2018

DRIVING BEHAVIOR ENERGY CONSUMPTION ESTIMATION FOR ELECTRIC VEHICLES

Niklas Schwertner
University of Rhode Island, n.schwertner@tu-bs.de

DRIVING BEHAVIOR ENERGY CONSUMPTION ESTIMATION FOR ELECTRIC VEHICLES

BY

NIKLAS SCHWERTNER

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

SYSTEMS ENGINEERING

UNIVERSITY OF RHODE ISLAND

2018

MASTER OF SCIENCE

OF

NIKLAS SCHWERTNER

APPROVED:

Thesis Committee:

Major Professor Gretchen Macht

Jyh-hone Wang

Christopher Hunter

George Tsiatas

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2018

ABSTRACT

Electric Vehicles (EV) sales are experiencing an increasing trend in many industrialized countries [1, 2]. Globally, at the end of 2017, there was an annual increase of one million EVs on the road, totaling to three million EVs on the road [3]. However, despite recent developments and the high potential of Battery Electric Vehicles (BEVs), the market penetration rate of EVs is still very low due to discrepancies between consumer expectations and knowledge, the limited range and long charging times [4, 5]. Recent research demonstrated that there is a significant difference in energy consumption of BEVs between aggressive and non-aggressive driving. This research additionally, provide evidence that the concept of eco-driving for Internal Combustion Engines (ICE) vehicles works well for describing energy efficient Driving Behavior (DB) for BEVs [6].

The goal of this research was to confirm the energy consumption clusters found in the literature, as well as to confirm and expand the clustering methodology executed for determining these clusters. The original literature executed a hierarchical clustering technique utilizing Ward's algorithm. In addition to verifying the hierarchical clusters, Latent Profile Analysis (LPA), a form of model-based clustering, is then introduced as the new clustering approach to explore alternative clusters through a more diverse clustering approach.

Based on the fact that Dataset 1 (from previous research) and Dataset 2 (from this body of work) were found to be statistically similar, they get merged into a more comprehensive dataset. This research confirmed the two energy consumption clusters

(i.e., efficient and inefficient drivers) found in previous research with Dataset 1 using Ward's method. Given the fact that the clusters were very similar for both Ward's method and LPA for Dataset 1, these results strongly affirm these previous results regardless of the methodological clustering approach. Clustering Dataset 2 with Ward's method resulted in three energy consumption clusters as well, providing proof that at least three clusters are significant. LPA for Dataset 2 revealed similar clusters providing evidence that Ward's method and LPA find similar cluster when the sample size within the clusters is sufficient large.

For the Combined Dataset, excluding the outlier driver 34.1, with a sample size exceeding 50 participants, Ward's method results in three significant clusters. This strengthens the argument that DB with respect to energy consumption can be clustered into at least three clusters. Expanding the cluster analysis by LPA provides a four and five component model with each equally shaped clusters, grouping drivers in accordance to what is known in the literature about the influence of DB on energy consumption.

This research provides a better understanding of how BEV drivers need to be clustered based on their mean energy consumption per mile and standard deviation. It provides strong evidence that the assumption from previous research, that at least 3 clusters are relevant when analyzing driving behavior with respect to energy consumption, is true. Additionally, further clusters are found on a more comprehensive dataset which go along with the perception of literature that acceleration and speed are main factors for explaining energy consumption of BEV driving behavior.

ACKNOWLEDGMENTS

First, I would like to thank my major Dr. professor Gretchen A. Macht for her invaluable guidance throughout the developing process of this research. Without her unrelenting commitment to make this research both feasible and valuable, the work completed would not have been possible. Her inspiring personality and her competent council have been an irreplaceable contribution to this research. I look forward to hearing about her continued success as a professor at the University of Rhode Island and wish her all the best.

I would also like to thank my inside committee member Dr. Jyh-hone Wang for aiding in statistical question around this research. I would also like to thank him for motivating me throughout the thesis developing process and his valuable counseling for the experimental part in this research.

Furthermore, I would like to thank Dr. Christopher Hunter, who served as my outside committee member. He has made great additions to this thesis that make the research even more relevant in the realm of transportation.

Lastly, I would like to thank my colleagues in the SIS Lab who supported me by reading my drafts, helping me with programming, supported me in visualization of the results and were valuable sparing partners for discussing my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 - INTRODUCTION

1.1 Background

Electric Vehicles (EV) sales show an increasing trend in many industrialized countries [1, 2]. China had a tremendous increase in EV sales and their annual sales volume increased by almost 7 times from 2014 to 2016 [7]. European countries show a similar trend in the number of EV sales which have increased by almost 200 percent form around 50,000 EVs in 2013 to around 150,000 EVs in 2015 [8]. Belgium alone has almost increased their EV sales 7-fold from 919 in 2013 to 6552 in 2017 and Sweden doubled their EV sales from 2014 to 2017 to almost 70,000 [9, 10]. The United States (US), have significantly increased their EV sales per year from around 50,000 in 2012 by 200 percent to 150,000 in 2016 and by more than 30 percent up to 200,000 in 2017 [11, 12]. For the US in 2018, this trend is predicted to be continued with an even stronger increase up to around 400,000 EV sales by the end of the year [11]. Overall, at the end of 2017, there was an annual increase of one million EVs on the road, totally to three million EVs on the road globally [3].

These rapid development in EV demand is provoked by socio-economic changes of increased urbanization, financial incentives, and political engagement for stricter environmental regulations [13], which spurred a redesign of transportation systems towards high-quality services for the customer with an minimal environmental footprint [14]. In that context electrification and on-demand services are two main driving forces within the current global automotive sector to meet this challenges [15]. Plug-in hybrid electric vehicles (PHEVs) and battery electric vehicles (BEVs) can contribute to significantly decreasing road traffic emissions and BEVs specifically can help to maintain zero local emissions [14, 16]. However, despite recent developments and the high potential of BEVs, the market penetration rate of EVs is still very low which is due to the discrepancies between consumer expectations and the limited range and long charging times [4, 5].

To address those issues EV manufacturers and service providers are working on increasing range limits and decreasing charging times. There is continuous research that focuses on improving the battery capacity [17], designing gearing configurations for better efficiency [18], or applying regenerative braking systems (RBS) [19]. Apart from these technical improvements on the EV itself, a lot of effort is put into optimization of charging infrastructure [20] and in energy efficient route planning [21]. Optimizing these factors and extending the overall range of EVs makes an analysis of the energy consumption of EVs essential [5]. In this context, recent development in information and communication technologies provide a basis for collecting data driving data in real-time from multiple vehicles at relatively low cost. This creates the potential for accurate energy prediction on-demand and thus decreases miles travel, energy consumed and the environmental impact. [22].

Understanding a users' future energy demand in combination with the energy left in the battery allows an estimation of the remaining vehicle's range, referred to as residual range [23]. There are various energy prediction models for internal combustion engine (ICE) vehicles and hybrids [21, 23, 24]. For ICE energy estimation models eco-routing has become a popular navigation method to determine the route between the start and a destination that consumes the least fuel and produces the least emissions [4]. Some of these approaches, however, could also be used for BEVs energy estimation since they are regardless of engine type but consider resistances of various types on vehicle which are omnipresent [4, 24]. A significant difference is that commercial EVs are generally equipped with RBS, which allows them to recuperate some of the kinetic energy when braking. A prediction model for EV energy consumption would need to take this into account since energy recuperation influences the residual range significantly.

Recent literature on energy estimation models for EVs is divided into detailed scientific approaches focusing on transparency and accuracy in the energy estimation models [4, 25] and in less complex approaches focusing on applications for route optimization based on energy consumption [26]. Other literature focuses on specific aspects of on energy consumption, like relation between speed and energy consumption, [27] or the sensitivity of a physics-based energy estimation model [28].

1.2 Research Goals

Besides the external factors that influence drivability, which are determined by the system that surrounds the EV (e.g., road type, street signs, traffic lights), DB as an internal factor has a significant impact on the overall energy consumption of BEVs [29]. Understanding the influence of DB on energy consumption of ICE vehicles eco-driving is a well-defined method to describe energy efficient driving. Eco-driving improves ICE vehicle efficiency through controlled rates of speed and acceleration which involves such things: (a) as moderately acceleration, (b) anticipating traffic flow and signals to avoid sudden starts and stops, (c) maintaining even driven pace, (d) driving at the speed limit, (e) and avoiding unnecessary idling. [30] Recent research demonstrated that there is a significant difference in energy consumption of BEVs between aggressive and non-aggressive driving; providing evidence that the concept of eco-driving works well for describing energy efficient DB for BEVs [6].

Previous literature [6] found, collecting real driving data from 30 participants driving an BEV, two clusters of drivers with respect to their energy consumption: energy *efficient* drivers and energy *inefficient* drivers. A third cluster was initially found for intermediate energy consuming drivers, however, this cluster was found to be not significantly different from the inefficient driver cluster. This conclusion was unclear as to whether this was due to the fact that the sample size of the dataset used was small or whether there are only two clusters for energy efficiency of BEVs. Thus, previous research recommends to expand the number of individual driving samples to investigate whether a third intermediate energy consumption cluster would be statistically significant at a larger sample size [6].

This research conducts more test drives under these same previous, rigorous conditions. This allows for the energy consumption for the new, Combined Dataset to be analyzed using the same methodology for clustering drivers based on their energy consumption as the previous research in order to confirmed or augment these various driving clustering types [6].

Furthermore, this research assumes that both datasets, with exception of the instructor, are conducted under the same conditions. Additionally, tests will verify the consistency between these two data sets; if found statistically similar, the comparison and combination of these datasets will occur for stronger results. Should this test confirm the assumption that, both datasets could be merged to one dataset, it would increase the samples size by almost 80%. Conducting a hierarchical clustering analysis, along with a detailed perspective of a LPA, will be able to provide further insights on driving profiles with respect to energy consumption.

This research tries to give a more detailed picture on the following: (1) clustering of drivers based on their energy consumption for a new dataset, (2) verification that both datasets are indeed similar, and (3) if proven to be similar conduct a hierarchical clustering and a LPA based on the Combined Dataset to increase the understanding of previous results.

To investigate energy consumption of a BEV with respect to DB, the following research questions being addressed in this thesis are:

1. Does the Combined Dataset show two significant groups of energy consumptive behavior?
2. Does the Combined Dataset support the claim that more profiles of energy consumptive behavior exist?

In order to answer these questions, this thesis is broken down into the following chapters.

Chapter 2 analyses the literature encompassing the technical background on EVs, energy consumption models for EVs, and suitable ICE energy consumption models that are used as a reference in this research.

Chapter 3 presents the methodological approach and the statistical structure used to explore the answers to the research questions. The chosen test route is presented as well as the technical setup for the data collection. Furthermore, ANOVA is presented as a tool to compare Dataset 1 and Dataset 2 and Ward's method as well as LPA is introduced for finding energy efficient driver clusters.

Chapter 4 discusses and presents the results found in this research with respect to the analysis tools presented in the previous chapter. The comparison of Dataset 1 and Dataset 2 reveals that both datasets are statistically similar, and Ward's method finds three statistically significant energy consumption clusters for both datasets. LPA demonstrates a different perspective than Ward's clustering indicating different profiles of energy consumptive behavior.

Finally, Chapter 5 summarizes the results in this research drawing conclusions, presenting the work's limitation and gives an outlook on further research.

CHAPTER 2 - REVIEW OF LITERATURE

As mentioned in the previous chapter the main barriers for mainstream acceptance of EVs are the long charging times and the limited range. Fast-DC charger have the ability to reduce the charging time significantly so EVs can be quickly recharged which make them also usable for longer trips e.g. on highways between cities [31]. Even though the number of fast chargers in the US is growing rapidly, the number of publicly available fast EV chargers almost tripled from 2,518 in 2014 to 6,267 in 2017, the density of DC fast charger is still low and beside of Tesla's fast charging system mainly limited to urban areas [32, 33]. In addition to the limited fast charging capacity, driving under highway conditions reduces the range of EVs significantly. [31, 34] Aerodynamic resistance might be one the major factors for high energy consumption under highway conditions [24]. While at speeds less than 50mph the engine power is mainly needed to accelerate the vehicle at speeds higher than 50mph the engine power is primarily used to overcome aerodynamic resistance [24]. This draws the assumption that driving on road types with a higher speed limit will result in higher energy consumption for EVs [24, 34].

Understanding energy consumption for EVs is a complex problem with various influencing factors that miscellaneously correlated and vary over time. [4] EV range can be increased in various ways, most of which focus on improvement of battery capacity [17], the design of gearing configurations [18], or the application of vehicle RBS [19]. In addition to optimizing the EV itself, efforts can be put into optimization of charging infrastructure [20] and energy efficient route planning [21]. Zhang and Yao [5] assert that energy consumption analysis is the basis for studying location of charging infrastructures, ICE vehicle eco-driving behavior, and energy-saving route planning, which all contribute to extend EV range. ICE vehicle eco-driving behavior forms the foundation of this research to understand the influence of DB on energy consumption.

For ICE vehicles changing DB has been discussed as a possible approach to reduce fuel consumption and thus the environmental impact [30]. The advantages of this is that these action could be applied by a great number of people and having an immediate effect without additional costs [30]. The goal is to change driving behavior in a way that eco-driving becomes the norm rather than the exception. It is estimated that eco-driving can reduce the fuel consumption by 10% up to even 20% [30, 35]. To reach that number a sophisticated, multidimensional approach would be required involving education regulation, fiscal incentives, and social norm enforcement. Especially, the use of feedback devices on DB is emphasized. Currently, actions in educating DB are not implemented in this scope in the United States [30].

Even though educating drivers in eco-driving is discussed in recent literature less is known about the effects of different learning methods [35]. Experiments with drivers giving them eco-driving advices on the one hand and providing them with comprehensive eco-driving training on the other hand are compared for their efficiency [35]. Both education types influenced fuel consumption, average speed, and average acceleration positively [35].

Speed and acceleration are fundamental parameters for describing the motion of a vehicles, therefore, there parameters are also crucial for describing DB [36]. To understand DB dynamic data of the vehicles motion in the scope of a real-road test is useful [36]. Smartphones equipped with a suitable application have proven to be a recording device for driving data but are also investigated as a possible feedback device for drivers [36].

Literature agrees that driving behavior has a great influence on energy efficiency. For ICE the difference in energy efficiency for aggressive drivers is estimated to be 40% higher than for non-aggressive drivers [37]. To reduce the environmental impact the idea arose to educate drivers to adopt an eco-friendly driving style [38]. Eco-friendly driving behavior could be achieved by avoiding strong acceleration or braking in longitudinal and lateral direction [39]. The evaluation of real road driving scenarios is a complex problem since there are various interconnected variables like road type but also road environments, road infrastructure, and traffic conditions. However, two main parameters have shown to be most significant for quantitative evaluation of aggressive driving behavior these are longitudinal and lateral acceleration and deceleration [40].

Furthermore, aggressive driving is considered to be related to two driving patterns, a) strong acceleration or braking and b) driving at high speed. To determine aggressive acceleration, previous research used a Safe Driving Region (SDR) within a friction circle [38]. The friction cycle represents an area of possible acceleration depending on road conditions (e.g. dry, wet, icy) and tire grip. The SDR is defined as an area within the friction cycle of wet roads that applies an amount of mental workload on drivers, which is mainly determined by acceleration and speed, that ensures safe driving [38].

In literature in which acceleration data was collected using a smartphone, aggressive participants had more than 10% of their acceleration measurement points outside the SDR. For the safe drivers it was less than 8% [38]. Thus, a threshold for a share of acceleration measurement points inside and outside of SDR, based on the boundaries given in literature, is set up to distinguish aggressive from non-aggressive acceleration [38].

Eco-driving behavior is considered to be a key issue in research for reducing fuel consumption in ICE vehicles [35–37]. Few efforts have been made in the field of EVs so far [41]. These behavior analyses should be conducted using real DB that need instrumentations on vehicles for data collection which is limited [41]. Also, for EVs smartphones are used to provide the sufficient data [41]. Comparison of smartphone data and onboard instrumentation confirm that both sources are equivalent and that former is sufficiently accurate [41].

Factors for energy consumption in EVs can be classified into three major categories: internal vehicle-specific elements, external environmental elements, and individual driver-specific elements. The internal vehicle-specific parameters include mass, rolling resistance, aerodynamics, powertrain efficiency, the operational strategy (e.g., degree of RBS), and auxiliary energy (e.g., heating or air-conditioning). External parameters are inherent attributes of a chosen route, such as road type, topography, and traffic conditions. Individual driver-specific elements include a driver's individual style of driving based on their skills and attitude, all of which can strongly affect the energy consumption. To determine the effects of these parameters on the estimate of SOC empirical data is needed.

CHAPTER 3 - METHODOLOGY

Previous research was limited by the fact that the energy consumption derived from the SOC was a stepwise scale due to the BEVs setup. In the Volkswagen (VW) e-Golf^{TM,} SOC is calculated responding in 0.5% steps of total SOC which occurred approximately every half mile. This configuration would only detect drops of energy over relatively large distances and without determining phases of energy recuperation due to braking. The first step, thus, was to improve data collection for this research by finding a methodology that determines energy consumption more accurately in order to gain a better understanding of how energy is consumed by individual drivers along the test route.

3.1 Experiment layout

Even with the current graphical user interfaces and electronic data on various devices, it was not possible to obtain information on energy consumption from the vehicle directly, thus a suitable proxy was required. The on-board computer does report the battery's current (measured in Amperage [A]) and voltage (measured in Volt [V]) at a high resolution which results in electrical power (measured in Watts [W]) when multiplied with each other, according to Ohm's Law [42]. A similar approach was used by Wu [34] to determine the energy consumption of an BEV which provides a simple, while still accurate, result. The relationship was used to calculate the energy consumption at a continuous level for this research. Even though energy consumption was calculated this way, the approach used in previous research was used for the analysis in this research to make results comparable.

A vehicles' computer is referred to as the On-Board Diagnostic system (OBD) which was used for the data collection. The OBD communicates over the Controller Area Network (CAN) bus which is the standard solution to realize fast and robust communication of microcontrollers in vehicles [43]. The 2015 VW e-Golf™ uses an OBDII port which is an improved version of the originally OBD in terms of an enhanced communication protocol and standardization. The CAN bus system can be accessed through the OBDII port, located inside of vehicles, where the information is optimized for machine reading (i.e., the data is encoded and not available in a readable alphabetical text). For older vehicles most of the codes are available online. For newer vehicles, and especially for the VW e-Golf™, the codes are strictly protected, likely to prevent reengineering on the car through competitors or potential hackers. In previous research, great lengths went into decoding this information and finding the values that represented these desirable parameters. Additionally, previous research data was collected from three different sources (the CAN bus system, a GPS responder, and a cellphone) and merged in order to prepare it for data analysis which required an effort.

After investigating several options, a new company was utilized to facilitate these issues and to obtain this data, FleetCarma, a company based in Waterloo, Canada, specializes in extracting real-time driving data from all types of vehicles, including EVs and BEVs. Choosing their solution provided one device responsible for the GPS data and CAN bus data, thus requiring no additional processing. Also, the collected data is available in real-time on an online portal which decreases the feedback time and any additional process time. Problems in the data collection process would be, therefore, discovered faster and could be solved sooner without having a significant loss of data.

A suitable test route needs to be representative of Rhode Island in terms of road type variation and landscape. In order to establish an adequate test route, several options were evaluated based on the opportunity to drive on different road types with varying levels of elevation while considering traffic volume. The route required

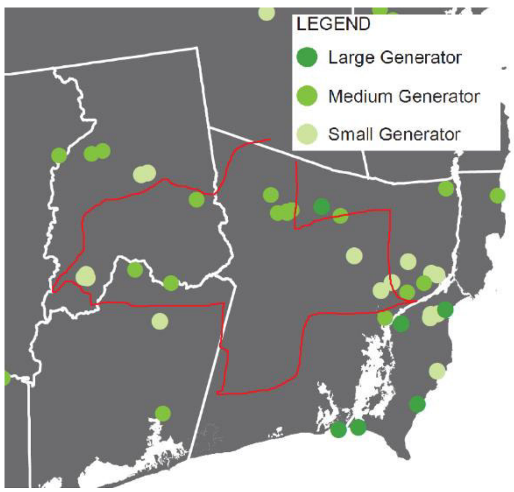


Figure 2: Traffic Generation [45]

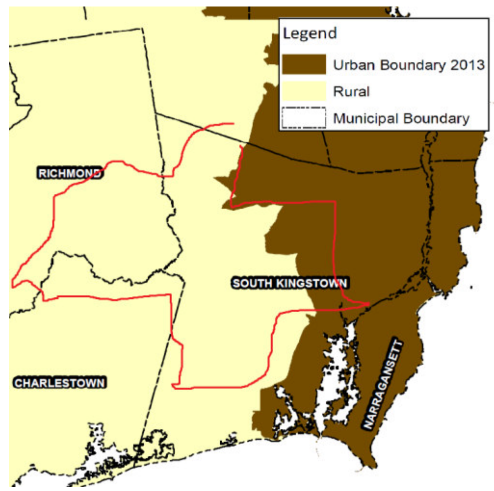


Figure 2: Urban and Rural Boundaries [45]

consistent traffic density considering it was not controlled for traffic concentration in the experiment due to its significantly complexity. Finding a route that is located in a low traffic density area would ensure that the variations in traffic concentration would range from low to medium which should minimally affect the traffic flow for the street network in the test area. Figure 2 illustrates the traffic generation in Washington County in the South Kingstown area using various color dots, representing low traffic generation (light green), medium traffic generation (green), high traffic generation (dark green). Traffic concentration is strongly determined by the time of the day with peaks generally in the morning and in the early evening due to work commutes [44].

To avoid these peaks and to ensure similar traffic conditions, test drives where conducted between 10am and 5pm. The test route progressed through urban and rural areas representing different road networks and development of infrastructure. Figure 2 illustrates rural boundaries (light brown) and urban boundaries, such as Providence's metropolitan area (dark brown) in the test area of South Kingstown.

Generally, roads are classified according to their function of either providing direct access to property or providing travel mobility. With respect to these two opposing functions the U.S Department of Transportation (DOT) distinguishes roads by six major classifications in descending order with respect to mobility: Freeway/Expressway, Principal Arterial, Minor Arterial, Major Collector, Minor Collector, and Local roads. Expressways, for example, exhibit high mobility with limited access with exit lanes, while local roads provide a high degree of land access [45].

The test route chosen selected was the same route used in a previous experiment [6] since all these factors were already considered. The past route covers a great range of road types (i.e., five out of six road types) while, progressing through an area of small to medium traffic generation that includes rural and urban areas. Conducting test drives along the same route also opens the potential towards combining Dataset 1 (from previous research [6]) and Dataset 2 (generated in this research) comparable if they are found to not be statistically different. Figure 3 depicts the test route starting and ending close to the University of Rhode Island (URI) including the different road types displayed in different colors, minor collectors (green), major collectors (blue), minor arterials (yellow), principal arterials (purple), and Expressways (black). The total mileage of the test route was 26.4 miles (42.5 km) from the start to end.

