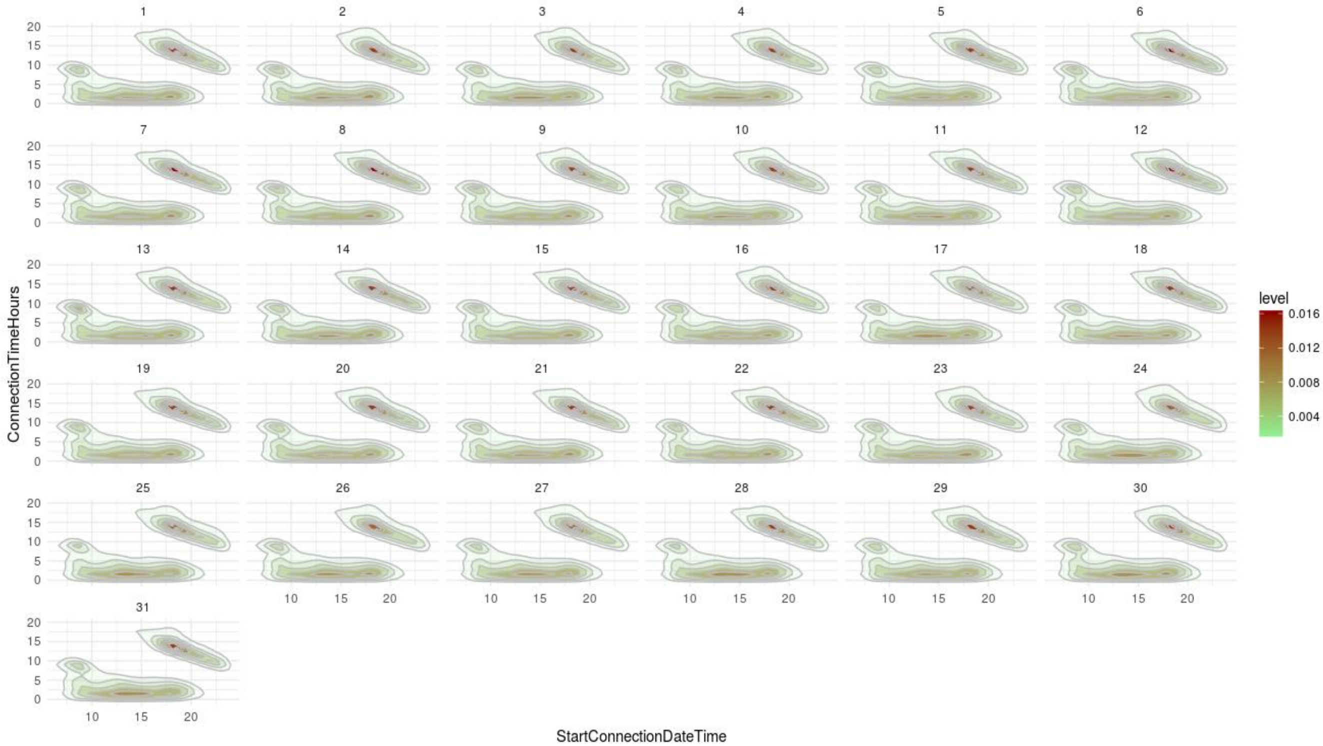


Figure 19: density plot per day of the month for Amsterdam (2015-2017)

If we make the density plots per day of the month over the three-year period, we can test the hypothesis that the day of the month influences how the clusters of sessions behave. We found that the effect is rather small on the different clusters.

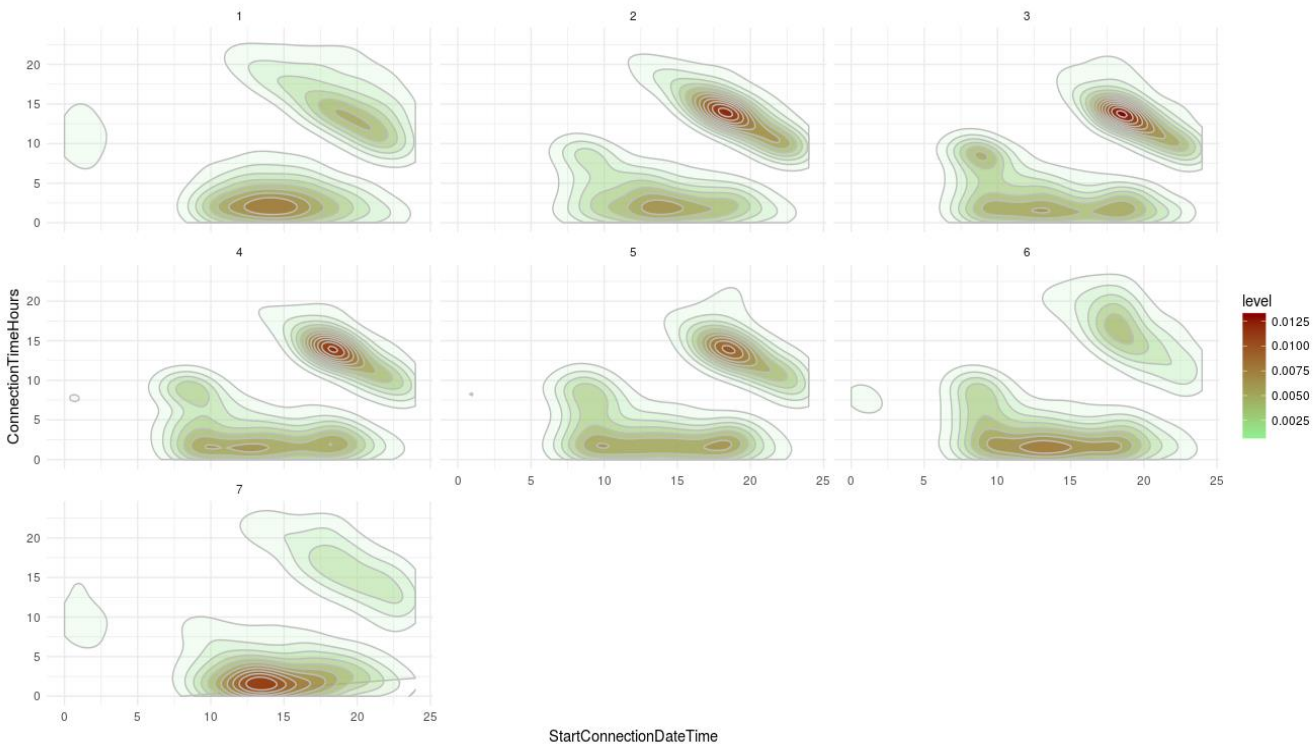


Figure 20: density plot per day of the week for Amsterdam (1= Sunday, 7= Saturday, 2015-2017)