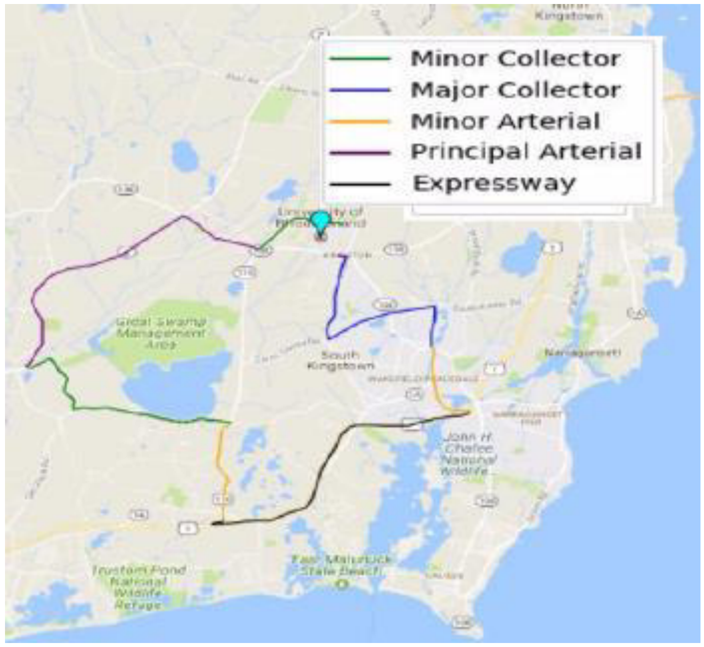


Figure 3: Test route [6]

Regarding the design of the experiment, drivers were recruited from the public. Primarily students and employees of URI participated. Participation in the experiment was voluntarily without financial compensations. Based on the self-selection of the participants, potential bias could exist based on their interest in electric vehicles or other sustainability related topics. Participants were encouraged to drive as they normally would in order to avoid purposeful driving issues due to being in an experiment. For example, more cautious driving in terms of energy efficiency due to driving another vehicle that is not their own. This process occurred the same for both datasets, Dataset 1 and Dataset 2, since participants were recruited the same way.

The conditions for the participants during the test drives were kept the same for both experiments. Additionally, the individual who executed Dataset 1 trained and advised the experimenter in Dataset 2. The data collected for both experiments included a timestamp, longitude and latitude, speed, altitude, acceleration, and SOC from the vehicle during the test drive. Battery current and battery voltage were obtained from the OBDII for a more accurate view of the actual energy consumption in Dataset 2. The experimental design was approved by the University of Rhode Island's Institutional Review Board (IRB). The documentation for the experiment can be found using the IRB reference number HU1617-055.

The exact same parameters for Dataset 2 were collected in previous research investigating the energy consumption on different routes for BEVs, indicating that these parameters are significant for understanding energy consumption [34].

3.2 Comparison of Dataset 1 and Dataset 2

To determine whether Datasets 1 and 2 are statistically similar, SOC is used as a proxy for total energy consumption over the entire test route during the separate drives. Table 1 gives an overview of the test drives conducted for both datasets including the number of male and female participants and the sample size that was used for data analysis after excluding poor data and outliers. However, Dataset 1 was analyzed for both scenarios, including the outlier and excluding the outlier.

Table 1: Break down of Dataset 1 and Dataset 2

Dataset	Collected	Poor data	Outlier	Male	Female	Sample data
1	38	11	1	21	13	29(30)
2	34	8	0	24	10	23

The fully charged battery contains 24.2 kWh which represents a 100% SOC. The consumed SOC for one test drive is related to this value, which results in the total energy consumption over the trip. Based on the reduction of SOC, both datasets are analyzed using an ANOVA to determine whether the two samples have a significantly different means. Both datasets are tested for normality and equal variance prior to the ANOVA. If, both datasets proved to be statistically similar for total energy consumption, then they could be combined to one larger dataset. This would increase the sample size in Dataset 1 (n=29, n=30) with those occurred in Dataset 2 to allow further investigation of statistically significant classes of drivers with respect to their energy consumption [6]. The combined dataset is analyzed regarding the variance of SOC classed by instructor (i.e. Dataset 1 and Dataset 2). The null hypothesis (H_0) is that the two datasets are statistically similar. Table 2 illustrates the SOC at the beginning of the test drives for both datasets. For Dataset 1 as well as for Dataset 2 approximately 2/3 of the test drives had a starting SOC higher than 70%. Since the discharging curve for lithium ionic batteries is fairly stable until 50% this ensures that the amount of energy drawn from the battery for different starting SOCs is similar.

Table 2: Stating SOC for Dataset 1 and Dataset 2

Dataset	100 - 90%	90 - 80%	80 – 70%	70 – 60%	60 – 50%	50 – 40%	Less than 40%	Sample data
1	9	7	4	4	2	4	0	29(30)
2	7	5	3	3	5	0	0	23

3.3 Dataset 2 Data Cleaning

Originally, for Dataset 2, there were 36 test drives performed. Significant thought was put into designing a data collection method that would be robust against errors by streamlining the process. Even with a specialized commercial device for capturing participants driving behavior along the route, only 16 test drives had all parameters collected without any issue. A total of 20 test drives had minor or major data collection problems, which limited their usage in this research. However, since in this context the aggregated data is analyzed over the entire test route, the GPS signal needed to be accurate only at the beginning and the end of the test route to determine start and end points of the test route.

There were several samples in Dataset 2 that matched these reduced requirements. Thus, seven more drives had an intermediate effort in post processing necessary due to either loss of GPS signal during the test drive or due to minor changes in the test route caused by construction going on at the on-ramp to Interstate 1. The GPS signal was still accurate enough to determine the coordinates where the vehicle was for the original test route and for an unexpected detour, which was consequently stripped from the dataset. This led to a loss of data (test drive 16 and 17) for a road segment of approximately 0.4 miles (650m), which, when compared to the entire test route did not have a significant effect on energy consumption. For 5 five other test drives, the test route was altered due to construction as well, but the GPS measurements where inaccurately recorded that the former method of removing the detour data was not applicable. Therefore, using these two different methods for determining the distance might have caused issues in further analysis.

Furthermore, seven test drives, in addition to the above-mentioned issues, resulted in data collection stopping at around 1/2 to 2/3 of the 26.4 miles (42.5 km) long test route so this data would have been available only for a part of the test route. For two samples, the data collection process stopped directly after the start or did not start at all so that there was no possibility to use this data.

The final result was 23 samples available in Dataset 2 of which 5 were tailored based on extraneous circumstances and 16 collected without issues. Each test drive contained approximately 5000 measuring points. As mentioned above energy consumption was determined based on SOC consumption, which could be recorded only in 0.5% intervals. SOC dropped by this value approximately every half mile, so, the exact energy consumption was known only at this point. To estimate the energy consumption for measuring point in between these drops of SOC, the theoretical difference in SOC between two values was calculated based on the distance traveled. This resulted in incremental SOC values (WeightSOC) for every measuring point. Since the fully charged battery contained 24.2kWh, the WeightSOC was used to calculate incremental energy consumption values (WeightkWh) for every measuring point using Equation 1.

Equation 1: Calculation for incremental energy consumption values

$$WeightkWh = totalenergy * \frac{WeightSOC}{totalSOC}$$

The distance between each measuring point was calculated using the Haversine formula (Equation 2) which determines the distance between two coordinates, latitude and longitude values of GPS, on a sphere. The formula for calculating the distance is shown in Equation 3.

Equation 2: Haversine formula

$$hav(\theta) = sin^2\left(\frac{\theta}{2}\right) = \frac{1 - \cos(\theta)}{2}$$

Equation 3: Distance calculation based on haversine formula

$$Distance = 2Rarcsin(\sqrt{hav(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\, hav(\delta_2 - \delta_1)})$$

Due to the fact that GPS was recorded at a lower resolution (every 10 seconds) than the other parameters of Dataset 2, the dataset got condensed so that for every measuring point there would be an individual GPS value available. This reduced the number of measuring points per individual per test drive from approximately 5000 to 153-305. Table 4 gives an overview of the number of measuring point for each test drive of Dataset 2. The reason for the few number of measuring points, especially for Driver 24, is that for some test drive GPS data got collected on an even lower resolution than motioned before. Table 3 shows the traveling time for each test drive of Dataset 2.

Table 3: Statistics for amount of time per individuum for finishing the test drive

Variable	Total	Mean	SE Mean	StDev	Minimum	Maximum
Time	23	51.670	0.966	4.632	44.270	63.040

Table 4: Statistics for number of measuring points of collapsed Dataset 2

Variable	Total	Mean	SE Mean	StDev	Minimum	Maximum
N	23	256.00	6.60	31.64	153.00	305.00

3.4 Determination of Energy Consumption Clusters

Two different types of clustering occurred in this research in order to identify the appropriate clusters of driving behavior based on energy consumption (difference in SOC in kWs) per mile of the experimental test route. These two methods are: (1) hierarchical clustering using Ward’s method and (2) model-based clustering using Latent Profile Analysis (LPA). In previous research, hierarchical clustering using Ward’s method was used to cluster drivers with respect to their energy consumption [6]. This method, also known as Ward’s Minimum Variance Method, begins with *n* clusters where each sample is in one group and then it merges two groups at each step and repeats until all samples are in a single group after n-1 steps. The criterion for choosing a pair of groups, from all grouping possibilities, is to merge these pair whose potential pairing minimizes the sum of squared distances between those two individual groups and the centroids of their respective group, summed over the resulting groups [46].

Equation 4: Ward's Minimum Variance Method [46]

$$W = \sum_{g=1}^{G} \sum_{i=1}^{n_g} \left| \left| x_i - \bar{x}_g \right| \right|^2 = \sum_{g=1}^{G} \sum_{i=1}^{n_g} \sum_{k=1}^{K} (x_{i,k} - x_{g,k})^2$$

Equation 4 is calculated for all possible pairs of groups and is exhaustive throughout evaluating a dataset. For each pair, the centroid (clustered or group mean) and the squared distances are calculated based on their new values. Ward's method minimizes the variance of within-groups variances, over the possible combinations (K), while maximizing the distances between groups [46].

Clustering drivers' behavior was based on their mean and standard deviation of their energy consumption per mile. In previous research, Ward's method revealed two significantly different clusters representing efficient and inefficient drivers [6]. Even though two clusters were found that were significant, there was strong evidence that there might be a third cluster of moderately efficient driving. However, this one was found to be not significantly different from the inefficient cluster after a rigorous validation process. The possible explanation provided for this in the literature is the limited sample size of Dataset 1 [6], hence the rationale for the expansion of Dataset 1 by executing Dataset 2.

The data from Dataset 1 is clustered again with Ward's method to validate these initial findings. In congruence with this process, the drivers from Dataset 2 are clustered according to their energy consumption using Ward's method (*Ward.D2* in the *hclust* package in R) to make the groups of drivers comparable for both datasets.

A disadvantage of hierarchical clustering algorithms, like Ward's method, is that they are largely heuristic and not based on formal models per se but Euclidean distance. Model-based clustering is proposed as an alternative [47]. The basic idea behind model-based clustering is that observations from the sample population arise from a distribution that is a combination of two or more components. Each component is described by a density function and is associated to a probability within the combination of components often a combination of multivariant normal distributions. These components represent the clusters and have a shape with the mean respective of the cluster [48, 49].

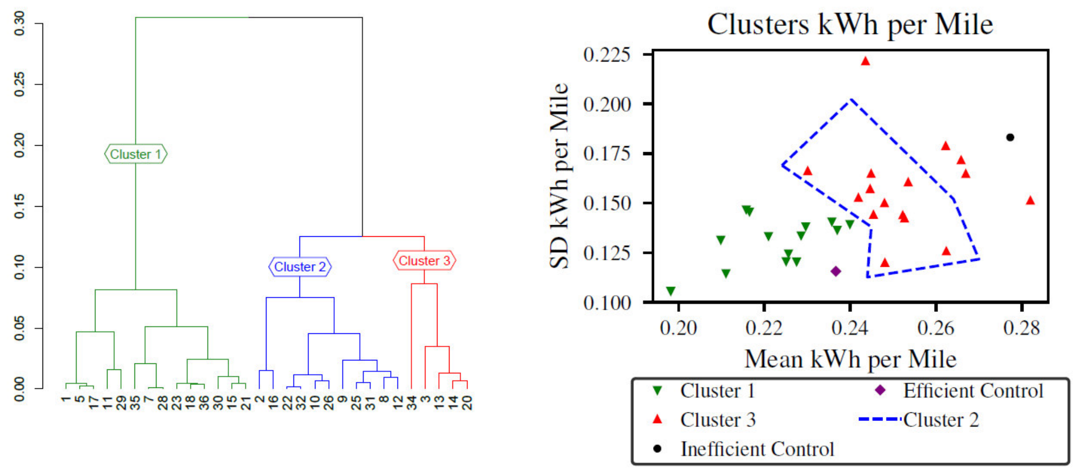
For implementing a model-based clustering algorithm in R the package *mclust* was used, which allows for a total combination of 10 different volumes, shapes, and orientations of the ellipsoidal shapes based on various Gaussian distributions. *Mclust* uses three different letters to describe the characteristics of the shapes E for equal, V for variable, and I for coordinate axes and reports the model in terms of volume, shape, and orientation. For example, an EEI model represents the resulting clustering groups have equal volume, equal shape, and their orientation is equal to the coordinate axes [50, 51]. This method of execution is sometime called LPA, which is ran in R using *mclust*, to determine parameter estimates and grouping according to these Gaussian distributions and shapes. The selection criterion of the Bayesian Information

Criterion (BIC) was used thus a model with a lower BIC fits the data better than one with higher BIC, also BIC penalizes large models [50].

CHAPTER 4 - RESULTS AND DISCUSSION

4.1 Determination of Energy Consumption Clusters for Dataset 1

In previous research two significant clusters for energy consumption were found, an energy efficient driver cluster and an energy inefficient driver cluster [6]. Figure 4 a



(a) Hierarchical Clusters (b) Clusters Spread

Figure 4: Dendrogram and Scatterplot of Dataset 1 for mean energy consumption per mile and standard deviation [6]

shows the dendrogram generated by Ward's method for pairs of mean energy consumption per mile and the respective standard deviation. The green branch in the dendrogram represents the energy efficient cluster with low mean and low standard deviation. The blue and the red branch in the dendrogram represent the energy inefficient cluster, whereas the blue branch represents the medium energy efficient cluster which was not found to be significant in previous research. In the scatter of Figure 4 b plot green triangles represent the points of the energy efficient cluster and red triangles represent points of the energy inefficient cluster whereas red triangles incorporated by the blue dashed line are part or the assumed but not significant medium energy consuming cluster. The black point represents the inefficient control the purple diamond represents the efficient control.

For this research the analysis on the data from Dataset 1 is repeated to confirm the results. Individuals are clustered the same way as in [6], by the distance of each point to the others based on mean energy consumption per mile and standard deviation using Ward's method. Figure 5 shows the resulting dendrogram. Individuals are clustered in

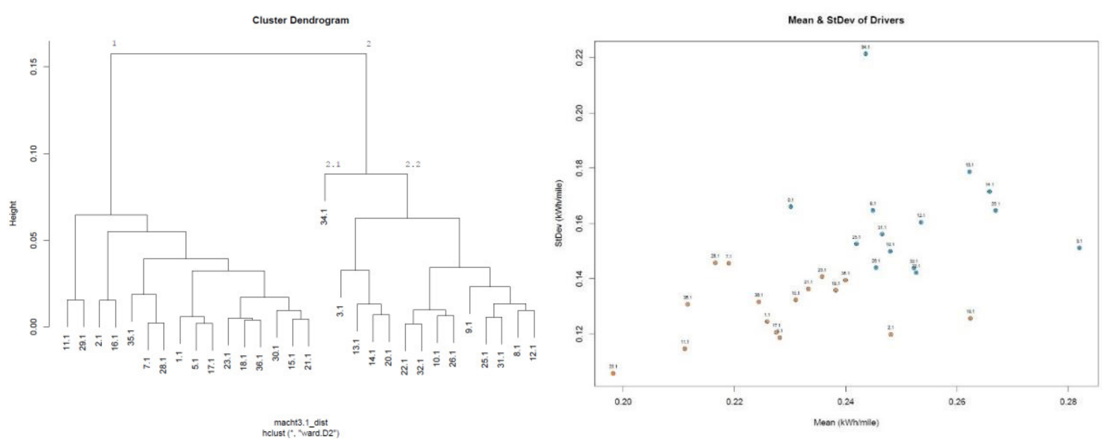


Figure 5: Dendrogram of Dataset 1 including individual 34.1

the same groups as in previous research with one exception. Individual 34.1 is clustered one group higher than in previous research, which is surprising since the same dataset and the clustering algorithm was used. As in previous research the clusters were found to be not uniformly normal distributed, therefore a non-parametric test, a Wilcox test, was used for validation of the clusters. For the Wilcox test only the two main clusters, cluster 1 and cluster 2, which divide the dataset into high energy consuming and low energy consuming drivers, were found to be significant. This is in accordance with previous research and confirms the results [6]

Ward's method, however, is sensitive to outliners. When analyzing the scatter plot in Figure 5, the point with the highest standard deviation, individuum 34.1, seems to be very far off. In order to obtain further insights on the whether a medium energy efficient cluster exist or not, this research conducts a LPA on Dataset 1. Analyzing

Dataset 1 with LPA creates a 2-component EVE (ellipsoidal, equal volume and orientation) model with a Bayesian Information Criterion (BIC) of 286.7642 and a log. likelihood of 158.6875, which results in two clusters of drivers. Looking at the scatter plot of this clusters in Figure 6, the shape of the clusters is very different. However, except for individual 20.1 the blue cluster incorporates the individuals of the most energy inefficient group and in contrast to the hierarchical clustering, LPA distinguishes this group from the rest of the drivers. This provides evidence that there is a difference between the inefficient drivers and the rest of the drivers. Noticeable is that individuum 34.1 has a significant higher standard deviation that the other drivers and is far off from the other points. Thus, the question remains whether this point is an

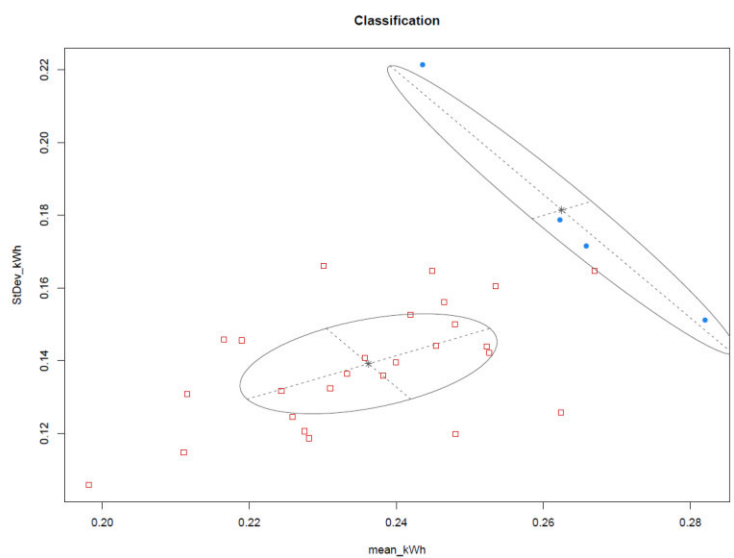


Figure 6: Latent Profile analysis of Dataset 1 including 34.1 for 2 clusters

outlier or whether it represents a group of highly inefficient drivers. In further analysis Dataset will be clustered without 34.1 and Dataset 1 will be merged with Dataset 2 and in these contexts the position of point 34.1 will be discussed again.

In order to see whether different clusters are found in Dataset 1 using Ward's method when individuum 34.1 is removed from the dataset, 34.1 is removed from

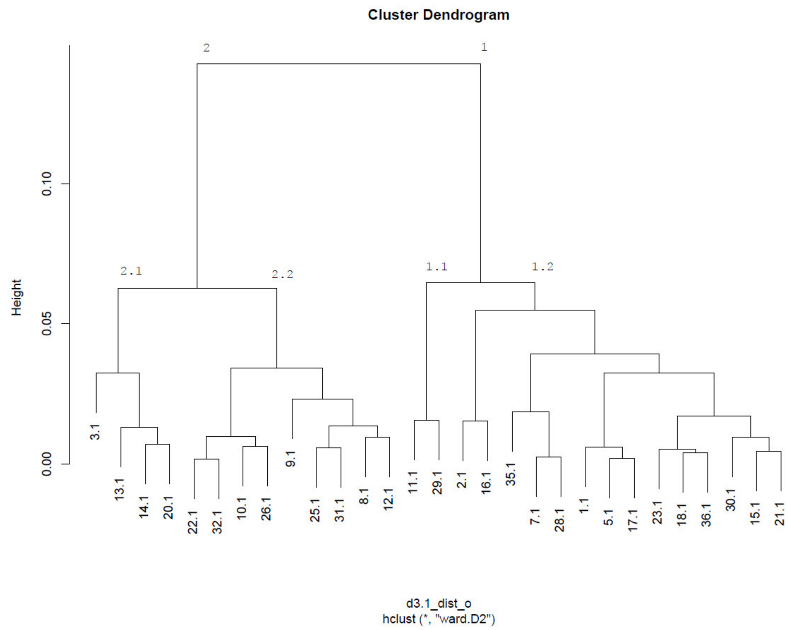


Figure 8: Dendrogram of Dataset 1 excluding individual 34.1

Dataset 1. Figure 8 shows the dendrogram for the hierarchical clustering of Dataset 1 excluding individual 34.1, using Ward's method, revealing that the clusters look exactly the same as before but without individuum 34.1 being a group on its own. The Wilcox test reveals that the clusters are significantly different for two clusters but also for three clusters (cluster 1.1, 1.2, and 2).

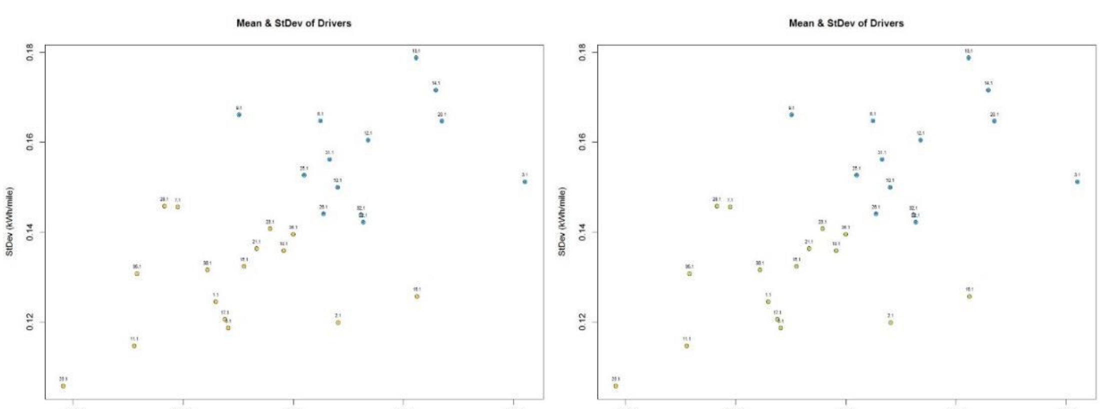


Figure 7: Scatter plot of hierarchical clustering for 2 clusters (left) and 3 cluster (right)

Figure 7 shows the scatter plots for two clusters on the left side and for three clusters on the right side, while individuals belonging to one group are colored in the same color. Finding three significantly different clusters for energy consumption of drivers using Ward's method, confirms the assumption from previous research that drivers can be divided into energy efficient drivers, medium energy efficient drivers, and energy inefficient drivers. These results drive the assumption that there are probably multiple driver clusters based on energy consumption. It might be possible that Ward's method is not an appropriate method to detected different groups of drivers, especially when clusters contain a small number of samples. LPA is used to get further insights on a possible cluster distribution within the dataset.

Performing the LPA on Dataset 1 with reduced sample size of 29, excluding 34.1, two options are found to be convincing based on distribution of clusters, based on BIC and log. likelihood. The first one is a 2-component EEI (spherical, equal volume)

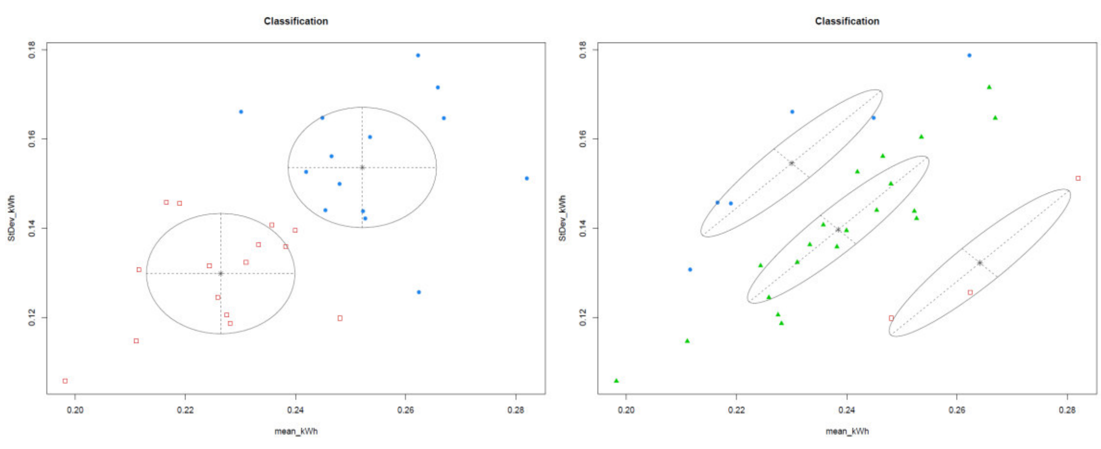


Figure 9: Latent Profile analysis of Dataset 1 excluding 34.1 for 2 clusters (left) and 3 clusters (right)

model with a BIC of 289.2305 and a log. likelihood of 154.7171. Beside of individuum 16.1, which is clustered to the energy inefficient group by LPA than to the energy efficient group according to the dendrogram, the clusters for Ward's method and the 2-component model LPA are the same.

The second one is 3-component EEE (ellipsoidal, equal volume, shape and orientation) model with a BIC of 287.3235 and a log. likelihood of 162.1819, which should be preferred over the first model, based on the BIC. However, the clusters for the 3-component model do not look very convincing based on what is known about the influence of driving behavior on energy consumption of BEVs.

4.2 Determination of Energy Consumption Clusters for Dataset 2

For the dataset generated in the scope of this research, Dataset 2, the drivers are clustered based on their energy consumption using Ward's method. Applying the hierarchical clustering algorithm to Dataset 2 results in the dendrogram seen in Figure 10.

Drivers are label based on their Driver ID, the decimal indicates that these samples are

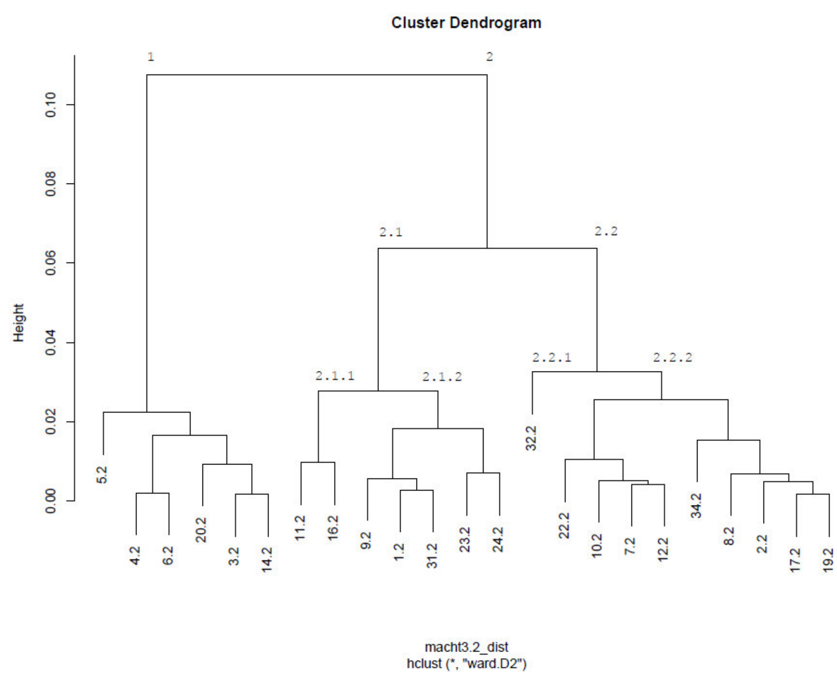


Figure 10: Cluster Dendrogram for Dataset 2 for overall SOC consumption

the drivers of Dataset 2. Clusters that were tested for significance are labeled on top of their branches. The clusters were found to be not uniformly normal distributed why a non-parametric test, a Wilcox test, was used to determine significantly different clusters. The Wilcox test found two clusters as well as three clusters (cluster 1, 2.1, and 2.2) to be significantly different with respect to mean energy consumption per mile and standard deviation.

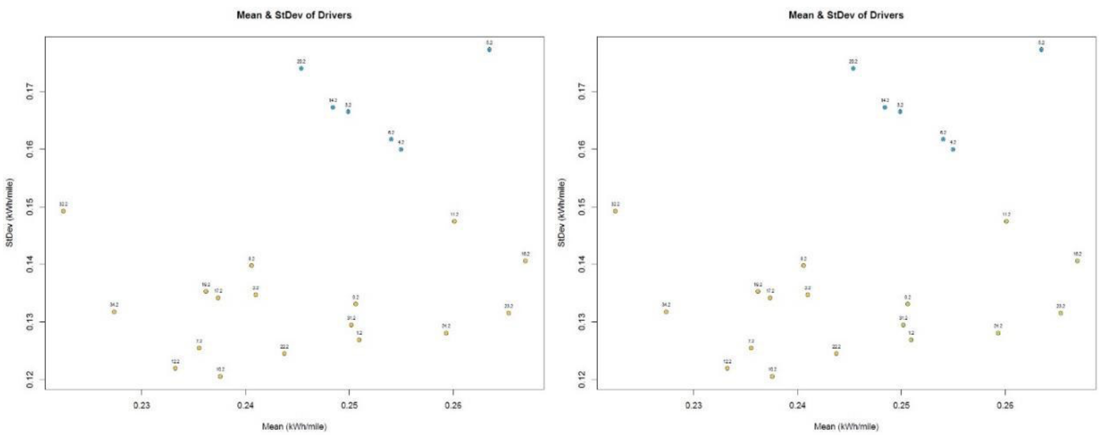


Figure 11: Scatter plot of HC for 2 clusters (left) and 3 cluster (right)

For the three clusters of Dataset 2, drivers are more equally distributed than for Dataset 1 which results in three clusters of energy consumption with a representative number of drivers for each cluster. With Dataset 2 having three significantly different clusters, this supports the assumption from previous research that there are three significant groups for clustering drivers, high energy consuming drivers, medium energy consuming drivers, and low energy consuming drivers.

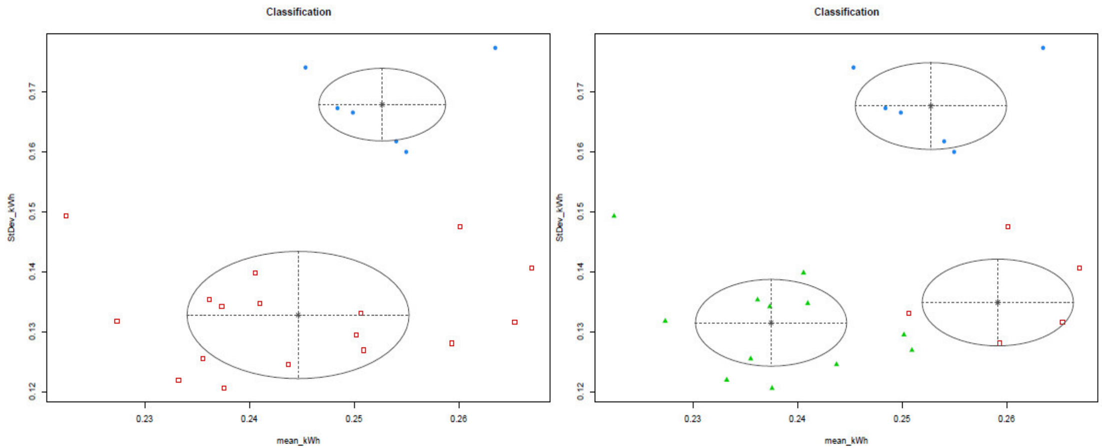


Figure 12: LPA for Dataset 2 2 component model (right) and 3 component model (left)

To prove the two clusters found by Ward's method, LPA is applied to Dataset 2. Comparing the LPA 2-component VII (spherical, varying volume) with the two clusters from the dendrogram reveals that both algorithms cluster the same drivers into the two clusters. This provides evidence that the two clusters found by Ward's method are correct based on a two-cluster monitoring.

Comparing the three clusters from the dendrogram with the 3-component EVV (ellipsoidal, equal volume) model from LPA, it shows the same clusters, except for individuals 31.2 and 1.2, which are part of the lower left cluster in the LPA. Figure 12 shows the 2-component LPA clustering on the left and the 3-component LPA clustering on the right side.

The comparison of the hierarchical clustering by Ward's method with LPA shows that beside of little differences Ward's method and LPA find the same or similar clusters which is a lead that Ward's method is an appropriate approach for clustering drivers based on their energy efficiency. For the clusters of Dataset 2, the clusters of Ward's method are more similar to the ones of LPA. One of the reasons could be that for Dataset 2 the population for three hierarchical clusters is more equally distributed between the three clusters than for Dataset 1. For two clusters the populations for Dataset 1 and Dataset 2 are similar distributed which also results in similar clusters for Ward's method and LPA in both cases. Merely individuum 16.1 is clustered differently in Dataset 1 when divided into two clusters and using Ward's method and LPA respectively. For Dataset 2 the two clusters from Ward's method and LPA are identical.

For both datasets there are two and three significantly different clusters found based on Ward's method. However, the two datasets, Dataset 1 and Dataset 2, have data in different areas. Dataset 1 provides measurements in the upper right quadrant where Dataset 2 lacks measurements. Dataset 2, on the other hand, provides measurements exclusively in the lower right quadrant. To obtain a more holistic understanding of how BEV drivers can be categorized based on their energy consumption, both datasets are test whether they are statistically similar and merge them into one Combined Dataset, if they should be found similar.

4.3 Comparison of Dataset 1 and Dataset 2

Since the experiment layouts for Dataset 1 and 2 were almost the same, beside the instructors and the two parameters Amperage and Voltage that got collected additionally, it was expected that both datasets are statistically similar. First both datasets were compared based on their mean SOC consumption over the entire test route.

Table 5: Descriptive Statistics for Dataset 1 and Dataset 2

Variable	INSTR	Mean	SE Mean	StDev	Variance
SOC	1	26.483	0.387	2.119	4.491
	2	27.826	0.329	1.578	2.491

Variable	INSTR	Minimum	Q1	Median	Q3	Maximum
SOC	1	21.500	24.875	27.000	27.500	31.500
	2	24.500	27.000	28.000	28.500	31.500

Comparing the means of SOC consumption of both datasets they are found to be close together. Dataset 2 has a slightly higher mean by almost 1.5% than Dataset 1 which is not a lot when considering that this is within one standard deviation of Dataset 2. For Dataset 1 the standard deviation is with 2.119 slightly higher than 1.578 from Dataset 2. The difference for the median is with exactly 1% even smaller than the mean. Table 5 illustrates the results of the descriptive analysis.

Analyzing the distribution of SOC consumption of both datasets for normality reveals that for both datasets SOC consumption is normal distributed. Furthermore, both datasets have equal variances based on SOC consumption. Conducting the ANOVA testing for Dataset 1, including 30 samples, and Dataset 2, including 23 samples, revealed that the two datasets were not statistically similar for SOC. Table 6 shows the ANOVA resulting in a p-value of 0.014 and Table 7 shows the results from Tukey test, indicating that both Datasets are significantly different. The reason that both datasets are different even though their means are close together is that Dataset 1 has a greater spread in SOC consumption than Dataset 2.

Table 6: Analysis of Variance of SOC for combined Dataset with 30 samples for Dataset 1 and 23 samples for Dataset 2

Source	DF	Adj SS	Adj MS	F-Value	P-Value
INSTR	1	23.47	23.473	6.47	0.014
Error	51	185.05	3.628		
Total	52	208.52			

Table 7: Comparison of SOC for combined Dataset with 30 samples for Dataset 1 and 23 samples for Dataset using Tukey test

INSTR	N	Mean	Grouping	
1	23	27.826	A	
2	30	26.483		B

Means that do not share a letter are significantly different.

This analysis was performed on overall SOC consumption showing that the aggregated energy consumption of individuals driving on the test route does not vary as much as assumed based on literature.

However, to understand how driving behavior influences energy consumption, energy consumption must be analyzed not as an aggregated value but trough out the test route. For this reason, incremental energy consumption was calculated for every measuring point of the test drive and the means as well as standard deviation was calculated for each measuring point for each driver. Performing the ANOVA for mean energy consumption per mile for both datasets results in a p-value of 0.118 revealing that both datasets, based on their mean energy consumption per incremental distance, are statistically similar. This result is confirmed by a Tukey test. Table 8 shows the results from ANOVA for mean energy consumption per mile and Table 9 show the result from Tukey test. These results show that for mean energy consumption per mile, which represents driving behavior at a more granular level, the two datasets are statistically similar meaning that they can be treated as one dataset.

Table 8: Analysis of Variance of mean energy consumption per mile for combined Dataset with 30 samples for Dataset 1 and 23 samples for Dataset 2

Source	DF	Adj SS	Adj MS	F-Value	P-Value
INSTR	1	0.000668	0.000668	2.53	0.118
Error	51	0.013439	0.000264		
Total	52	0.014107			

Table 9: Comparison of mean energy consumption per mile for combined Dataset with 30 samples for Dataset 1 and 23 samples for Dataset 2 using Tukey test

INSTR	N	Mean	Grouping
2	23	0.24672	A
1	30	0.23956	A

Means that do not share a letter are significantly different.

For Dataset 1 it was not clear whether there was an intermediate energy efficient driver cluster since it was found to be not significantly different from two inefficient and efficient groups. It was assumed to be a result of insufficient sample size [6]. This research proved that there are three clusters for Dataset 1, when individuum 34.1 is excluded, and for Dataset 2. However, since both datasets provide data in complimentary areas the question remains whether the found clusters are consistent when merging both datasets. Therefore Dataset 1 and Dataset 2 are tested from similarity based on their mean energy consumption per mile and are found to be statistical similar which allows to merge both datasets into one Combined Dataset to increase the sample size of test drives to draw a clearer picture of how to cluster BEV drivers based on their energy consumption.

4.4 Determination of Energy Consumption Clusters for Combined Datasets

To see whether a larger sample size would confirm or reject multiple clusters for energy consumption both datasets get, based on the fact that they are statistically similar, merged into one large Combined Dataset with a total sample size of 53.

The Combined Dataset was first clustered using Ward's method. Figure 13 shows the resulting dendrogram. Drivers are labeled based on their Driver ID, the decimal indicating origin from either Dataset 1 or Dataset 2. Clusters that were tested for significance, using a Wilcox test, are labeled on top of their branches. This test also found cluster 1 and cluster 2 to be significant, as well as clusters 1, 2.1, and 2.2 which

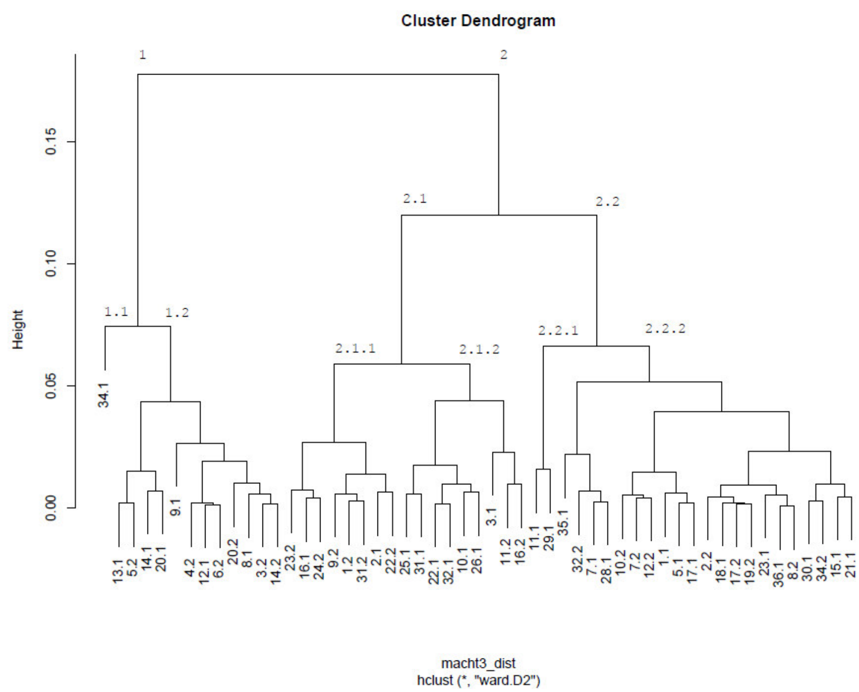


Figure 13: Dendrogram for Combined Dataset with 53 samples using Ward's method

confirms the assumption from previous research that there are three clusters, a high energy consuming, medium energy consuming, and a low energy consuming driver cluster. Figure 14 shows the scatter plot for the two hierarchical clusters on the left side and for three hierarchical clusters on the right side.

LPA is used to test whether based on cluster number found by Ward's method the clusters will look the same which is an indication that those are the actual clusters.

The 2-component EII (spherical, equal volume) model for LPA clusters the drivers into almost the same clusters as Ward's method., except for individuum 31.1.

The 3-component EEV (ellipsoidal, equal volume and shape) model, however, clusters a main part of the population in two clusters divided by standard deviation over a wide range of means and individuum 34.1 in a cluster of its own. Even though the populations are equally distributed between the clusters in the dendrogram, LPA considers individuum 34.1 to be a cluster on its own, which contradicts earlier assumptions that Ward's clustering and LPA would produce similar clusters as long as there are sufficient samples in a cluster for Ward's method.

However, the reason LPA clusters the population for the 3-component model differently could be that 34.1 is an outlier to which Ward's method is sensitive and should not be respected in the clustering process.

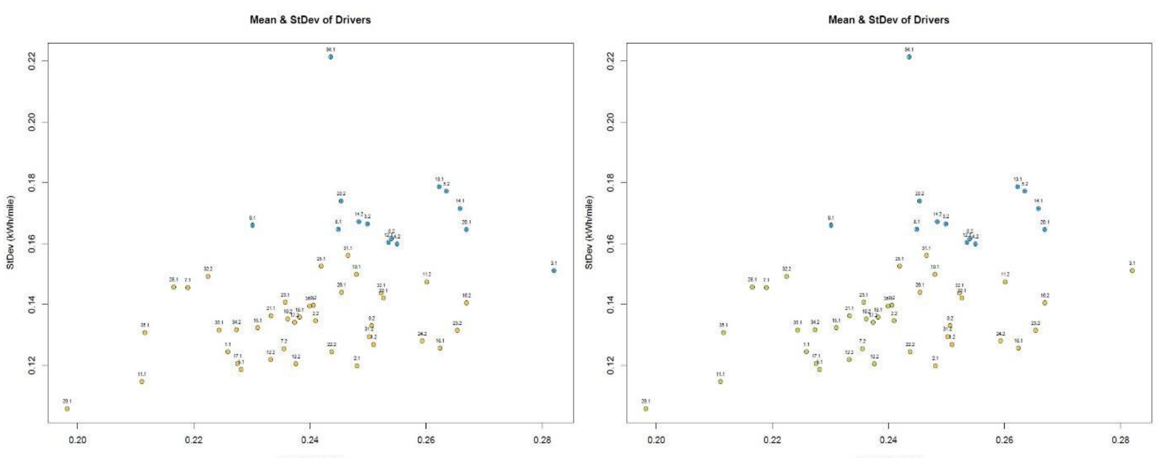


Figure 14: Scatter plot of hierarchical clustering using Ward's method for 2 clusters (left) and 3 cluster (right)

To understand whether more than three clusters might be possible LPA is extended to multiple clusters. Computing the LPA for a 4-component EVE (ellipsoidal, equal volume and orientation) model with a BIC of 520.486 and a log. likelihood of 293.9905 and a 5-component EEE (ellipsoidal, equal volume, shape and orientation) model with a BIC of 524.017 and a log. likelihood of 295.756 reveals additional possible clusters. The 4-component model clusters the population in three equal shape clusters for high mean and high standard deviation in the left scatter plot in Figure 15: LPA of Combined Dataset for 4-component model (left) and 5-component model (right) and one long elliptical shaped cluster for low mean and

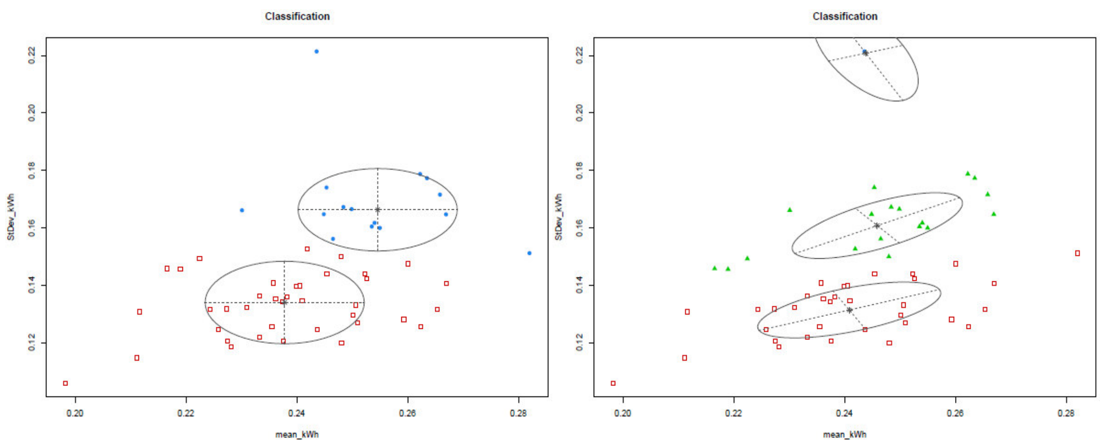


Figure 16: LPA analysis of Combined Dataset for 2 component model (left) and 3 component model (right)

various standard deviation including 34.1. The long elliptical shaped cluster could be

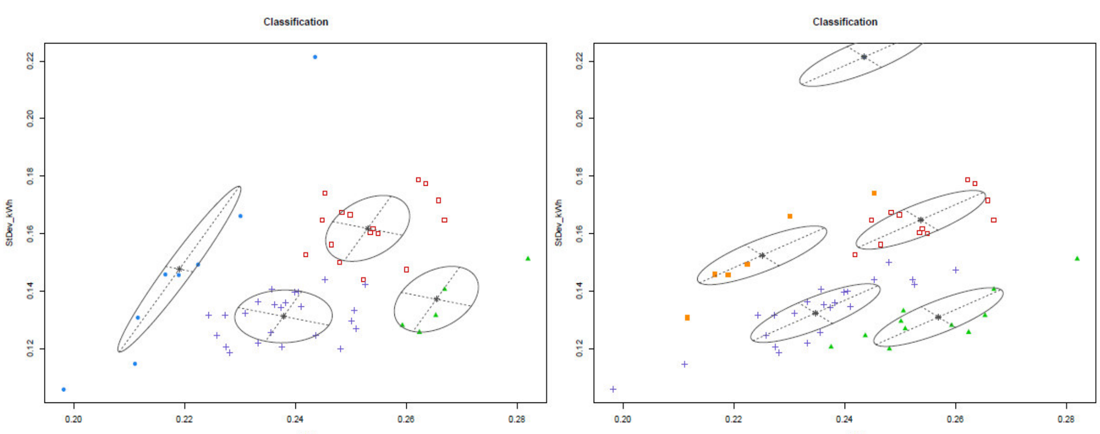


Figure 15: LPA of Combined Dataset for 4-component model (left) and 5-component model (right)

attenuated to the shape of the other clusters by excluding 34. The 5-component model clusters the population in four equal shaped clusters for the main part of the population and 34.1 in a cluster of its own. Both clustering models give evidence that 34.1 is an outlier and should be excluded from the population.