When we look at the different days of the week though, significant differences between week and weekend days are shown. We observe that clusters can vary strongly. In the weekend, there are barely any morning chargers, but a lot of charging sessions throughout the day. Also the evening chargers category is significantly smaller. On Monday till Thursday, the three different clusters are clearly visible.

The question is whether we can use clustering techniques to distinguish these clusters. In order to use algorithms to determine the optimum Smart Charging strategy the charge sessions characteristics need to be understood. As can be seen in figures 17, 18, 19 and 20 the charging sessions are concentrated at different points. Therefore it can be assumed that charging sessions closer to each other have similar characteristics.

Figure 17-20 show the *distribution* of charging sessions with respect of connection start time and connection duration. In order to identify the different clusters we used density based clustering algorithms.

The first algorithm we used is DBSCAN (Density-Based Spatial Clustering of Applications with Noise), a widely used clustering algorithm to cluster data points based on density. The second algorithm we used is Gaussian Mixture Models (GMMs). Since GMMs are parametric models it allows us to capture more information in the data. The GMMs fit different Gaussian distributions at different points in the data.

In figure 21 and 22 the results of using DBSCAN to cluster the charging sessions are shown. In figure 21 the value for the DBSCAN parameters are 0.9 and 50 for Epsilon and MinPts respectively. In figure 21 three clusters are identified and the red points are considered to be noise.

In figure 22 a slightly different values of the parameters are chose: 0.95 and 95 for Epsilon and MinPts respectively. As can be seen in figure 22, less points are considered to be noise, but this goes at the expense of the number of clusters that can be identified.

The DBSCAN algorithm is not robust enough for clustering charging sessions. A slight change in parameters results in a significant change in the number of clusters.

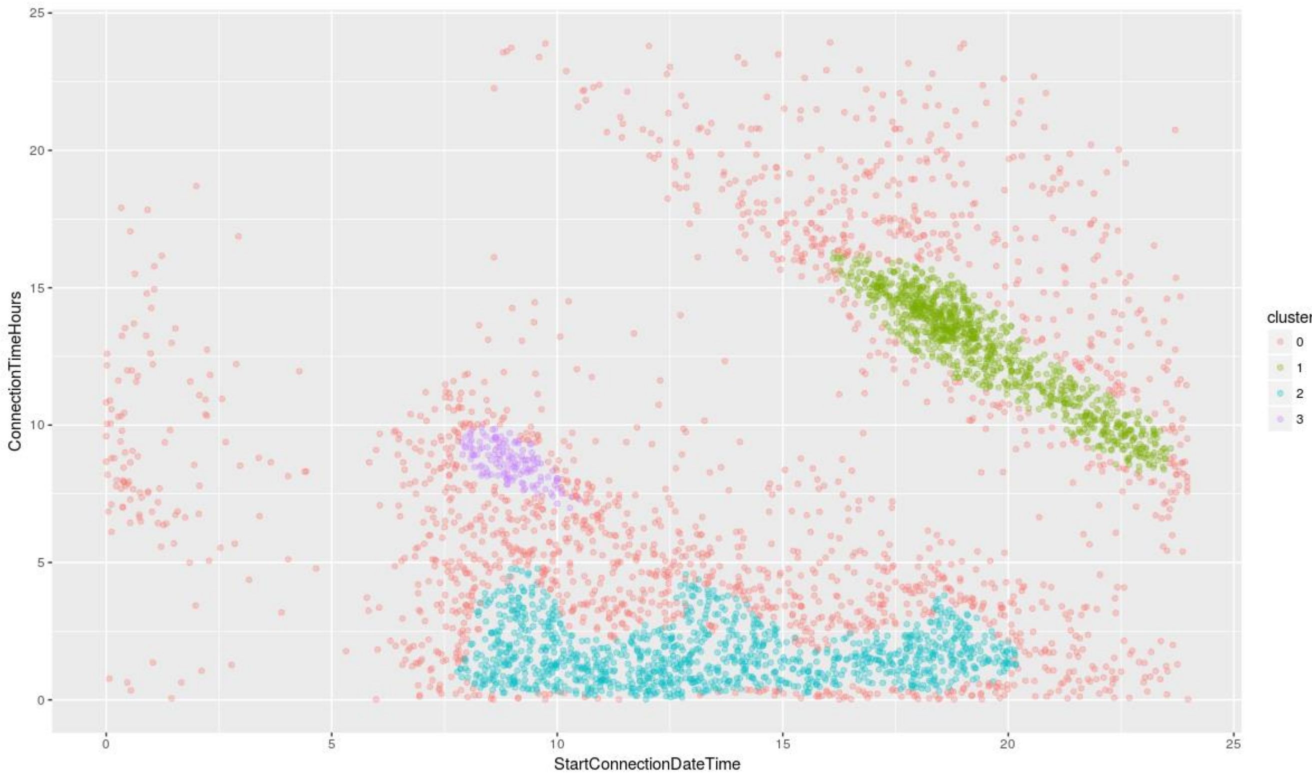


Figure 21: DBSCAN with 3 clusters visualization of charging sessions in Amsterdam. The clustering is done using DBSCAN with parameter values Epsilon = 0.9 and MinPts = 50.