Based on the findings, that the Combined Dataset is clustered again with Ward's method and LPA, excluding 34.1, in order so see whether the earlier assumptions are confirmed or rejected.

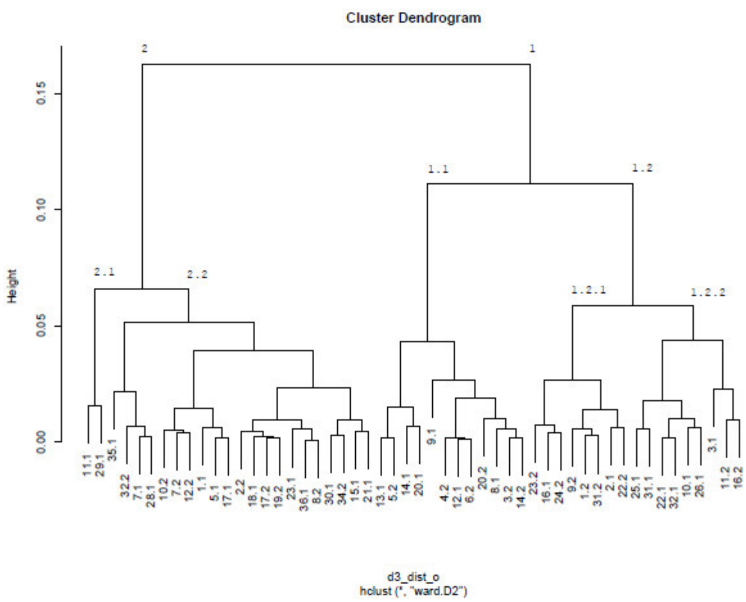


Figure 17: Dendrogram for Combined Dataset using Ward's method

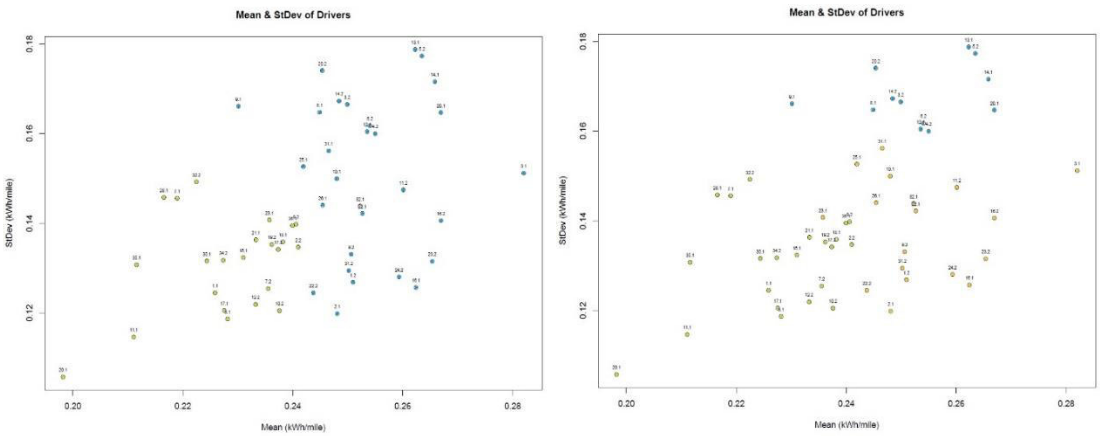


Figure 18: Scatter plot of hierarchical clustering using Ward's method for 2 clusters (left) and 3 cluster (right)

Clustering the Combined Dataset with reduced sample size (n=52) with Ward's method results in the dendrogram shown in Figure 17: Dendrogram for Combined Dataset using Ward's method. Validating possible cluster with Wilcox test reveals two significant different clusters, cluster 1 and cluster 2, as well as three significant clusters (1.1, 1.2, and 2), resulting in the same clusters that were confirmed for the bigger population of n=53. Figure 18: Scatter plot of hierarchical clustering using Ward's method for 2 clusters (left) and 3 cluster (right) shows the scatter plots for two clusters on the left side and for three clusters on the right side.

LPA is used to confirm the clusters found by using Ward's method. While Ward's method divides the dataset for two clusters between high and low mean, LPA divides it by high and low standard deviation. Both groups of clusters do not explain driving behavior very well based on what is known about energy consumption for BEVs in literature.

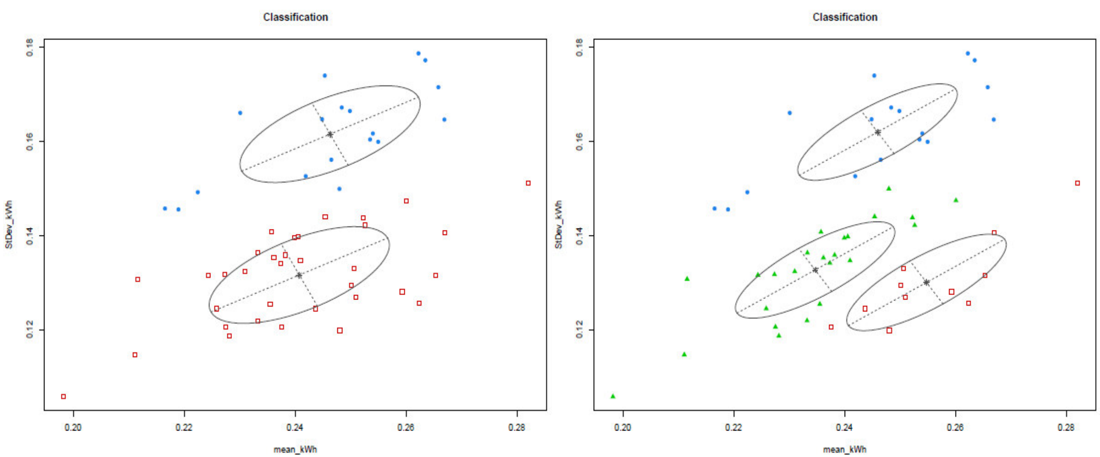


Figure 19: LPA of reduced Combined Dataset for 2-component model (left) and 3-component model (right)

Performing the LPA for 4 clusters results in a 4-component EII (spherical, equal

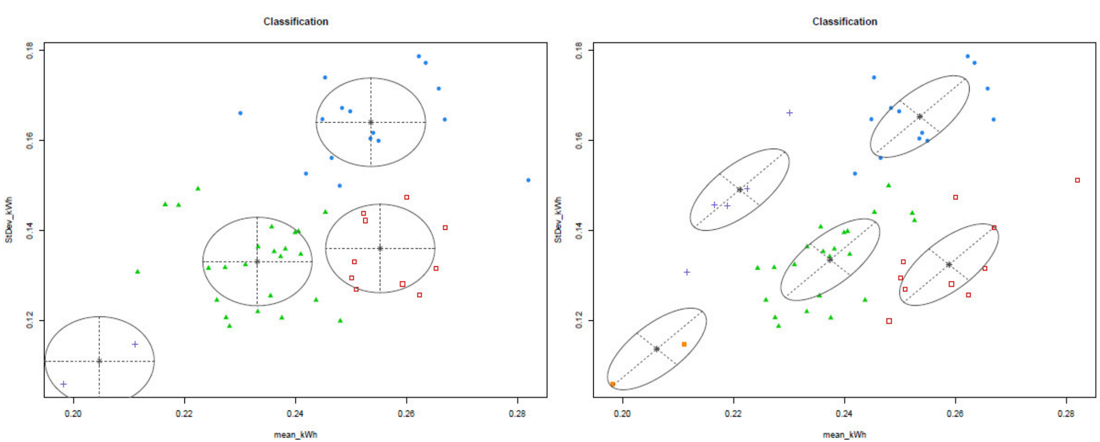


Figure 20: LPA of reduced Combined Dataset for 4-component model (left) and 5-component model (right)

volume) model that separates the population into 4 equally shaped clusters resulting in the four equally shape clusters that were assumed for the 4-component model of the non-reduced Combined Dataset when individuum 34.1 is excluded. A 5-component

EEE (ellipsoidal, equal volume, shape and orientation) model, in addition to the clusters found by the 4-component model, puts the individuals in the upper left corner in a group of its own.

Both models seem to provide a suitable clustering from BEV drivers based on their energy consumption based on the agreement in literature about the influence of driving behavior on energy consumption for BEVs. Hence, that energy consumption increases with mean energy consumption per mile but also with a high degree of variation in the mean energy consumption per mile indicating an agitated driving style.

Table 10: Energy consumption per cluster of LPA 4-component model for Combined Dataset (n=52) over entire route

Variable	LPA 4 clusters	Total	Mean	StDev	Min	Max	Mix. Prob.
Energy cons.	1	2	5.384	0.257	5.203	5.566	0.2897
	2	23	6.2973	0.3451	5.6870	7.0180	0.2086
	3	10	6.873	0.351	6.413	7.623	0.4645
	4	17	6.8401	0.3317	6.2920	7.6230	0.0373

Table 11: Energy consumption per cluster of LPA 4-component model for Combined Dataset (n=52) centroids

Variable	LPA 4 clusters	Total	Mean	SE Mean
Energy cons.	1	2	5.384	0.181
	2	23	6.2973	0.0720
	3	10	6.873	0.111
	4	17	6.8401	0.0804

Table 10 shows the average energy consumption per cluster of the LPA 4-component model. (1 - purple cluster, 2 - green cluster, 3 - red cluster, 4 - blue cluster). The energy consumption increases from cluster 1 to cluster 4, while cluster 3 and 4 have almost the same mean energy consumption. The individuals in cluster 1 consumed considerably less energy than the rest of the drivers. Table 11 displays mean energy consumption per mile and standard deviation for the LPA 4-component model's centroids.

To give an overview about the effect of driving behavior on the annual energy consumption of BEVs, calculating the kWh per year, the yearly fueling bill, and the yearly number of charging events occurred *a posteriori*. Based on the national average of 13,475 miles traveled per year and the mean energy consumption per mile as a function of the cluster results in the total annual energy consumption [52]. This total annual energy consumption multiplied by the rate of 13.1¢/kWh results in the total amount of money spent on recharging the BEV [53]. Lastly, the mean energy consumption per mile is divided by the total BEV battery capacity of 24.2kWh in order to understand the potential number of charges per cluster, assuming charging at a full charge each time.

There is a significant difference in the number of charges between the four clusters. Also, the amount spent for charging differs significantly, especially when considering mean and standard deviation.

Table 12: Driving Estimation for BEV Driving Behavior

Event	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
kWh/mile	0.204741	0.110993	0.233174	0.1331	0.255296	0.136004	0.253612	0.164074
kWh per Year	2759	1496	3142	1793	3440	1833	3417	2211
Yearly Fuel Bill	$ 361.39	$ 195.91	$ 411.57	$ 234.93	$ 450.62	$ 240.06	$ 447.65	$ 289.61
Yearly charging Event	114	62	130	74	142	76	141	91

CHAPTER 5 - CONCLUSION

The goal of this research was initially to understand the driver profiles for battery electric vehicles based on empirical driving behavior. This was executed in a four-pronged approach: (1) confirm previous literature driver profiles of energy consumption, (2) validate profiles by applying an additional clustering method, (3) expanding the original dataset, and (4) re-assess those energy consumptive behavior profiles.

(1) This research confirmed the two energy consumption clusters found in previous research for Dataset 1 using Ward's method, a cluster of energy efficient drivers and a cluster of energy inefficient drivers. Furthermore, a potential third cluster that was discussed in previous research was also found to be not statistically different when clustering the entire Dataset 1 with Ward's method. Since Ward's method is sensitive to outliers, potential outliers are discussed. This research finds driver 34.1 to be an outlier which is removed from Dataset 1 and Dataset 1 is clustered again. For this reduced Dataset 1 (n=29) two significant and three significantly different groups of drivers for energy consumption were found using Ward's method, an energy efficient, an energy inefficient, and an intermediate energy efficient group. Therefore, this research shows that even for Dataset 1 using Ward's method an intermediate energy efficient cluster exists when driver 34.1 is excluded from the dataset and thus, confirms the assumption of previous research.

Two and three statistically different clusters, which are similar to the clustering of the reduced Dataset 1, were found when hierarchical clustering was applied to Dataset 2. The reproducing of the results of Dataset 1 using a different dataset (Dataset 2) strengthens the argument that three clusters are significant.

(2) Furthermore, LPA is introduced as a new clustering approach to check the clusters found using hierarchical clustering. The introduction of LPA to the analysis methods augments the clustering procedure through a more advanced approach to find alternative clusters to the ones found in previous research [47]. Clustering Dataset 1 with LPA revealed a different picture of the clusters, the energy inefficient cluster was different from the rest of the driver population. This supports the assumption of previous research that the individuals of most energy inefficiency should be treated as a group of its own. The 2-component model from LPA performed on the reduced Dataset 1 confirms the two clusters found through hierarchical clustering by creating similar. Merely driver 16.1 is assigned to a different cluster than previously. For a 3-component model LPA finds different clusters. However, the clusters found in the 3-component LPA model are not convincing with respect to the literature available on the influence of driving behavior on energy consumption of BEVs.

Clustering Dataset 2 with an LPA 2-component model produces the exact same two clusters found by hierarchical clustering which provides evidence that the drivers are clustered correctly based on a hierarchical 2 cluster model. A 3-component model finds groups similar to the hierarchical clustering with the mere exception of drivers 31.2 and 1.2. These two drivers were assigned to the energy efficient group rather than the intermediate energy efficient group. Given the fact that the clusters are very similar for both clustering methods in Dataset 2, this research draws the assumption that Ward’s method and LPA produce similar clusters under the condition that there are sufficient samples in each Ward’s cluster.

(3) In addition, since the vehicle, the test route, and the instructions for the drives were in the experiment held constant when comparing to previous research, it was assumed that both datasets are similar. This research provides evidence that both datasets, Dataset 1 (generated in previous research [6]) and Dataset 2 (generated in the scope of this research), are statistically similar based on their mean energy consumption per mile. Based on this fact Dataset 1(including 30 samples) and Dataset 2 (including 23 samples) are merged into one Combined Dataset with a total sample size of 53 to have a comprehensive dataset.

(4) The Combined Dataset is clustered based on their mean energy consumption per mile and standard deviation, using Ward’s method (used in previous research), as well as introducing LPA as a new clustering approach in this research. Clustering the Combined Dataset with both clustering methods revealed two insights: (1) The cases when Ward’s method and LPA produce similar clusters, and (2) groups of BEV drivers, generated using LPA, based on their energy consumption are easily relatable to the literature.

Clustering the Combined Dataset with Ward’s method confirms two and three significant different clusters for the Combined Dataset. For a 2-component model, LPA finds similar clusters to the hierarchical clustering except from driver 31.1 (which is assigned to the energy inefficient cluster in LPA). The 3-component model for LPA, however, finds two clusters for the main part of the population and one cluster of its own for driver 34.1. This provides evidence that driver 34.1 is an outlier. In addition, a 4-component LPA model resulted in three equally shaped clusters and a long elliptical shaped cluster for low mean energy consumption (including driver 34.1). For a 5-component LPA model the main part of the population is clustered in equally shaped clusters except for driver 34.1 who is assigned a group on its own. Both models suggest that 34.1 should be excluded.

Given the previous results, driver 34.1 is excluded from the Combined Dataset. Clustering the reduced Combined Dataset (n=52) with Ward's method results in the same three clusters as for the non-reduced Combined Dataset just without driver 34.1. Applying a 2-component LPA and 3-component LPA model reveals significantly different clusters. This might be due to the fact that additional parameters apart from mean energy consumption per mile and standard deviation are relevant for understanding energy consumption based on driving behavior that are not respected here. Interesting is that the clustering pattern seen for the 3-component model is similar to the one for Dataset 1, which provides evidence that these clusters are relevant. It might be possible that the reasons for this pattern become clearer by including additional information into the clustering (e.g., adding a z-axis to the graph with acceleration or speed). A 4-component LPA model and 5-component LPA model find clusters resulting in equally shaped clusters. Drivers are grouped in accordance to what is known in literature and provide evidence that speed and acceleration are main factor for describing driving behavior and energy consumption of BEVs.

In the scope of this research the 4-component LPA model is promoted as the most suitable one for clustering the comprehensive dataset. There is a cluster of high mean and high standard deviation which represents the group of high energy consumption. There are two clusters of high and low mean at medium standard deviation which represent the intermediate energy consuming group. Then there is a group of two individuals with low mean and low standard deviation representing the energy efficient group. There is a significant difference in energy consumption between drivers based on their driving style. In the context of eco-driving, for ICE vehicles, education methods have been discussed to increase eco-driving behavior throughout ICE vehicle drivers. Based on this research also for BEV drivers educating energy efficient driving seems to have considerable benefits. In terms of understanding how BEV drivers would need to drive in order to be the most energy efficient, the individuals in the energy efficient cluster can be used as a reference since they consumed considerable less energy than the rest of the test drivers.

In summary, this research provides a better understanding of how BEV drivers need to be clustered based on their mean energy consumption per mile and standard deviation. However, even though acceleration and speed data (key indicators for describing the vehicle motion and thus driving behavior) are collected in the scope of this research they are not included in the clustering analysis. Including these parameters into the analysis would allow a clearer understanding of the characteristics of driver groups and based on these driving characteristics derive recommendations for educating BEV drivers for energy efficient driving.

5.1 Limitations and Further Research

This research intended to provide a more detailed analysis of energy consumption based on the concept of eco-driving, commonly used for ICE vehicles. This is realized by collecting energy consumption data at a resolution of one second. However, due to the complications in the data collection process for Dataset 2, only approximately 2/3 of the collected data was usable. In addition, even though the data from Dataset 1 was already pre-prepared, some important information was missing that needed to be added in post-processing and the data needed to be aggregated for further analysis in this research. This is why this research focused on an advanced clustering of drivers based on mean energy consumption per mile and standard deviation rather than an analysis of the parameters speed and acceleration on energy consumption.

With respect to the energy consumption clusters found in this research, it is assumed that speed and acceleration are main factors for describing driving behavior. However, to explain which driving parameters are responsible for the energy consumption pattern seen in this research, acceleration and speed data would need to be connected to this analysis.

Previous research has used the concept of an SDR within a friction circle to determine aggressive driving, which is considered to be a main factor for high energy consumption for ICE vehicles and EVs alike. The same concept could be used for further in-depth analysis of speed which is the other important parameter for determining aggressive driving as defined for eco-driving. In this context, a speed analysis could be implemented by determining a threshold for every speed limit along a road segment. For these road segments the speed measuring point inside the threshold and outside the threshold could be compared and would give a result for speed similar to the approach presented in previous research for acceleration. The advantage of this approach, in comparing mean speeds, is that only aggressive speeding above the threshold would contribute to the aggressive driving analysis, leaving out varying speeds within a boundary of safe driving caused by traffic concentration or road conditions.