As can be seen in figure 21, three clusters are identified using DBSCAN.
The red points are considered to be noise, which is in this case quite a lot. If we want to assign these points to a cluster, it means that the number of clusters is reduced. What we then get is the following:

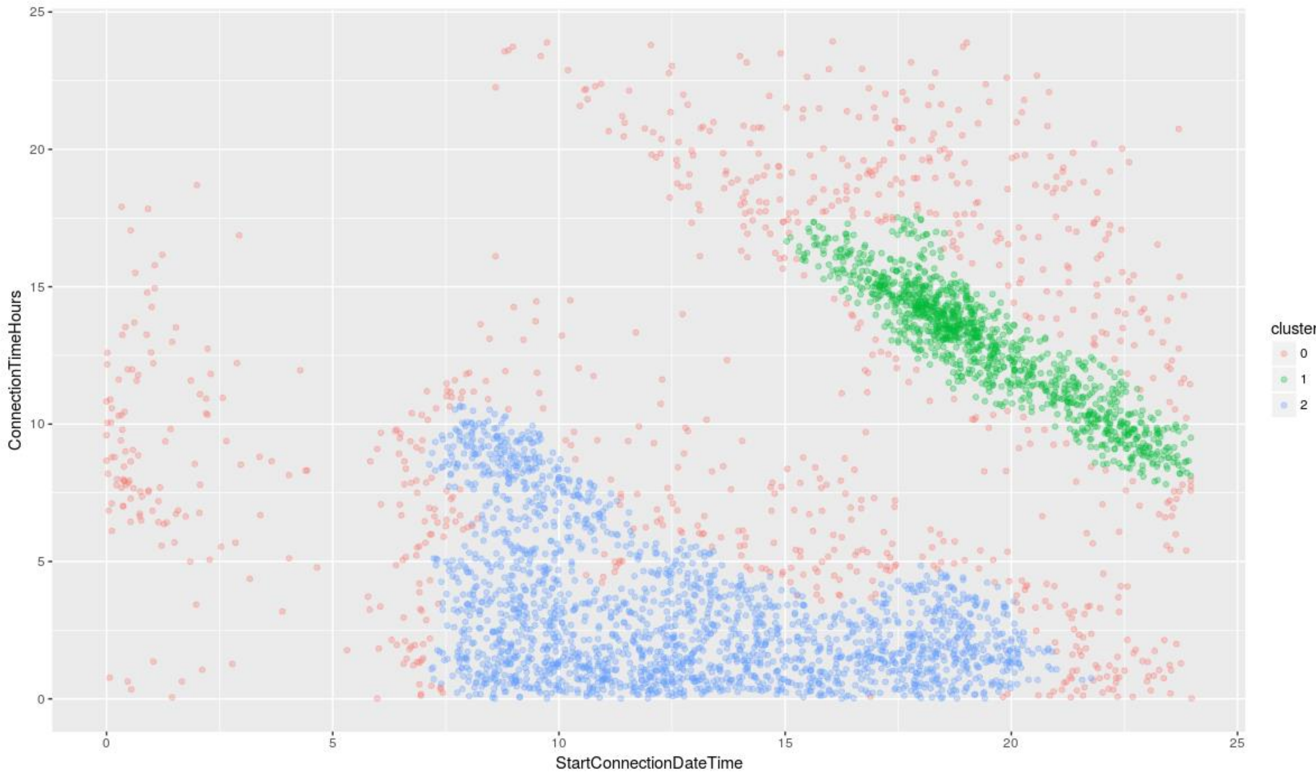


Figure 22: DBSCAN with 2 clusters visualization of charging sessions in Amsterdam. The clustering is done using DBSCAN with parameter values Epsilon = 0.95 and MinPts = 95.

Gaussian Mixture Models (GMMs) work better for recognizing the different clusters. GMMs use multiple probability density functions that are fitted to different parts of the data. This model parameterization allows us to have a more robust clustering algorithm.

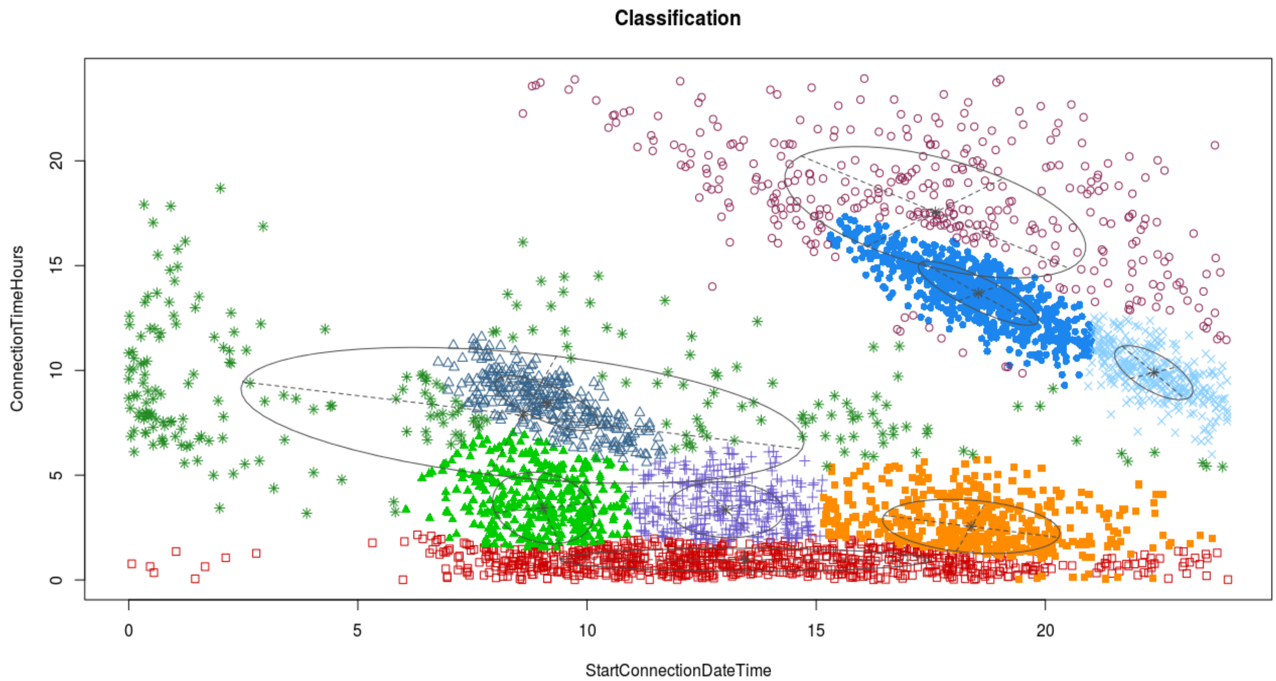


Figure 23: GMM visualization with 9 clusters of charging sessions in Amsterdam (2015-2017)

In figure 23 the result of using GMMs are shown. The nine clusters identified are based on the value of the Bayesian Information Criterion (BIC). As the number of clusters is increased, the BIC value increases too which means that more information is captured as new clusters are added. At a certain point the value of the BIC does not increase significantly anymore. The number of clusters at which this occurs is the optimum number of clusters that should be used.

In this research nine clusters are chosen although more research and analysis should be done to determine the optimum number of clusters using domain knowledge besides the BIC.

2.4 Optimizing charging sessions

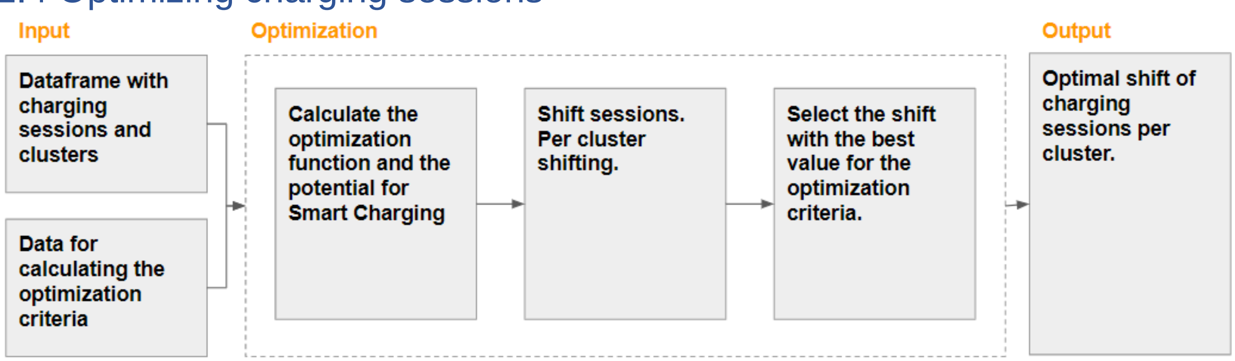


Figure 24: optimizing charging sessions pipeline

After clustering the data, we focused on optimizing Smart Charging strategies for the individual clusters. For the sessions in each cluster a single optimum was computed meaning that the sessions in each cluster are treated as a homogeneous set. This is a reasonable assumption since sessions with similar characteristics in terms of starting time and connection duration are similar in terms of charging needs and hence the optimum. Optimizing Smart Charging strategies requires two important inputs. First, it requires data on the charging sessions. Second, it requires data on the optimization criteria. Optimization criteria can be different, and sometimes conflicting as mentioned before.

We will look at optimizing the postpone strategy. Optimizing here means finding the right shift by which the sessions in each cluster should be shifted (postponed) in order to optimize for the chosen optimization criterion. The optimization itself consists of a number of steps. First the optimization cost function is defined based on the optimization criterion. The optimization cost function assigns a reward or a penalty to each kwh charged at a given time during the day. For example, when optimizing for solar energy, charging at noon has lower cost than charging at night since at noon there is more solar power available, as you can see below in figure 25. Secondly, the smart charging potential is computed for each session. The smart charging potential is an indication of how much room there is to shift a charging session.

After defining the cost function and Smart Charging potential, the sessions are shifted from 0% to 100% of the smart charging potential. Which means that at 0% the session is not shifted and at 100% the session is shifted completely given the permissible interval. It is important to keep in mind that shifting charging sessions happens within the start and end times of the connection known from the historical data. This is a way to ensure that the optimization is realistic and takes the needs of the end user into account.

After computing the costs for the different possible shifts, the optimal shift per cluster is then selected.