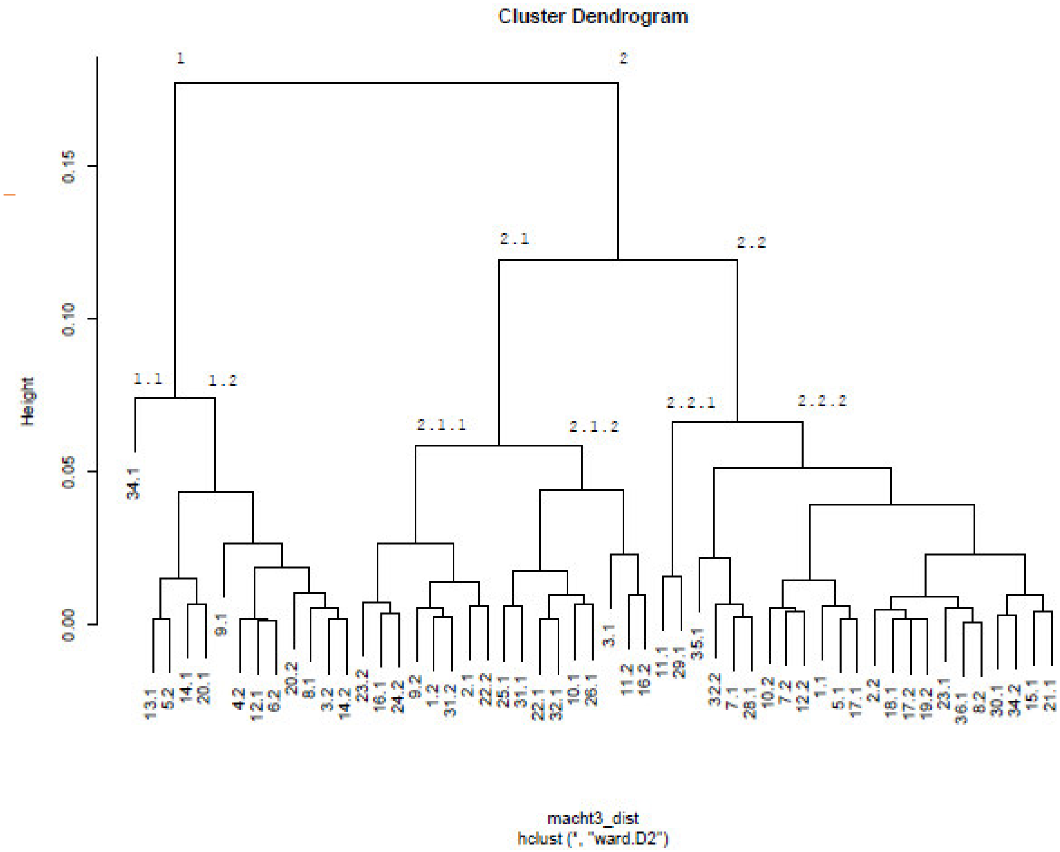
Relating these two measures to the energy consumption clusters would deliver a clearer understanding of which driving behavior would provoke what kind of energy consumption pattern for driving BEVs.

This research focused on clustering drivers based on their energy consumption over the entire test route. To reduce variability and generate more robust results braking down the test route into segments for each road type would be beneficial. This would reveal for which road type there might be the largest difference in energy consumption and where adjusting the driving style would generate the most impact in terms of energy savings.

Furthermore, in the scope of this research the traffic concentration was controlled indirectly by making sure the test route progressed through an area of low to medium traffic generation and by conducting test drives between 10am and 5pm to avoid commute rush hours. This experiment setup resulted in minor stops due to high traffic concentration during the test drives. For the Combined Dataset, there was no data available in the upper left corner of low mean energy consumption and high standard deviation, which might have been due to the experimental design. To test whether this area might be theoretically feasible, future research is needed to conduct test drives in a high traffic density area outside the time interval of low traffic concentration between 10am and 5pm. Due to low speeds with abrupt start and stop movements (e.g., in traffic jams), low means with high standard deviations could be possible. This approach would generate data in an area where there is no data available yet and would give further insights on driving behavior influences energy consumption in high traffic concentration conditions.

APPENDICES

Latent Profile Analysis for Combined Dataset

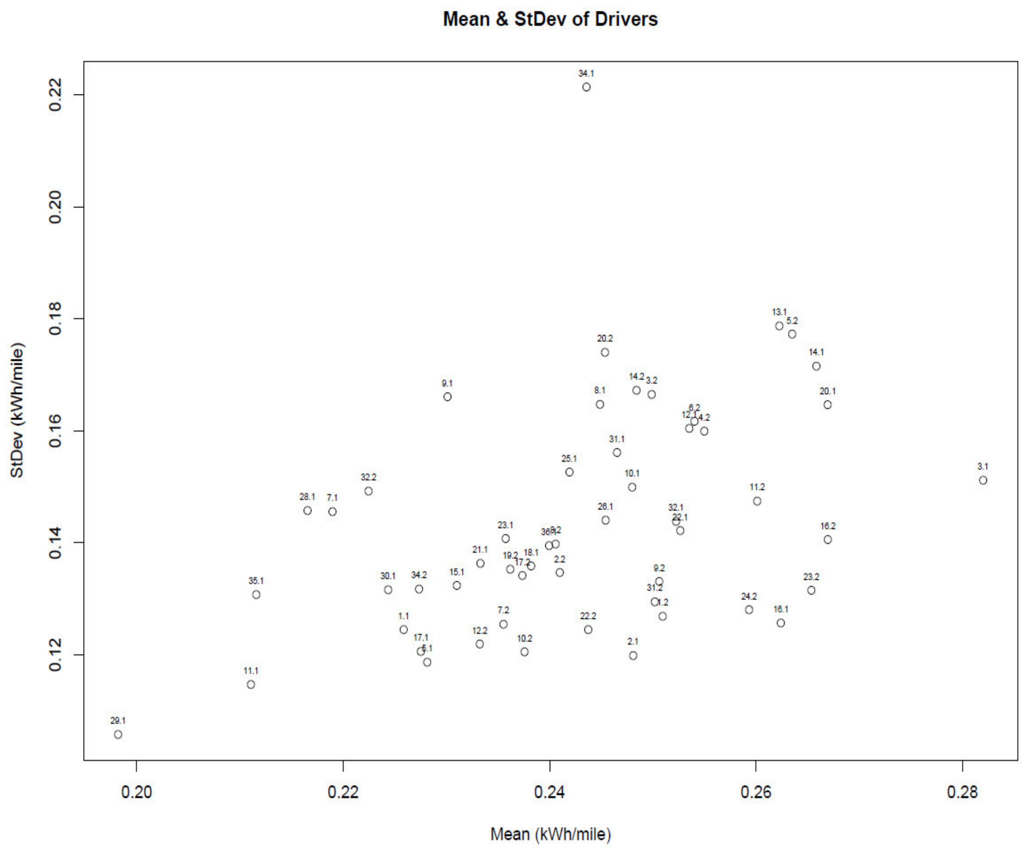


Appendix 1: Dendrogram Combined Dataset

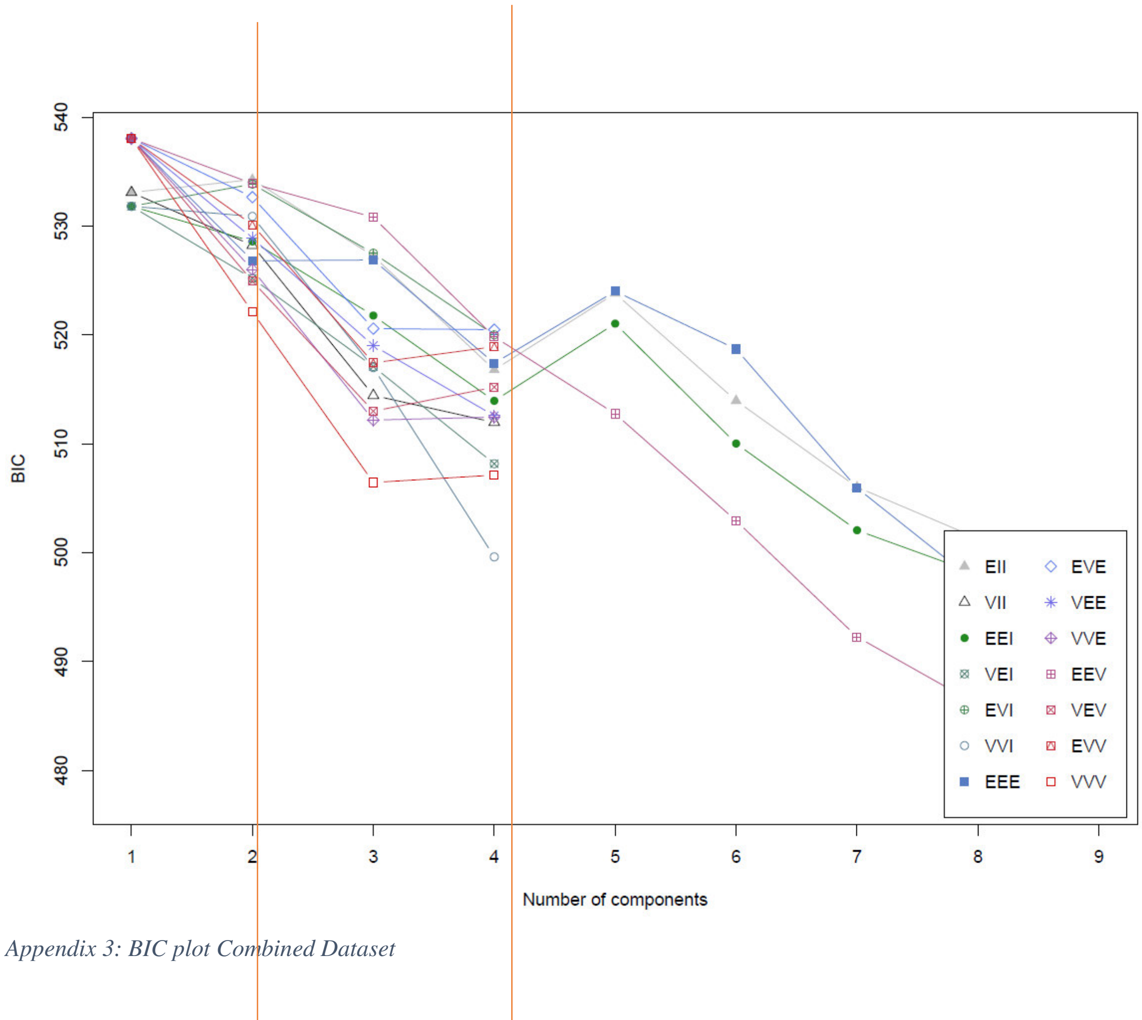
Based on the Dendrogram a reasonable number of cluster seems to be either:

- 2 cluster, -> 1/2 different
- 3 clusters, -> 1/2.1 different, 1/2.2 different, 2.1/2.2 different
- 4 clusters, -> 1.1/1.2 same, 1.1/2.1 same, 1.1/2.2 same, 1.2/2.1 different, 1.2/2.2 different, 2.1/2.2 different

Different is defined as clusters being statistically different based on Wilcox test
Same is defined as clusters being statistically different based on Wilcox test



Appendix 2: Scatter plot Combined Dataset



Appendix 3: BIC plot Combined Dataset

Based on the Latend Profile output:

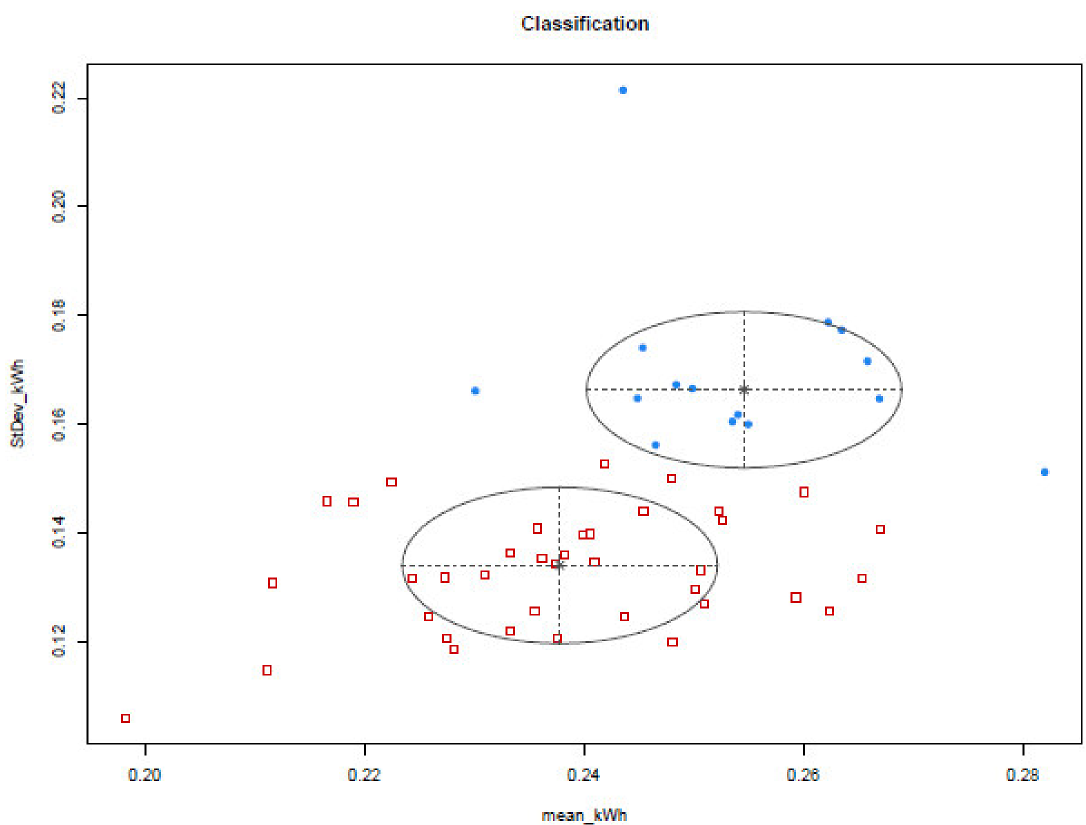
- 2 clusters looks good
- 4 clusters weirdly shaped

```
> LP_2 <- Mclust(macht3, 2)
fitting ...
  |========================================================================
========================================| 100%
> summary(LP_2, parameters=TRUE)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EII (spherical, equal volume) model with 2 components:

 log.likelihood  n df      BIC      ICL
       279.0472 53  6 534.2727 523.2111

Clustering table:
 1  2
38 15
```

Mixing probabilities:
        1   2
0.7104175 0.2895825

Means:
              [,1]   [,2]
mean_kWh 0.2378054 0.2545888
StDev_kWh 0.1340071 0.1662910

Variances:
[,,1]
             mean_kWh   StDev_kWh
mean_kWh 0.0002062574 0.0000000000
StDev_kWh 0.0000000000 0.0002062574
[,,2]
             mean_kWh   StDev_kWh
mean_kWh 0.0002062574 0.0000000000
StDev_kWh 0.0000000000 0.0002062574



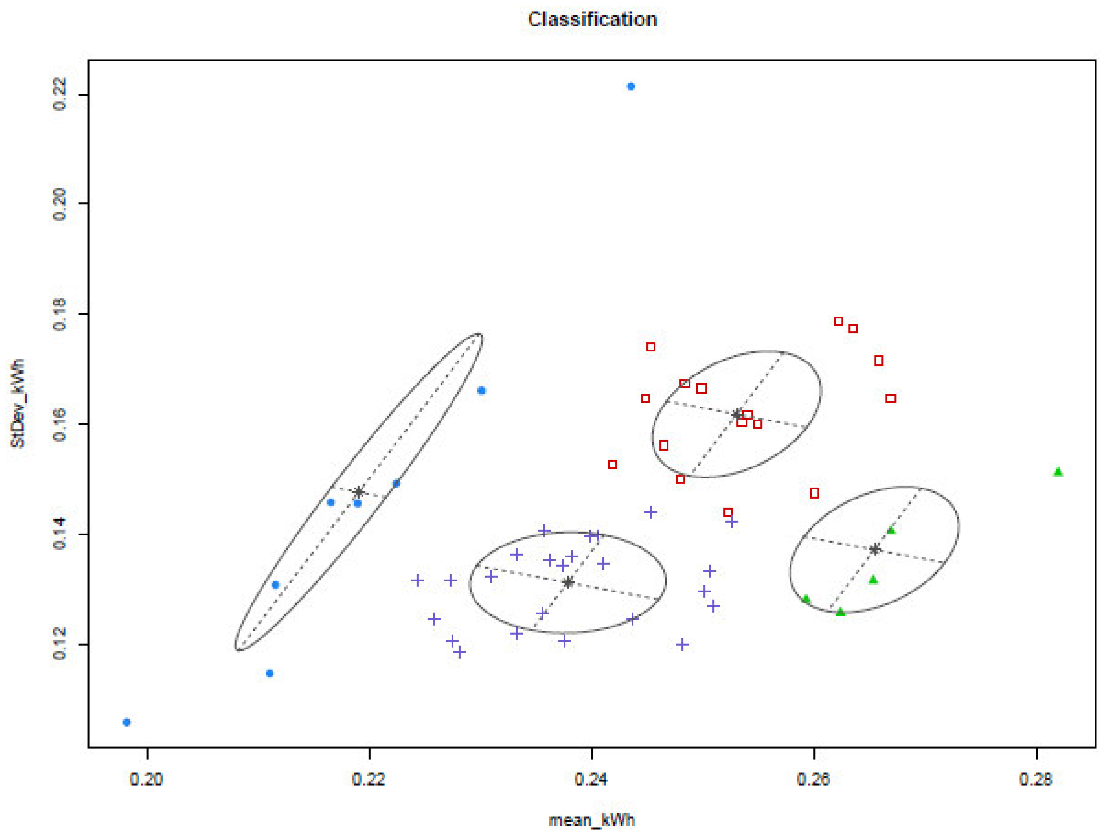
Appendix 4: LPA 2-component model Combined Dataset

> LP_4 <- Mclust(macht3, 4)
fitting ...
  |==
===| 100%
> summary(LP_4, parameters=TRUE)

Gaussian finite mixture model fitted by EM algorithm

Mclust EVE (ellipsoidal, equal volume and orientation) model with 4 components:

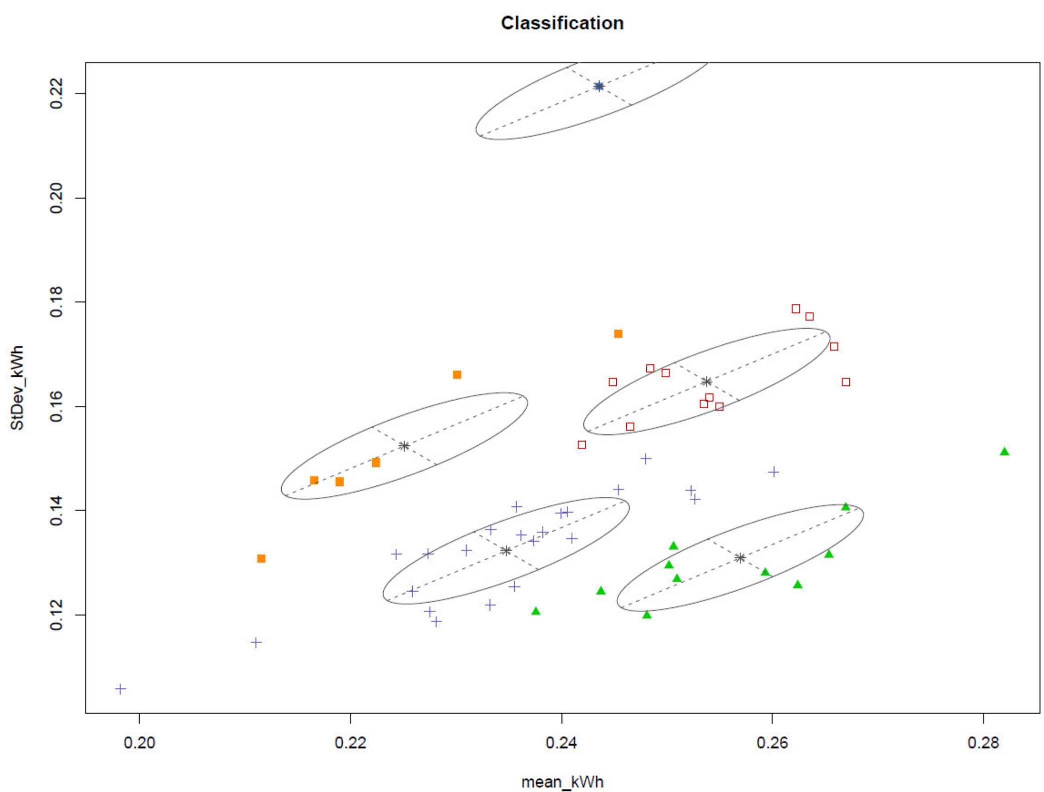
```
 log.likelihood  n df     BIC      ICL
       293.9905 53 17 520.486 510.9118

Clustering table:
 1  2  3  4 
24  5  8 16 

Mixing probabilities:
        1         2         3         4 
0.4494706 0.1075613 0.1441479 0.2988202 

Means:
               [,1]      [,2]      [,3]      [,4]
mean_kWh  0.2378713 0.2654454 0.2190753 0.2530571
StDev_kWh 0.1311981 0.1371618 0.1476446 0.1618040

Variances:
[,,1]
              mean_kWh    StDev_kWh
mean_kWh  7.757598e-05 2.900464e-06
StDev_kWh 2.900464e-06 8.419488e-05
[,,2]
              mean_kWh    StDev_kWh
mean_kWh  5.771901e-05 3.201044e-05
StDev_kWh 3.201044e-05 1.307673e-04
[,,3]
              mean_kWh    StDev_kWh
mean_kWh  0.0001236294 0.0003103536
StDev_kWh 0.0003103536 0.0008318604
[,,4]
              mean_kWh    StDev_kWh
mean_kWh  5.760717e-05 3.232429e-05
StDev_kWh 3.232429e-05 1.313716e-04
```