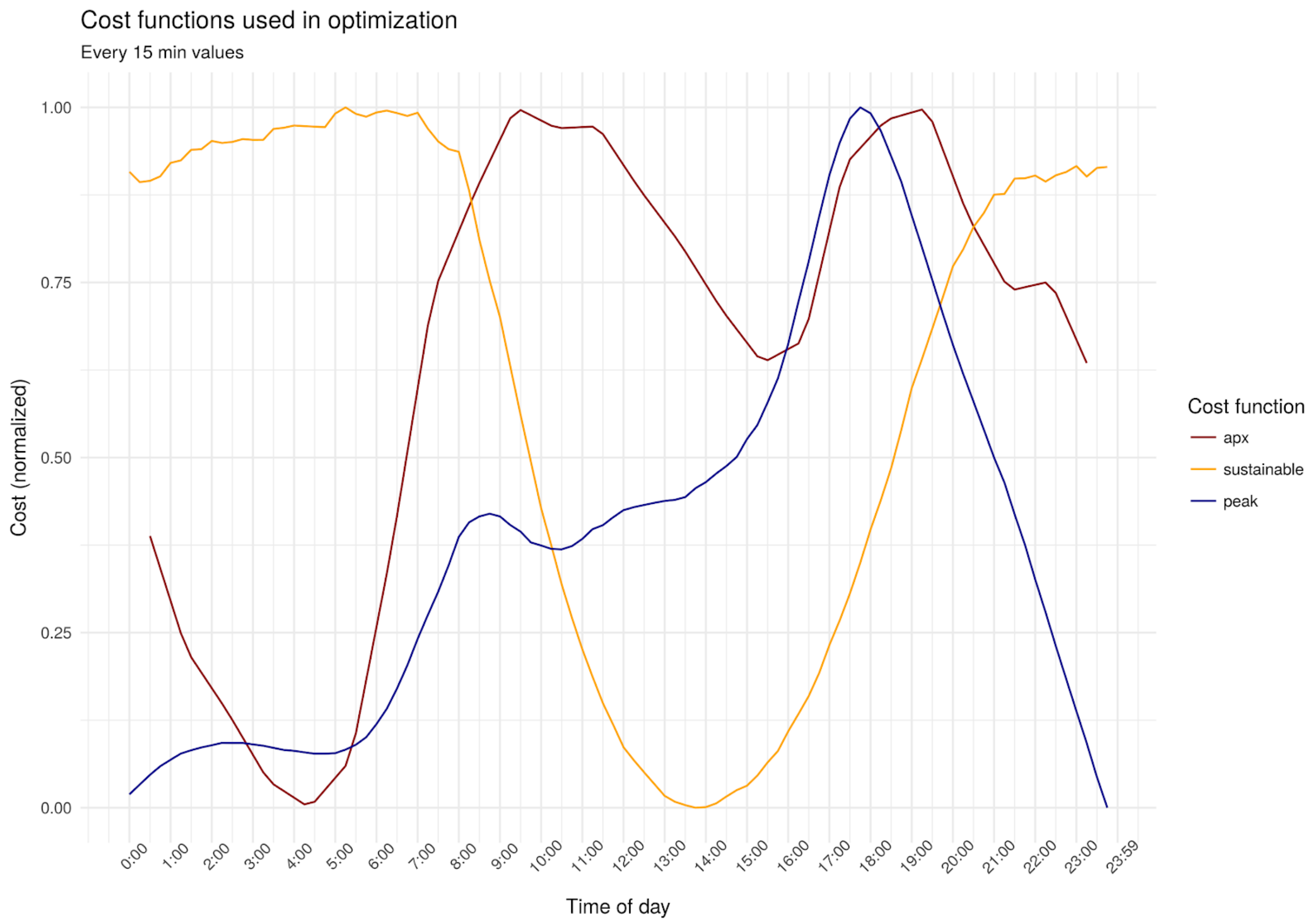


Figure 25: cost functions for APX price, sustainable energy and grid peaks

In figure 25 the three different cost functions are shown, all based on their own data input. Apart from these examples, we can vary in any cost functions used in our model. Basically we can adapt the cost functions for specific optimizations. If for example we want to optimize for local grid load in a certain city and we have the peak load data available, we could use that for that specific optimization.

For this research we used averaged cost functions to generate day profiles for each cost function. Averaging in this case is not accurate. But the aim here was to create optimization models that easily generalize to include cost functions of any kind.

Optimization goal	**Cost function:**	**Data used:**	**Limitations**
Net impact	Peak loads in grids	Total charging demand based on all charging sessions in the dataset used for this research	No household data
Renewable energy match	Solar and on-shore wind generation data	Entsoe, 2015-2018	No off-shore wind data used

Energy costs	APX electricity price index	APX prices (2015) – averaged over the entire year	Averaging the data over 1 year does not always take the fluctuations into account

The APX example is from the average APX-prices of 2015. The sustainable energy example is form average profiles over the years 2015-2018. The peak power demand is based on average loads over 2015-2017. As soon as recent data is available, the same optimization models can be used without having to use average cost function profiles.

As can be seen the different cost functions can be conflicting –in above figure illustrated by sustainable energy with different cost functions both during daytime and nighttime. As a result optimization for all objectives is not possible, making smart charging strategies per definition a question of prioritizing on which optimization goals are preferred over others.

2.5 Combining clustering and optimization

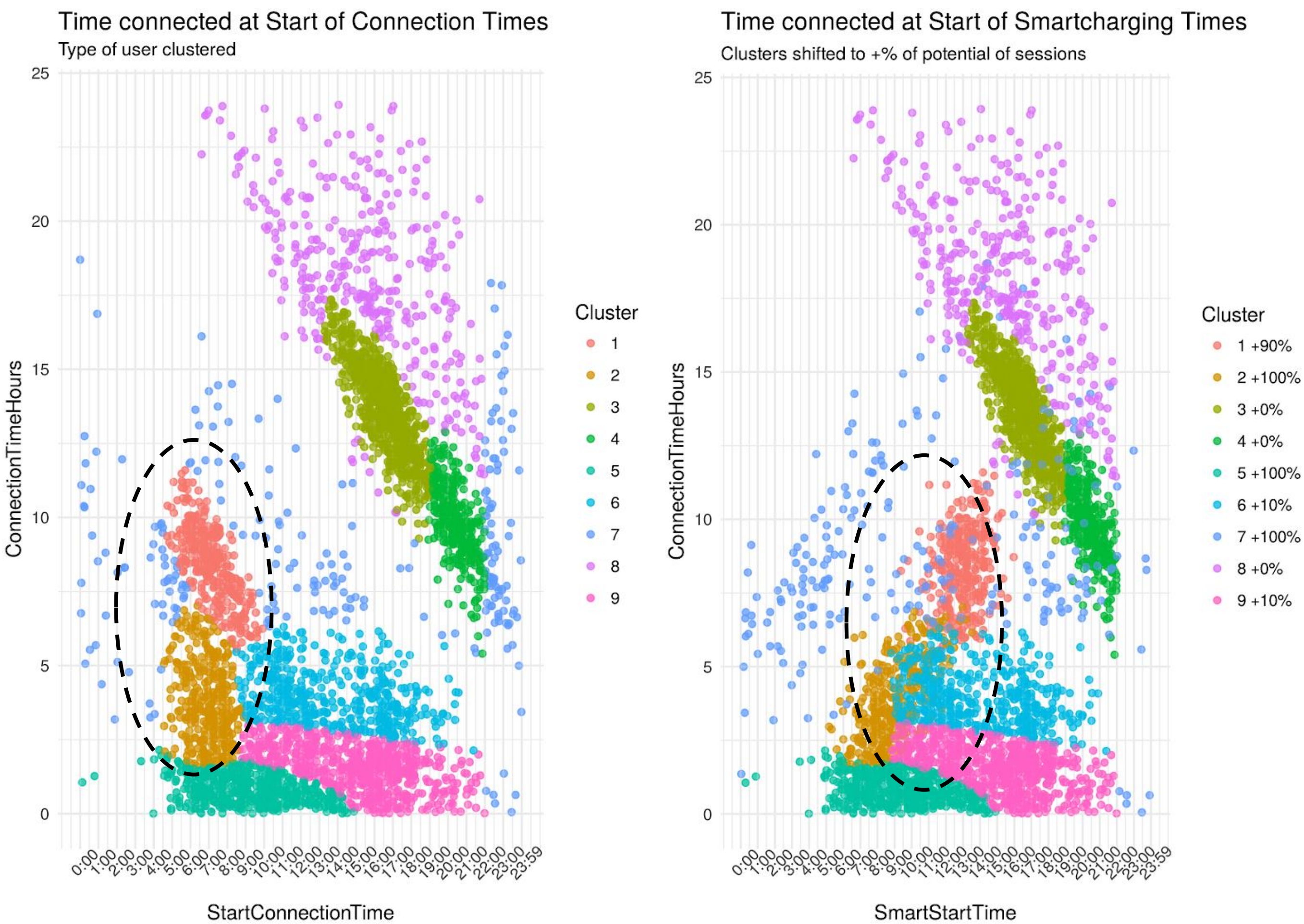


Figure 26: Smart Charging optimization of the 9 clusters for renewable energy supply. The percentage indicates the shift of the sessions measured in the portion of the Smart Charging potential

When we combine both the clustering and the optimization results, we can model how optimization plays out in the starting times of different clusters. In this figure we see the nine clusters defined using a GMM model. On the left we see the distribution of charging sessions when no smart charging is applied. On the right we see the optimal postpone strategy optimized for renewable energy generation. As we can see on the right figure, not surprisingly mainly the morning sessions are shifted to match solar irradiation profiles being highest during daytime. Apparently, when optimizing for solar energy generation, morning sessions are most affected.

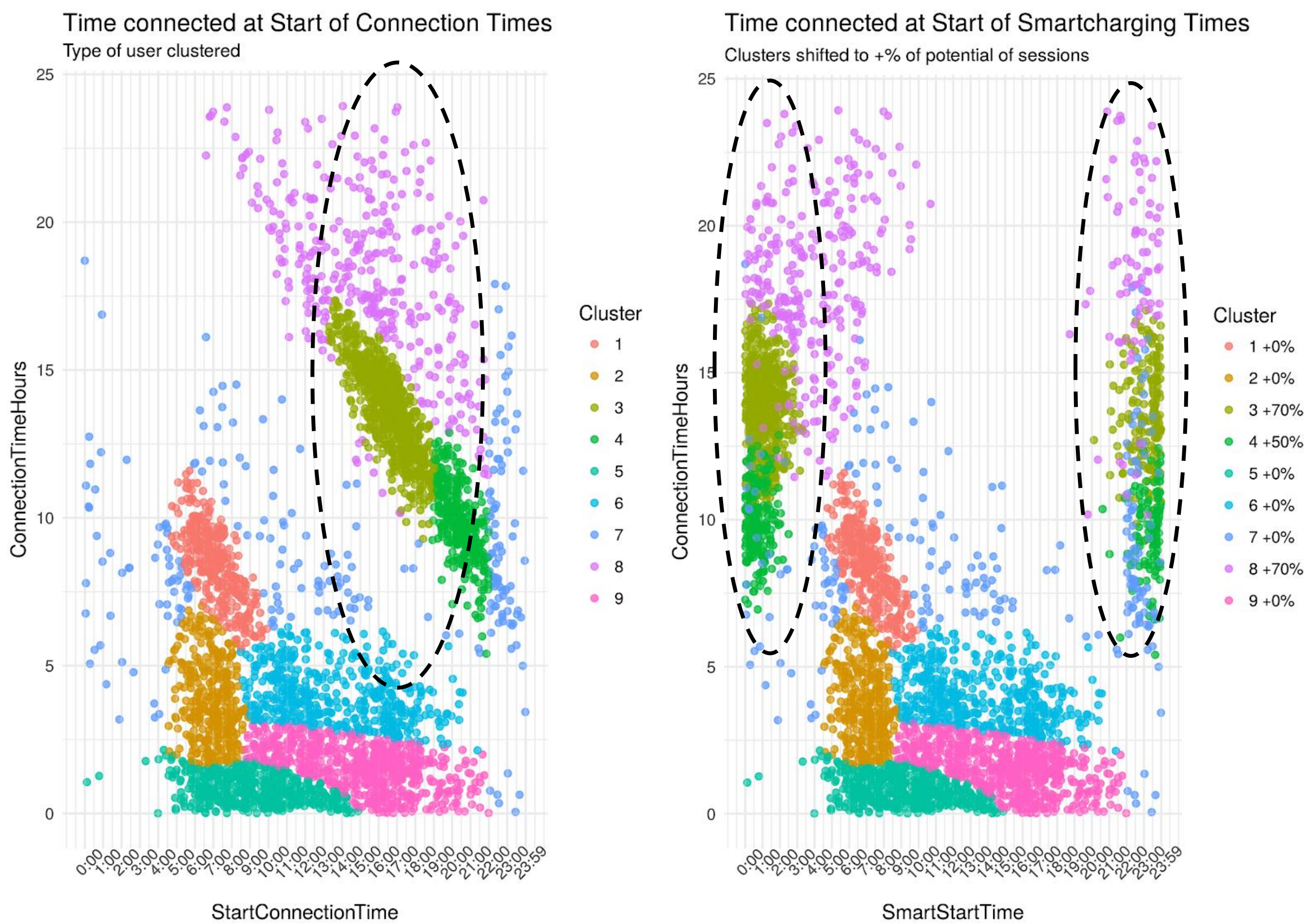


Figure 27: Smart Charging optimization for grid capacity

The same optimization is applied for the energy demand (net congestion reduction). And as can be observed in above figure, mainly the evening sessions are shifted. That is because the home chargers are the only clusters providing enough flexibility at the hours when the most demand occurs. (roughly between 17.00h and 20.00h)