Appendix 5: LPA 4-component model Combined Dataset

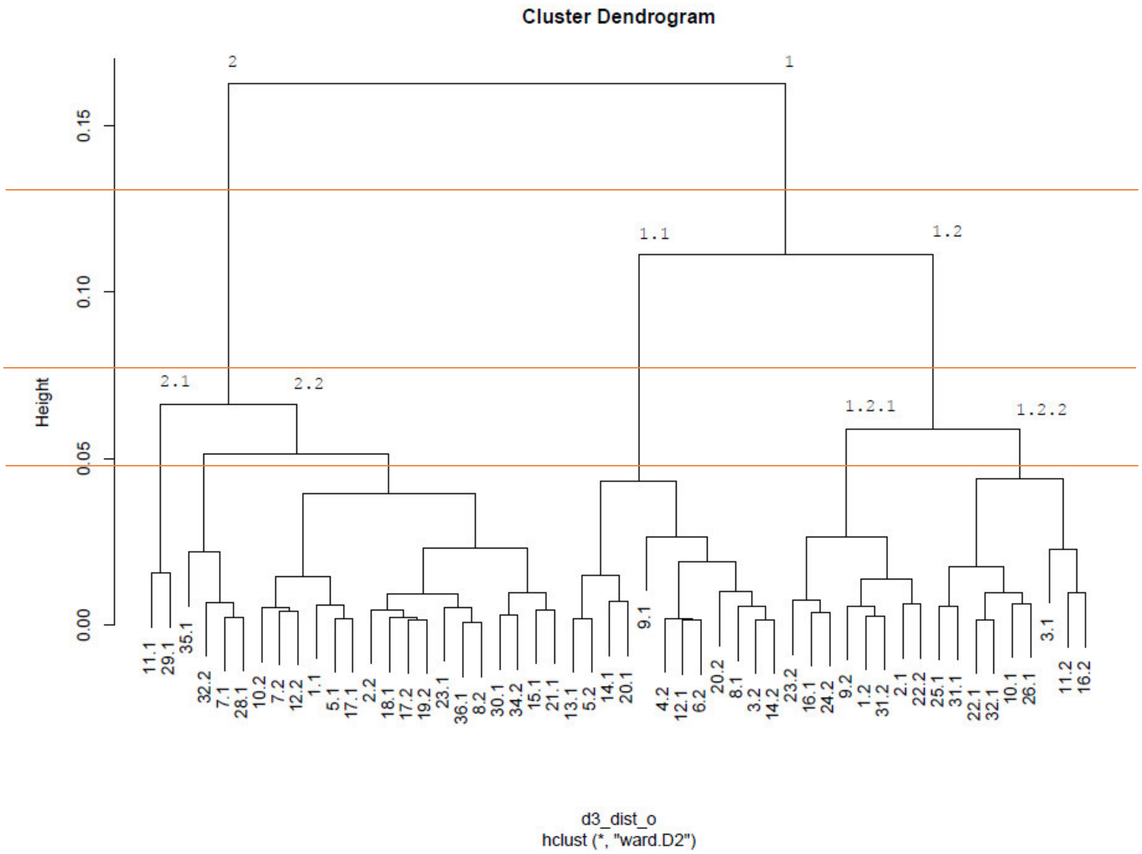
```
> LP_5 <- Mclust(macht3, 5)
fitting ...
  |===========================================================================
=========================================| 100%
> summary(LP_5, parameters=TRUE) # 34.1 a group on its own
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EEE (ellipsoidal, equal volume, shape and orientation) model with 5 components:

 log.likelihood  n df      BIC      ICL
        295.756 53 17 524.017 515.7674

Clustering table:
 1  2  3  4  5
 1 12 11 23  6

Mixing probabilities:
         1          2          3          4          5
0.01886792 0.23027724 0.19809647 0.44624895 0.10650942

Means:
              [,1]      [,2]      [,3]      [,4]      [,5]
```

```
mean_kWh  0.2435669 0.2537720 0.2569445 0.2347485 0.2251071
StDev_kWh 0.2214319 0.1647889 0.1309563 0.1323313 0.1524388

Variances:
[,,1]
            mean_kWh     StDev_kWh
mean_kWh  1.365545e-04 9.661497e-05
StDev_kWh 9.661497e-05 1.044259e-04
[,,2]
            mean_kWh     StDev_kWh
mean_kWh  1.365545e-04 9.661497e-05
StDev_kWh 9.661497e-05 1.044259e-04
[,,3]
            mean_kWh     StDev_kWh
mean_kWh  1.365545e-04 9.661497e-05
StDev_kWh 9.661497e-05 1.044259e-04
[,,4]
            mean_kWh     StDev_kWh
mean_kWh  1.365545e-04 9.661497e-05
StDev_kWh 9.661497e-05 1.044259e-04
[,,5]
            mean_kWh     StDev_kWh
mean_kWh  1.365545e-04 9.661497e-05
StDev_kWh 9.661497e-05 1.044259e-04
```



Appendix 6: LPA 5-component model Combined Dataset

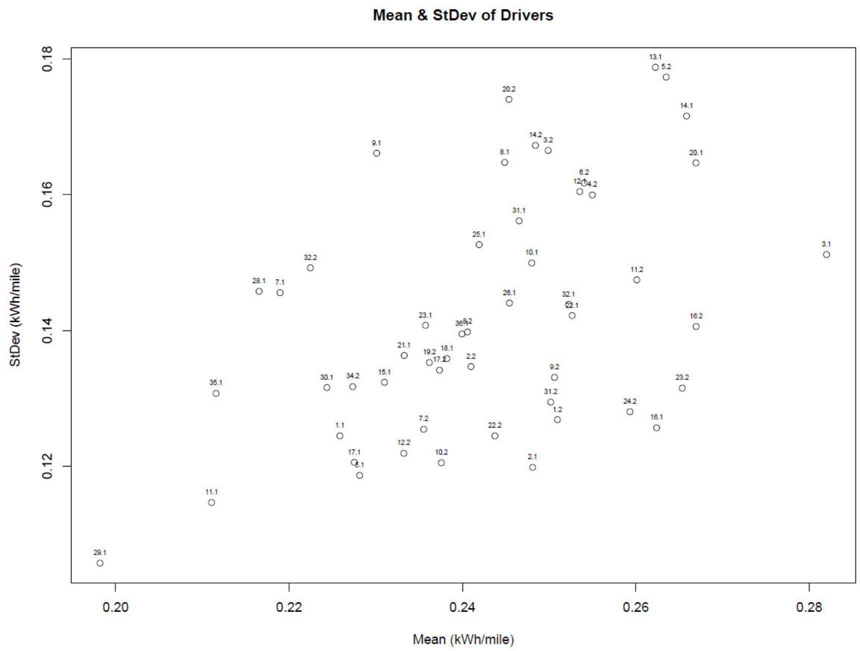
Latent Profile Analysis without 34.1 for Combined Dataset



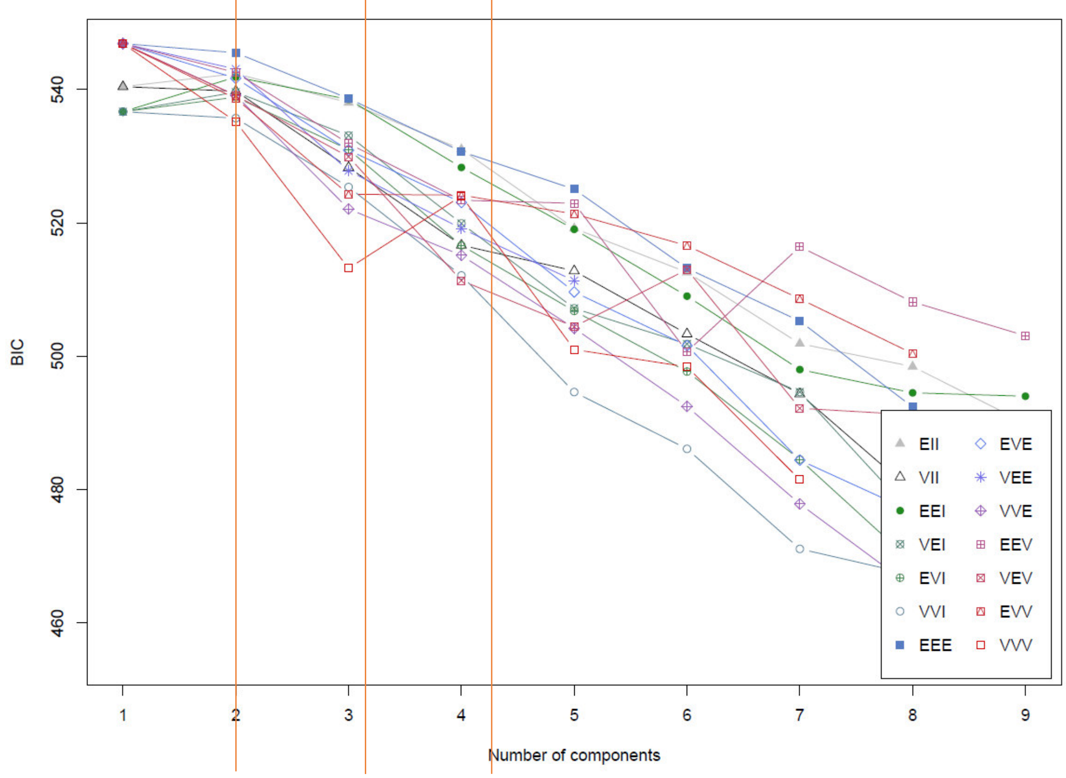
Appendix 7: Dendrogram Combined Dataset without 34.1

Based on Dendrogram

- 2 cluster -> 1/2 different
- 3 clusters, -> 1.1/1.2 different, 1.1/2 different, 1.2/2 different
- 4 clusters, -> 1.1/1.2 different, 1.1/2.1 different, 1.1/2.2 same, 1.1/2.2 different, 1.2/2.1 different, 1.2/2.2 different, 2.1/2.2 different



Appendix 8: Scatter plot Combined Dataset without 34.1

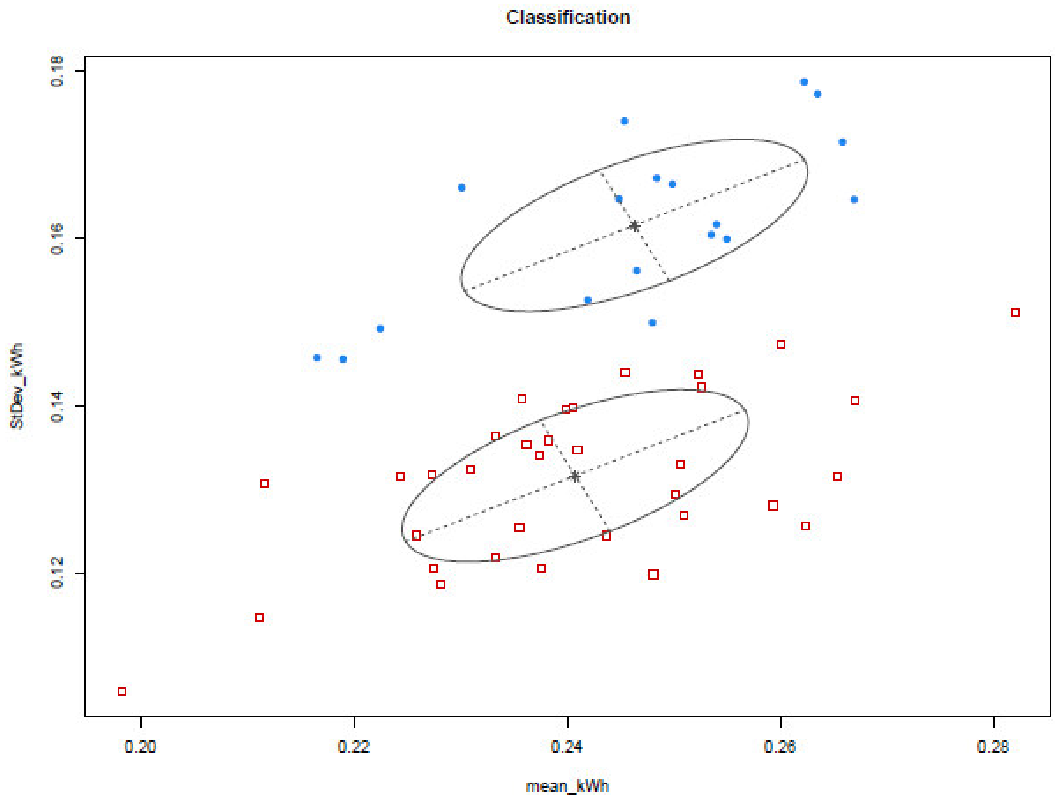


Appendix 9: BIC plot Combined Dataset without 34.1

Based on Latent Profile Analysis

- 2 cluster, possible
- 3 clusters, look good, although highes cluster convers almost entire range of means
- 4 clusters, looks good

```
> LP_2_o <- Mclust(d3_o, 2)
fitting ...
  |==========================================================================
========================================| 100%
> summary(LP_2_o, parameters=TRUE)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EEE (ellipsoidal, equal volume, shape and orientation) model with 2 components:

 log.likelihood  n df     BIC      ICL
       288.557 52  8 545.504 542.0696

Clustering table:
 1  2
34 18

Mixing probabilities:
        1         2
0.6593286 0.3406714

Means:
               [,1]      [,2]
mean_kWh  0.2407567 0.2463093
StDev_kWh 0.1316544 0.1615956

Variances:
[,,1]
             mean_kWh    StDev_kWh
mean_kWh  0.0002643498 0.0001036132
StDev_kWh 0.0001036132 0.0001059172
[,,2]
             mean_kWh    StDev_kWh
mean_kWh  0.0002643498 0.0001036132
StDev_kWh 0.0001036132 0.0001059172
```

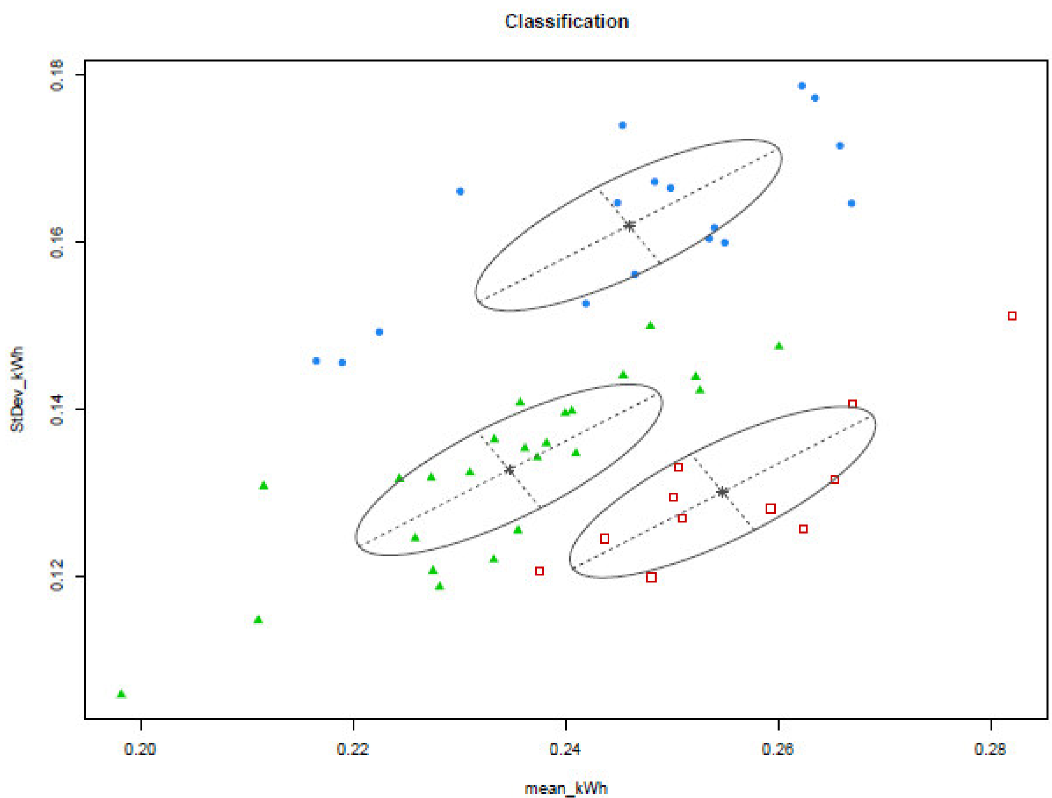


Appendix 10: LPA 2-component model Combined Dataset without 34.1

```
> LP_3_o <- Mclust(d3_o, 3)
fitting ...
  |=========================================================================
========================================| 100%
> summary(LP_3_o, parameters=TRUE)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EEE (ellipsoidal, equal volume, shape and orientation) model with 3 components:

 log.likelihood  n df      BIC      ICL
       291.0495 52 11 538.6352 530.7024

Clustering table:
 1  2  3
24 11 17

Mixing probabilities:
        1         2         3
0.4584138 0.2109577 0.3306285

Means:
               [,1]      [,2]      [,3]
mean_kWh  0.2346870 0.2547832 0.2459438
```

StDev_kWh 0.1327247 0.1300744 0.1620292

Variances:
[,,1]
            mean_kWh   StDev_kWh
mean_kWh 0.0002075641 0.0001158101
StDev_kWh 0.0001158101 0.0001052216
[,,2]
            mean_kWh   StDev_kWh
mean_kWh 0.0002075641 0.0001158101
StDev_kWh 0.0001158101 0.0001052216
[,,3]
            mean_kWh   StDev_kWh
mean_kWh 0.0002075641 0.0001158101
StDev_kWh 0.0001158101 0.0001052216



Appendix 11: LPA 3-component model Combined Dataset without 34.1

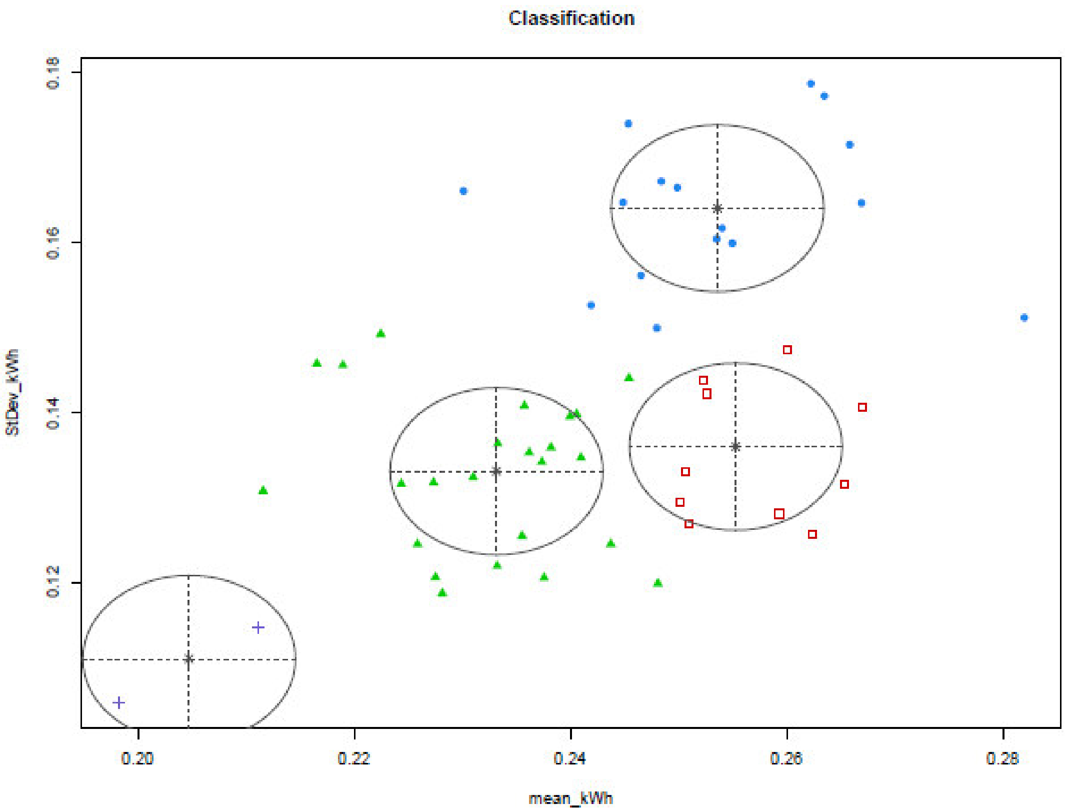
> LP_4_o <- Mclust(d3_o, 4)
fitting ...
  |==
===| 100%
> summary(LP_4_o, parameters=TRUE)

Gaussian finite mixture model fitted by EM algorithm

Mclust EII (spherical, equal volume) model with 4 components:

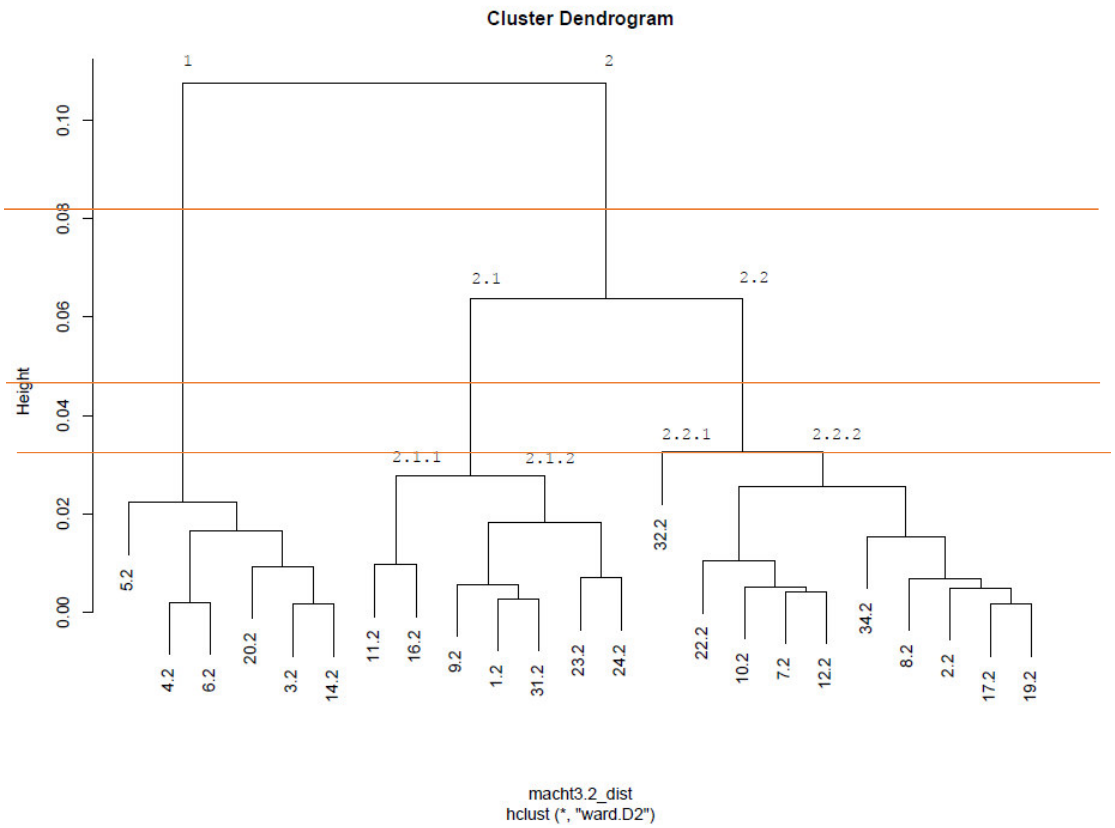
 log.likelihood n df   BIC   ICL

```
    289.2242 52 12 531.0335 514.3711

Clustering table:
 1  2  3  4 
24 10  2 16 

Mixing probabilities:
         1          2          3          4 
0.46446940 0.20855692 0.03728042 0.28969326 

Means:
              [,1]      [,2]      [,3]      [,4]
mean_kWh  0.2331737 0.2552956 0.2047409 0.2536122
StDev_kWh 0.1330998 0.1360042 0.1109926 0.1640743

Variances:
[,,1]
             mean_kWh    StDev_kWh
mean_kWh  9.691648e-05 0.000000e+00
StDev_kWh 0.000000e+00 9.691648e-05
[,,2]
             mean_kWh    StDev_kWh
mean_kWh  9.691648e-05 0.000000e+00
StDev_kWh 0.000000e+00 9.691648e-05
[,,3]
             mean_kWh    StDev_kWh
mean_kWh  9.691648e-05 0.000000e+00
StDev_kWh 0.000000e+00 9.691648e-05
[,,4]
             mean_kWh    StDev_kWh
mean_kWh  9.691648e-05 0.000000e+00
StDev_kWh 0.000000e+00 9.691648e-05
```



Appendix 12: LPA 4-component model Combined Dataset without 34.1

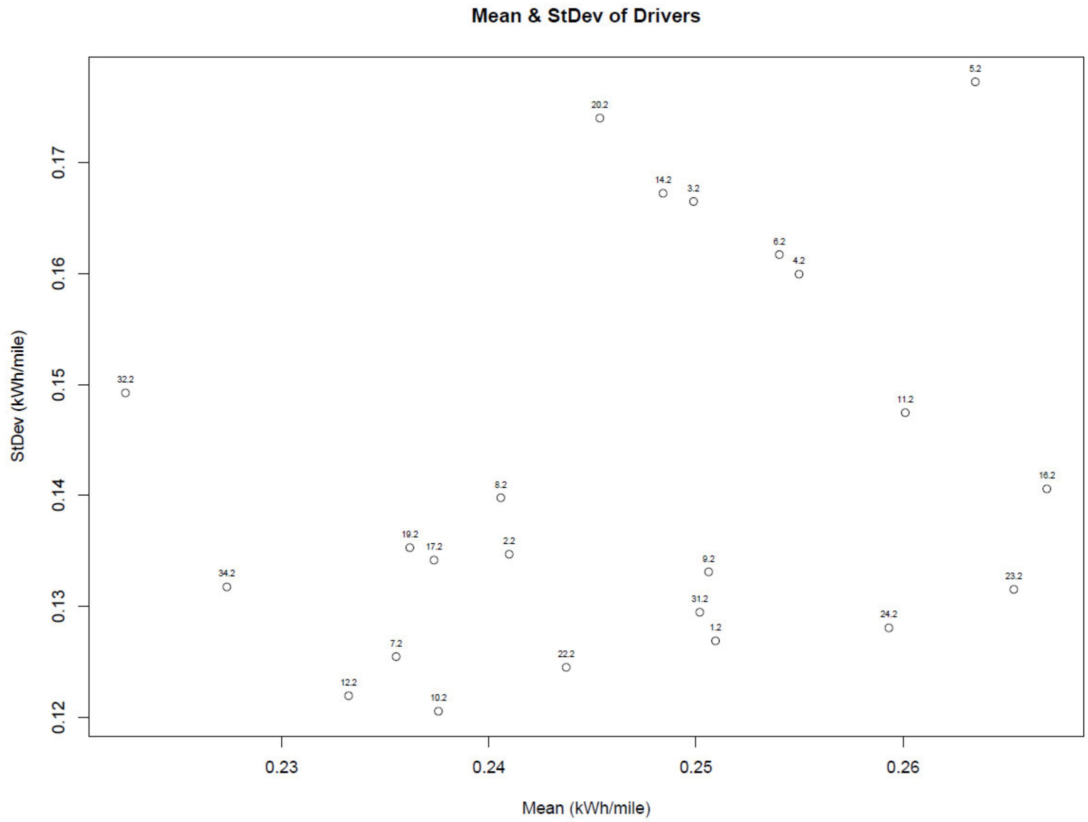
Latent Profile Analysis for Dataset 2



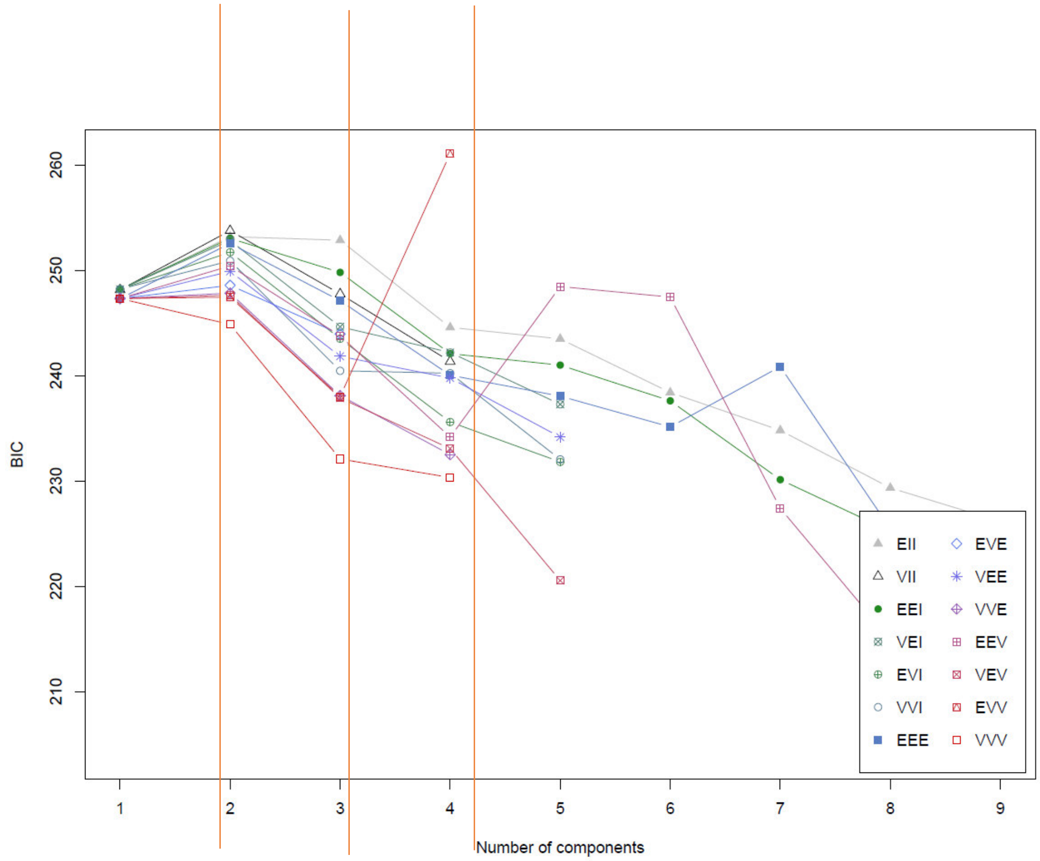
Appendix 13: Dendrogram Dataset 2

Based on Dendrogramm

- 2 cluster -> 1/2 different
- 3 clusters -> 1/2.1 different, 1/2.2 different, 2.1/2.2 different
- 4 clusters -> 1/2.1 different, 1/2.2.1 same, 1/2.2.2 different, 2.1/2.2.1 same, 2.1/2.2.2 different, 2.2.1/2.2.2 same



Appendix 14: Scatter plot Dataset 2



Appendix 15: BIC plot Dataset 2

Latent Profile Analysis

- 2 clusters, looks good
- 3 clusters, looks good, 1 low mean low StDev cluster, and 2 high and low mean and high and l ow StDev clusters, evenly low/low cluster around double the size of the other clusters
- 4 clusters, doesn´t look very convincing

```
> LP_2.2 <- Mclust(macht3.2, 2)
fitting ...
  |==========================================================================
=========================================| 100%
> summary(LP_2.2, parameters=TRUE)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust VII (spherical, varying volume) model with 2 components:

 log.likelihood  n df      BIC      ICL
       137.8516 23  7 253.7548 253.5422

Clustering table:
 1  2
17  6

Mixing probabilities:
        1         2
```

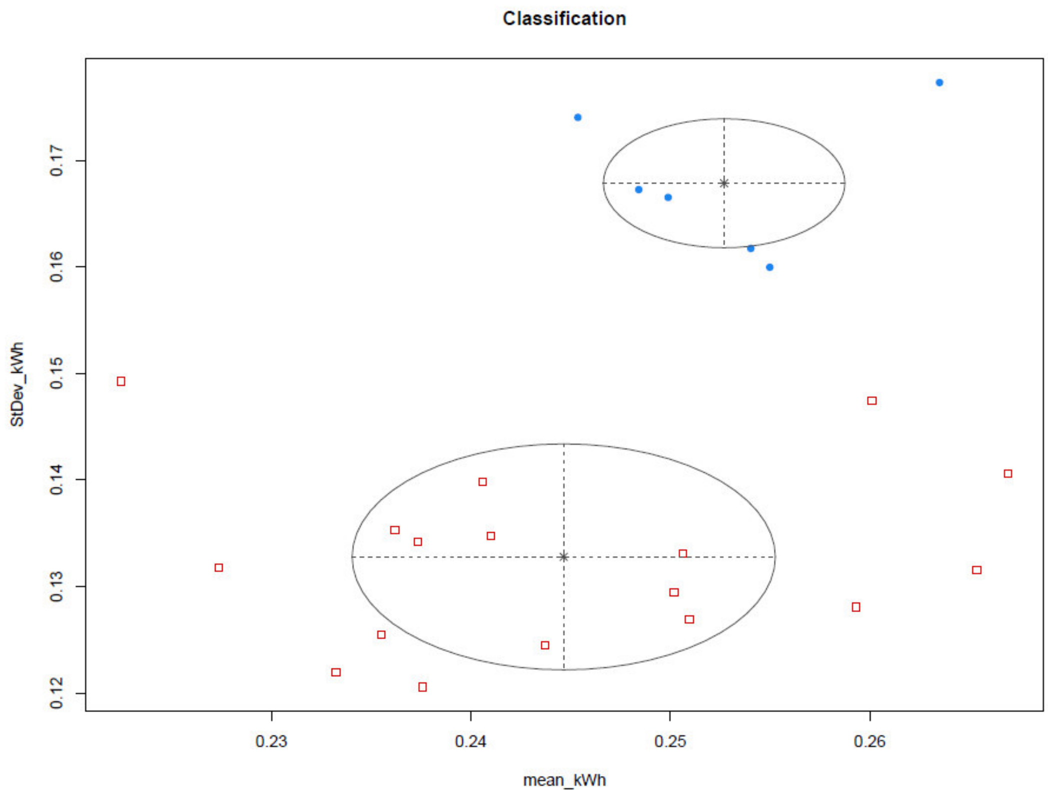
0.7424812 0.2575188

Means:

```
           [,1]      [,2]
mean_kWh  0.2446488 0.2526898
StDev_kWh 0.1327696 0.1678454
```

Variances:

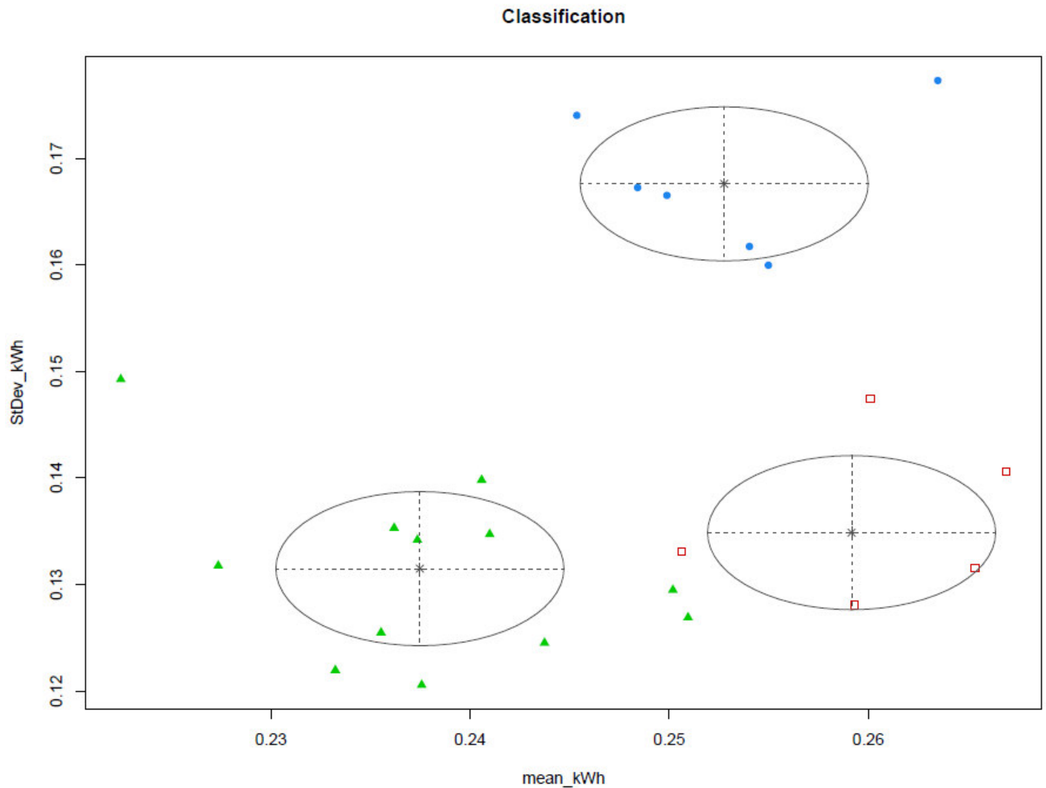
```
[,,1]
            mean_kWh    StDev_kWh
mean_kWh  0.0001125321 0.0000000000
StDev_kWh 0.0000000000 0.0001125321
[,,2]
            mean_kWh    StDev_kWh
mean_kWh  3.666054e-05 0.000000e+00
StDev_kWh 0.000000e+00 3.666054e-05
```



Appendix 16: LPA 2-component model Dataset 2

```
> LP_3.2 <- Mclust(macht3.2, 3)
fitting ...
  |=============================================================================
==========================================| 100%
> summary(LP_3.2, parameters=TRUE)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EII (spherical, equal volume) model with 3 components:
```

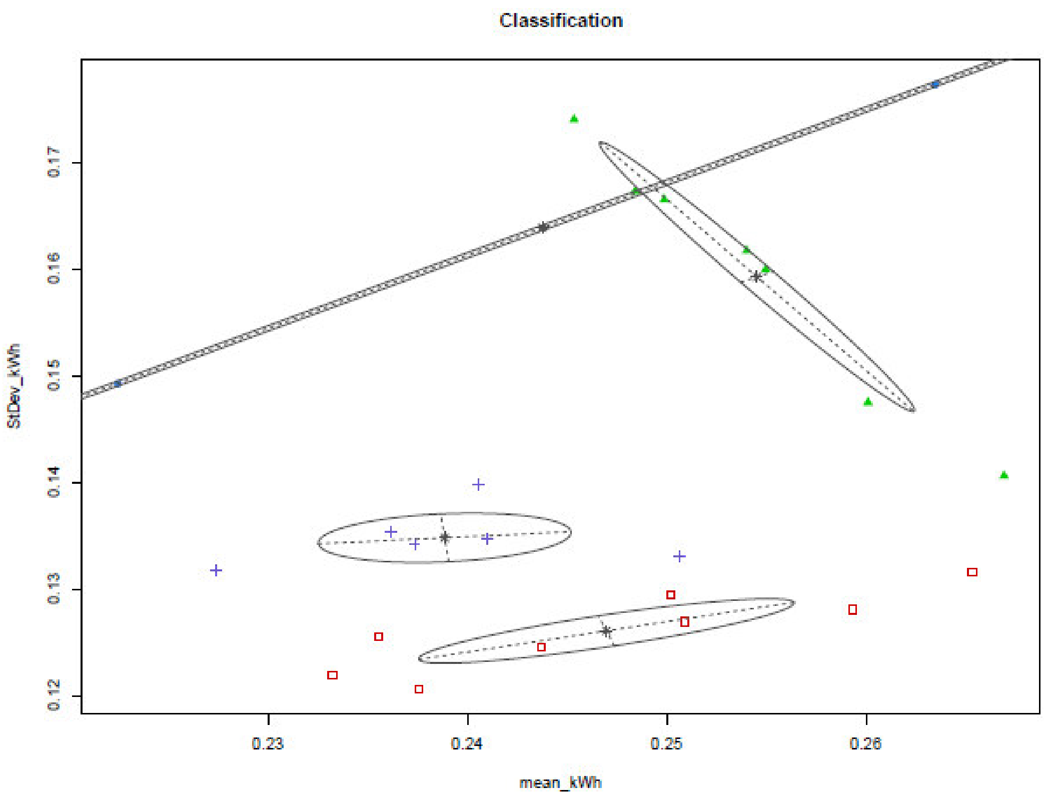
```
 log.likelihood  n df     BIC      ICL
       140.5317 23  9 252.844 248.6559

Clustering table:
 1  2  3 
 5 12  6 

Mixing probabilities:
        1         2         3 
0.2407532 0.4959752 0.2632716 

Means:
               [,1]      [,2]      [,3]
mean_kWh  0.2591571 0.2374764 0.2527587
StDev_kWh 0.1348584 0.1314781 0.1676018

Variances:
[,,1]
             mean_kWh    StDev_kWh
mean_kWh  5.232965e-05 0.000000e+00
StDev_kWh 0.000000e+00 5.232965e-05
[,,2]
             mean_kWh    StDev_kWh
mean_kWh  5.232965e-05 0.000000e+00
StDev_kWh 0.000000e+00 5.232965e-05
[,,3]
             mean_kWh    StDev_kWh
mean_kWh  5.232965e-05 0.000000e+00
StDev_kWh 0.000000e+00 5.232965e-05
```



Appendix 17: LPA 3-component model Dataset 2

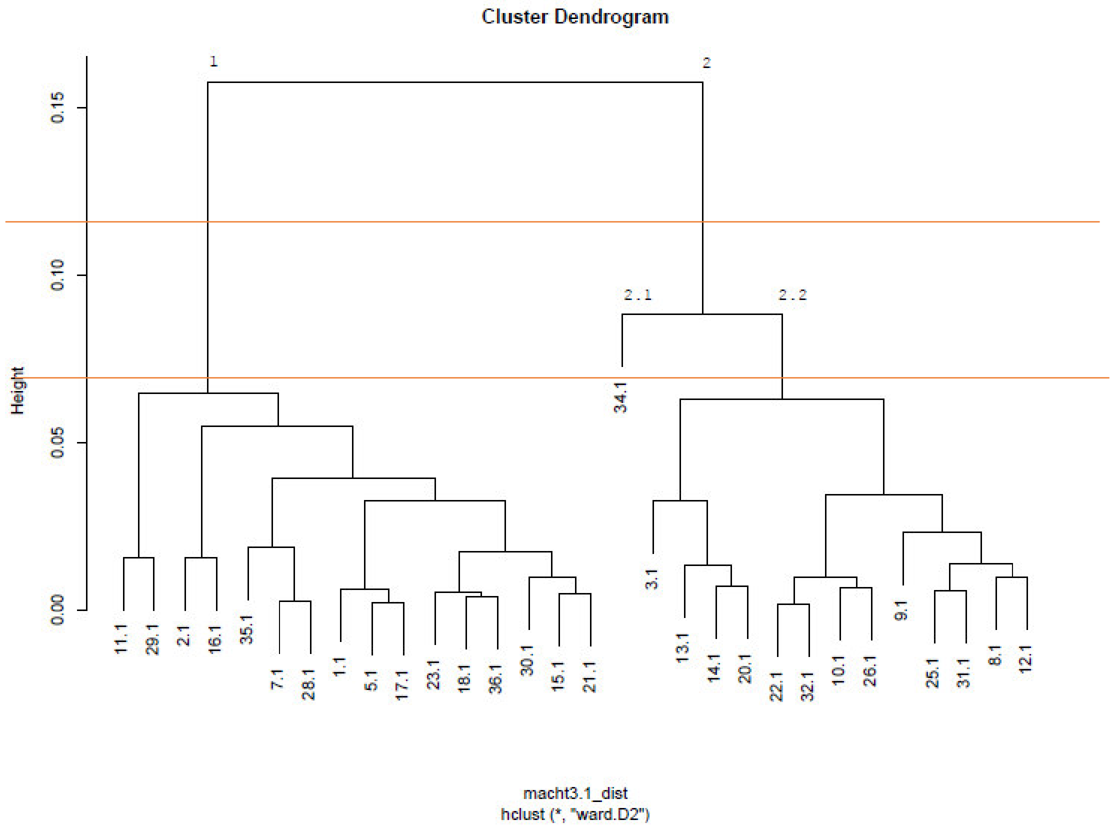
```
> LP_4.2 <- Mclust(macht3.2, 4)
fitting ...
  |===========================================================================
=========================================| 100%
> summary(LP_4.2, parameters=TRUE)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EVV (ellipsoidal, equal volume) model with 4 components:

 log.likelihood  n df      BIC      ICL
       161.8902 23 20 261.0704 260.1904

Clustering table:
1 2 3 4
8 6 2 7

Mixing probabilities:
        1         2         3         4
0.3473287 0.2613694 0.1018124 0.2894895

Means:
               [,1]      [,2]      [,3]      [,4]
mean_kWh  0.2469795 0.2388604 0.2437652 0.2545422
StDev_kWh 0.1260571 0.1347978 0.1638560 0.1592611
```

```
Variances:
[,,1]
             [,1]         [,2]
[1,] 8.886808e-05 2.470083e-05
[2,] 2.470083e-05 9.177887e-06
[,,2]
             [,1]         [,2]
[1,] 4.030396e-05 3.144028e-06
[2,] 3.144028e-06 5.343776e-06
[,,3]
            [,1]        [,2]
[1,] 0.003010840 0.002059617
[2,] 0.002059617 0.001408985
[,,4]
              [,1]          [,2]
[1,]  6.272763e-05 -9.920041e-05
[2,] -9.920041e-05  1.601561e-04
```



Appendix 18: LPA 4-component model Dataset 2

Latent Profile Analysis for Dataset 1 (Dans Dataset)