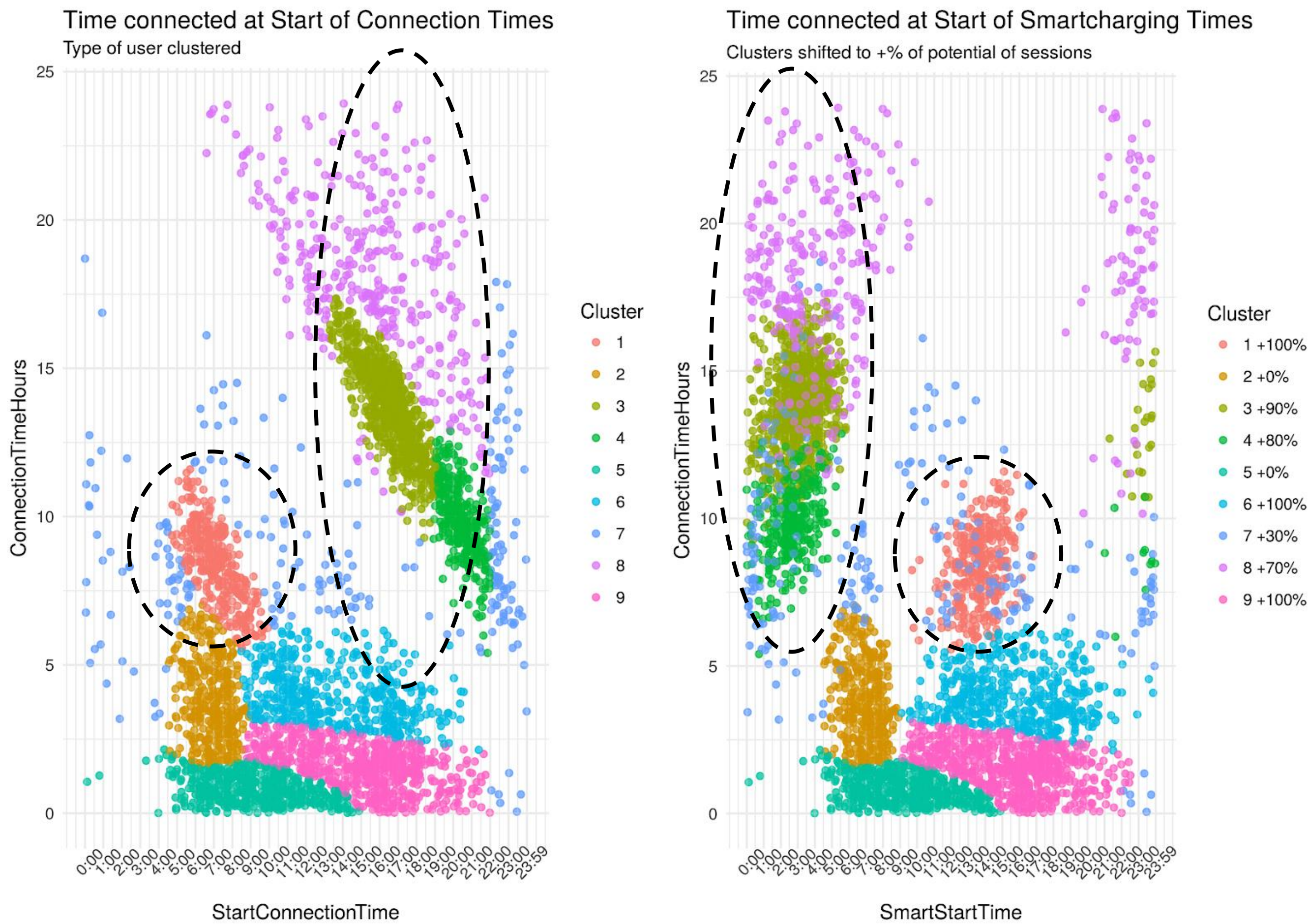


Figure 28: Smart Charging optimization for APX prices

When optimizing for the APX index, a mix of morning and evening sessions are shifted. Intuitively that does make sense. The APX index is an energy price index determined by supply and demand of energy. The supply is influenced by the availability of sustainable energy sources, and the demand is influenced by the relatively high demand in the early morning and evening hours by households and EVs. This means that those low supply-high demand hours will be the most expensive ones. Given that fact, we want to charge the EVs preferable around noon when demand is lower and (renewable) supply on average is significantly higher, or at night when demand is even lower.

So far we have seen the optimization when considering a single optimization strategy (postpone). The question for the next part of our research is whether we can develop generalized optimization methods that can deal with different Smart Charging strategies, handle different optimization cost functions and determine the optimum for each cluster.

3. Conclusions and recommendations

3.1 Conclusions

This research focused on predicting charging and connection time and optimizing charging profiles using the postpone Smart Charging strategy. The main conclusions of this research are:

- For prediction five different methods are used. In the first two methods averages are used to predict connection times, which rates every session the same way and averages out any pattern in the data. The GLM model is better in capturing these patterns but it is not needed (and too difficult) to predict the connection time to the minute, this results in high errors. Therefore, we can conclude that the classification methods are the best approach and also resulted in the best accuracy.
- Predicting charging and connection times using these different methods resulted in a varying accuracy between 78% and 82%. Although predicting charging and connection times is challenging, this is a promising first result. In the future this result can be improved using other modelling techniques and creating new features from the original dataset.
- Clustering charging sessions is done using two different methods (DBSCAN and Gaussian Mixture Models). Clustering using GMMs resulted in a more detailed clustering with more groups. The challenge is to interpret the found clusters and link these to situations in reality.
- Optimization charging profiles for the different clusters is based on the *Postpone* Smart Charging strategy. Depending on the optimization objectives, the different clusters are shifted in different ways. When optimizing for sustainable energy, mainly the morning sessions are affected, when optimizing for power demand peaks the evening sessions are shifted and when optimizing for the energy price a combination of morning and evening sessions are shifted.
- Ideally we would like to predict the class a new charging sessions belongs to and based on this prediction apply the optimal Smart Charging Strategy for that cluster. In this research the prediction of the class to which a charging sessions belgons was done based on the 9 clusters identified using GMMs. The models which were used to accomplish this can still be finetuned to improve the accuracy.

3.2 Recommendations

The results of this research project form a solid basis for further development of Smart Charging algorithms. To improve the results obtained here, we would like to make the following recommendations:

- The developed models within this project can be further developed in the Simulaad project since the project objectives regarding Smart Charging development overlap and the results in this project form a solid basis.
- Better quantify the effect of applying Smart Charging strategies using the number of charging sessions affected and the charges in the charging profile.
- Develop generalized optimization methods that can deal with different Smart Charging strategies, handle different optimization cost functions and determine the optimum for each cluster. This means that in addition to the *postpone* strategy used in this report, the optimization function can handle the cut-and-divide and slower charging strategy.

- In this research nine clusters of charging sessions are chosen although more research and analysis should be done to determine the optimum number of clusters using domain knowledge besides the BIC.
- Choose smaller time windows for the classes of charging and connection times in order to predict the charging and connection times in a fine grained manner. Choosing smaller time windows should not affect the accuracy. Preferably aim for an accuracy of at least 90%.
- The models are created using historical data and is still a theoretical framework. In order to test the performance of the model and gather new data, the models should be tested in practice within Smart Charging pilots. In addition to this, we want to provide relevant market actors with our conclusions and make our algorithms usable for them.
- In Smart Charging pilots, user acceptance of Smart Charging is key. Users need to be able to overrule a proposed Smart Charging session when they want. Ideally, Smart Charging is provided as a service to users in which they do not notice different outcomes of charging sessions, but do take profit of e.g. lower energy prices.
- Clustering and prediction should be better tuned to each other. In the following research start with clustering first and use the clusters to classify the sessions and optimize for each clusters.
- Use a Bayesian approach to predict distributions of charging and connection times. This approach allows to have an indication of the uncertainty in the predictions. This uncertainty can be used as a measure to decide whether to apply Smart Charging or not.
- Use dimensionality reduction to reduce the number of levels in some of the features. Some features/variables have a lot of levels, for example: there are more than 6800 charge points (i.e 6800 levels). Most Machine Learning models are limited in the number of levels they can handle. Reducing the number of levels, allows the research team to include more features and use more models.
- Optimizing for the grid load is approximated using the charging demand. However, this is not in line with the reality. In reality the household consumption has a significant effect on the grid load. For this research there we haven’t fitted realtime grid load or household consumption data into our models yet. In order to improve the models further, realtime grid load and household data should be used.

4. Further reading on this topic

For the interested reader in this topic, we would like to refer to the following publications.

- A. Hoekstra & N. Refa, Characteristics of Dutch EV drivers, EVS30, 2017.
- A. Beltramo, A. Julea, N. Refa et al., Using electric vehicles as flexible resource in power systems: A case study in the Netherlands, 2017 International Conference on the European Energy Market, Dresden.
- C. Bikcora, N. Refa, L. Verheijen and S. Weiland, Prediction of availability and charging rate at charging stations for electric vehicles,2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Beijing, 2016, pp. 1-6.
- C. Develder, N. Sadeghianpourhamami, M. Strobbe and N. Refa, Quantifying flexibility in EV charging as DR potential: Analysis of two real-world data sets, 2016 IEEE International Conference on Smart Grid Communications (SmartGridComm), Sydney, NSW, 2016, pp. 600-605.
- J.R. Helmus, J.C. Spoelstra, N. Refa et al., Assessment of public charging infrastructure push and pull rollout strategies: The case of the Netherlands, Energy Policy, volume 121, October 2018, pp. 35-47.
- N. Sadeghianpourhamami, N. Refa, M. Strobbe and C. Develder, Quantitive analysis of electric vehicle flexibility: a data-driven approach, International Journal of Electrical Power and Energy Systems, volume 95, February 2018, pp. 451–462.
- S. Hardman et al., A review of consumer preferences of and interactions with electric vehicle charging infrastructure, Transportation Research Part D: Transport and Environment, volume 62, July 2018, pp. 508-523.

Appendix

Prediction features

List of features used to predict connection times.

DayOfWeekStart	Day number of the week of the session
MonthStart	Month number of the start of the session
WeekStart	Week number of the start of the session
HourChargingStart	Hour number of the start of the session
dayPart	We defined multiple parts of the day (hours: 0-7, 7-12, 12-16,16-24), daypart is a categorical feature that represents these classes
rfid_wday_mean_connected	Historical mean connection time of the sessions with the same weekday of the new session
rfid_wday_sd_connected	Historical standard deviation of the connection time of the sessions with the same weekday of the new session
rfid_wday_median_connected	Historical median of the connection time of the sessions with the same weekday of the new session
rfid_wday_n	Historical number of sessions with the same weekday of the new session
rfid_dp_mean_connected	Historical mean connection time of the sessions with the same daypart of the new session
rfid_dp_sd_connected	Historical standard deviation of the connection time of the sessions with the same daypart of the new session
rfid_dp_median_connected	Historical median of the connection time of the sessions with the same daypart of the new session
rfid_dp_n	Historical number of sessions with the same daypart of the new session
cp_wday_mean_connected	Historical mean connection time of the sessions from a certain charge point with the same weekday of the new session
cp_wday_sd_connected	Historical standard deviation of the connection time of the sessions from a certain charge point with the same weekday of the new session
cp_wday_median_connected	Historical median of the connection time of the sessions from a certain charge point with the same weekday of the new session
cp_wday_n	Historical number of sessions on the charge point with the same weekday

cp_dp_mean_connected	Historical mean connection time of the sessions from a certain charge point with the same daypart of the new session
cp_dp_sd_connected	Historical standard deviation of the connection time of the sessions from a certain charge point with the same daypart of the new session
cp_dp_n	Historical number of sessions on the charge point with the same daypart
cp_dp_median_connected	Historical median of the connection time of the sessions from a certain charge point with the same daypart of the new session
MaxPower	The max power from all historical sessions from the user