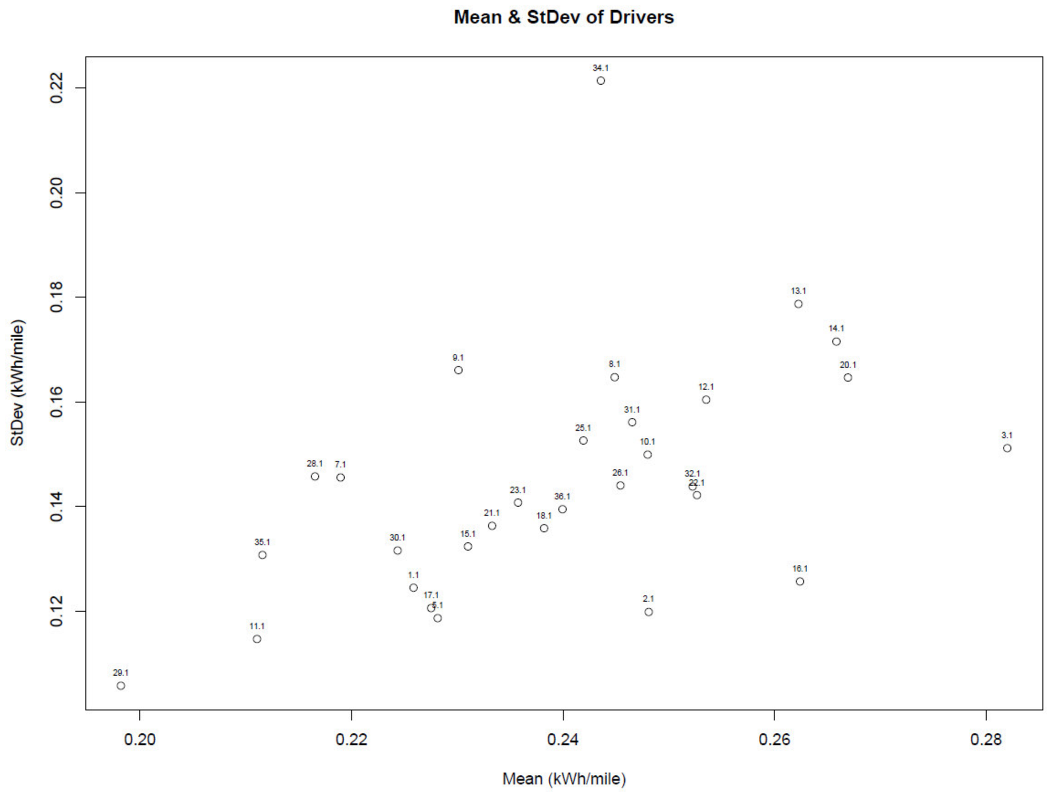


Appendix 19: Dendrogram Dataset 1

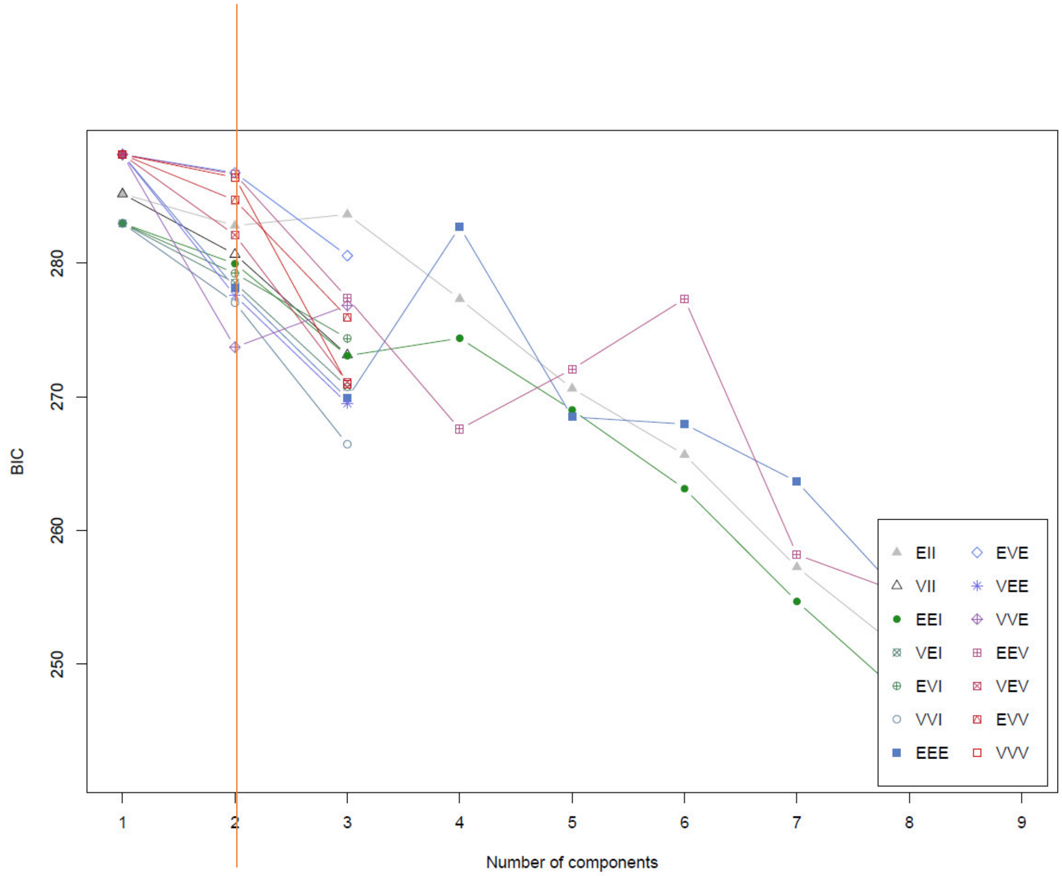
Based on Dendrogramm:

- 2 clusters -> 1/2 different
- 3 clusters -> 1/2.1 same, 1/2.2 different, 2.1/2.2 same

  Dans and my Dendrogram are the same with distinction of the 34 being one branch higher in my Dendrogram
  Also heights are different I guess



Appendix 20: Scatter plot Dataset 1



Appendix 21: BIC plot Dataset 1

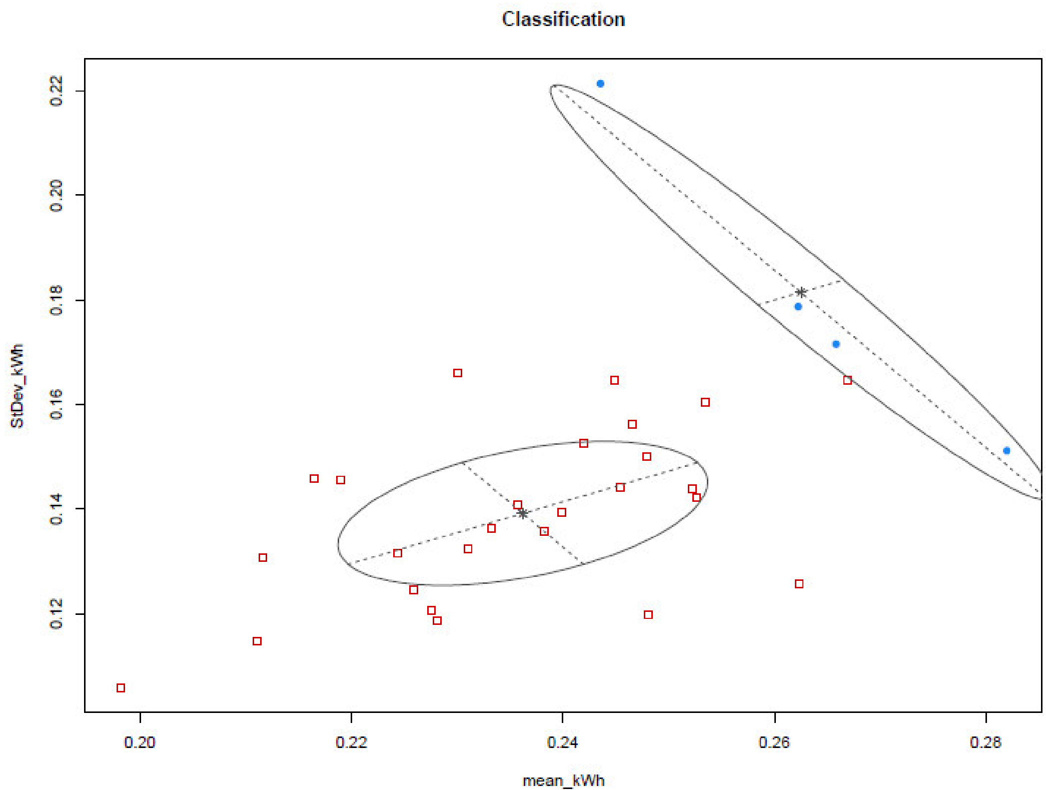
Based on Latent Profile Analysis

- 2 clusters do not look convincing, very few values in the second cluster (high mean, high StDev)

```
> LP_2.1 <- Mclust(macht3.1, 2)
fitting ...
  |==========================================================================
=========================================| 100%
> summary(LP_2.1, parameters=TRUE)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EVE (ellipsoidal, equal volume and orientation) model with 2 components:

 log.likelihood  n df      BIC      ICL
       158.6875 30  9 286.7642 284.2076

Clustering table:
 1  2
26  4

Mixing probabilities:
        1         2
0.8727979 0.1272021
```

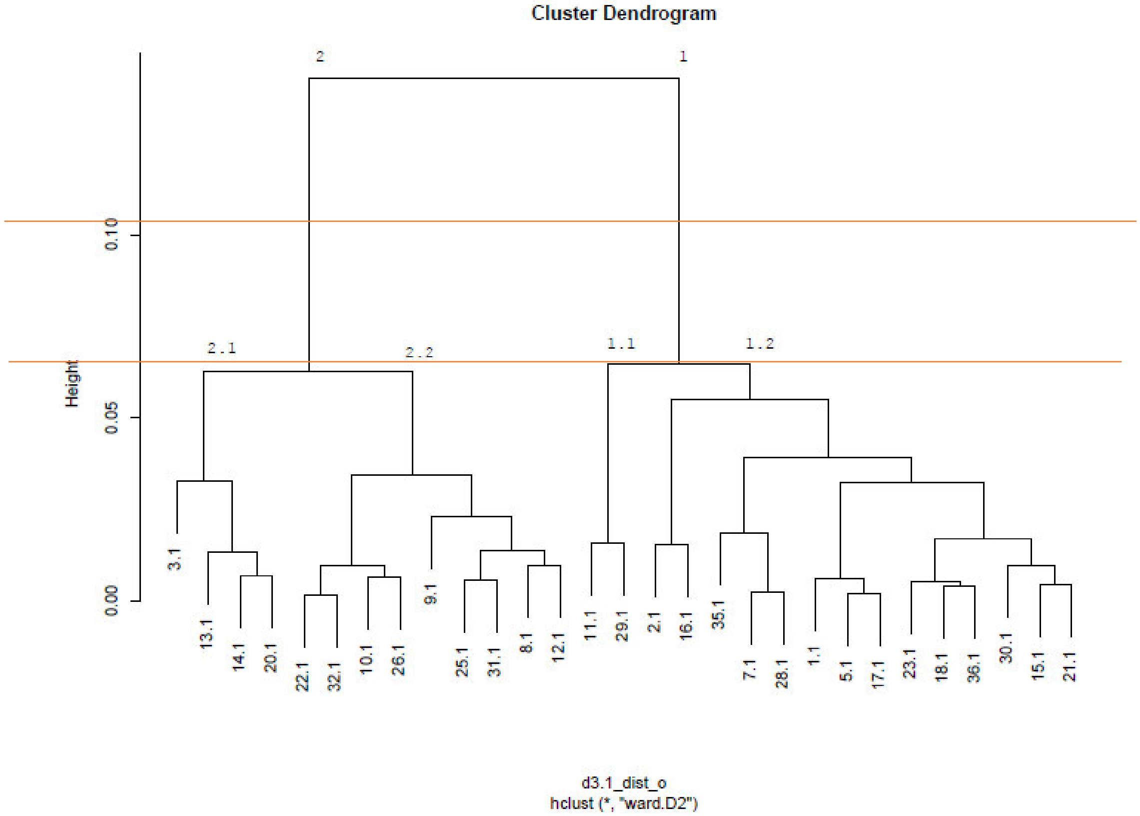
Means:
[,1] [,2]
mean_kWh 0.2362131 0.2625063
StDev_kWh 0.1391670 0.1814629

Variances:
[,,1]
mean_kWh StDev_kWh
mean_kWh 0.0003060054 0.0001049814
StDev_kWh 0.0001049814 0.0001896629
[,,2]
mean_kWh StDev_kWh
mean_kWh 0.0005627130 -0.0009174876
StDev_kWh -0.0009174876 0.0015794913



Appendix 22: LPA 2-component model Dataset 1

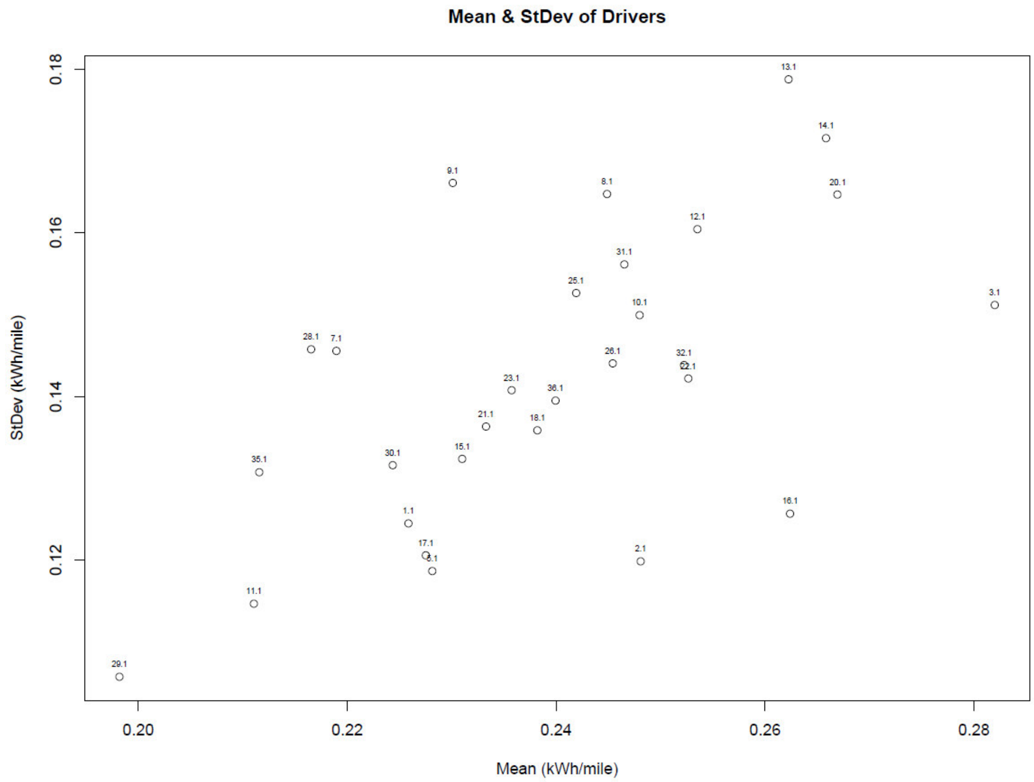
Latent Profile Analysis without 34.1 for Dataset 1 (Dans Data)



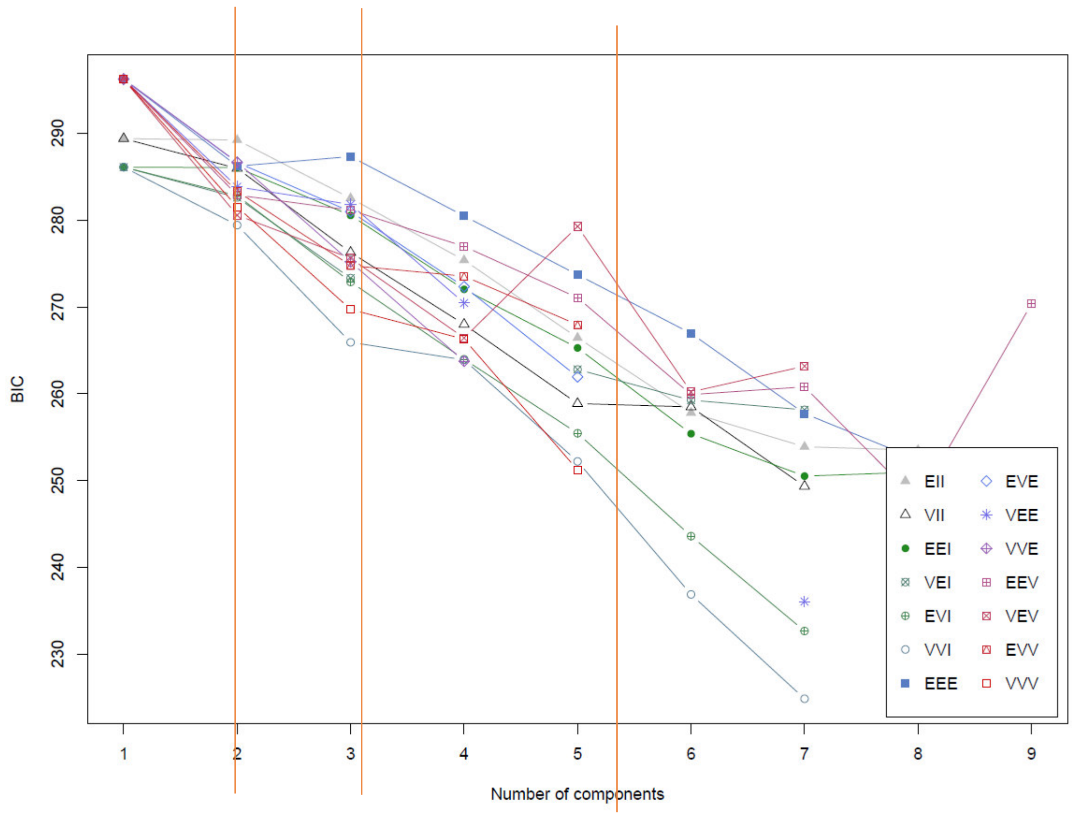
Appendix 23: Dendrogram Dataset 1 without 34.1

Based on Dendrogram:

- 2 clusters -> 1/2 different
- 3 clusters -> 1.1/1.2 different, 1.1/2 different, 1.2/2 different
- 4 clusters -> 1.1/1.2 different, 1.1/2.1 same, 1.1/2.2 different, 1.2/2.1 different, 1.2/2.2 different, 2.1/2.2 different



Appendix 24: Scatter plot Dataset 1 without 34.1



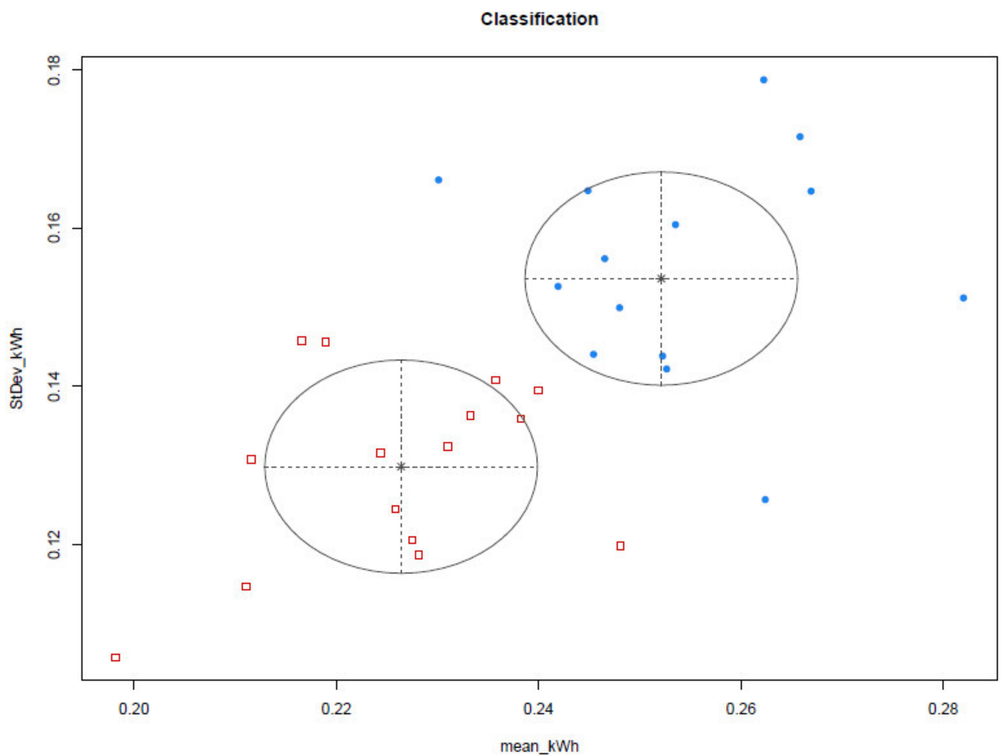
Appendix 25: BIC plot Dataset 1 without 34.1

Based on Latent Profile Analysis

- 2 clusters, looks most convincing
- 3 clusters, medium group covers entire scope of energy consumption
- 5 clusters, does not look convincing

```
> LP_2.1_o <- Mclust(d3.1_o, 2)
fitting ...
  |==========================================================================
=========================================| 100%
> summary(LP_2.1_o, parameters=TRUE)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EII (spherical, equal volume) model with 2 components:

 log.likelihood  n df      BIC      ICL
       154.7171 29  6 289.2305 282.5396

Clustering table:
 1  2
15 14

Mixing probabilities:
```

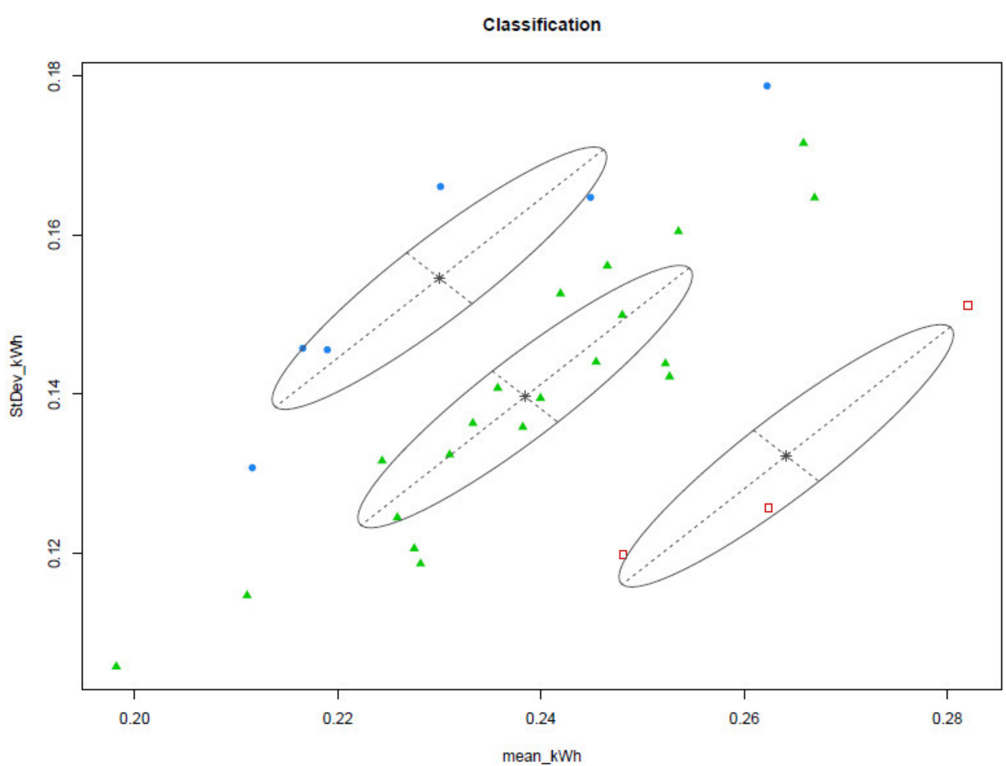
```
        1         2
0.4938913 0.5061087

Means:
              [,1]      [,2]
mean_kWh  0.2264037 0.2521208
StDev_kWh 0.1298654 0.1536360

Variances:
[,,1]
             mean_kWh    StDev_kWh
mean_kWh  0.0001816956 0.0000000000
StDev_kWh 0.0000000000 0.0001816956
[,,2]
             mean_kWh    StDev_kWh
mean_kWh  0.0001816956 0.0000000000
StDev_kWh 0.0000000000 0.0001816956
```



Appendix 26: LPA 2-component model Dataset 1 without 34.1

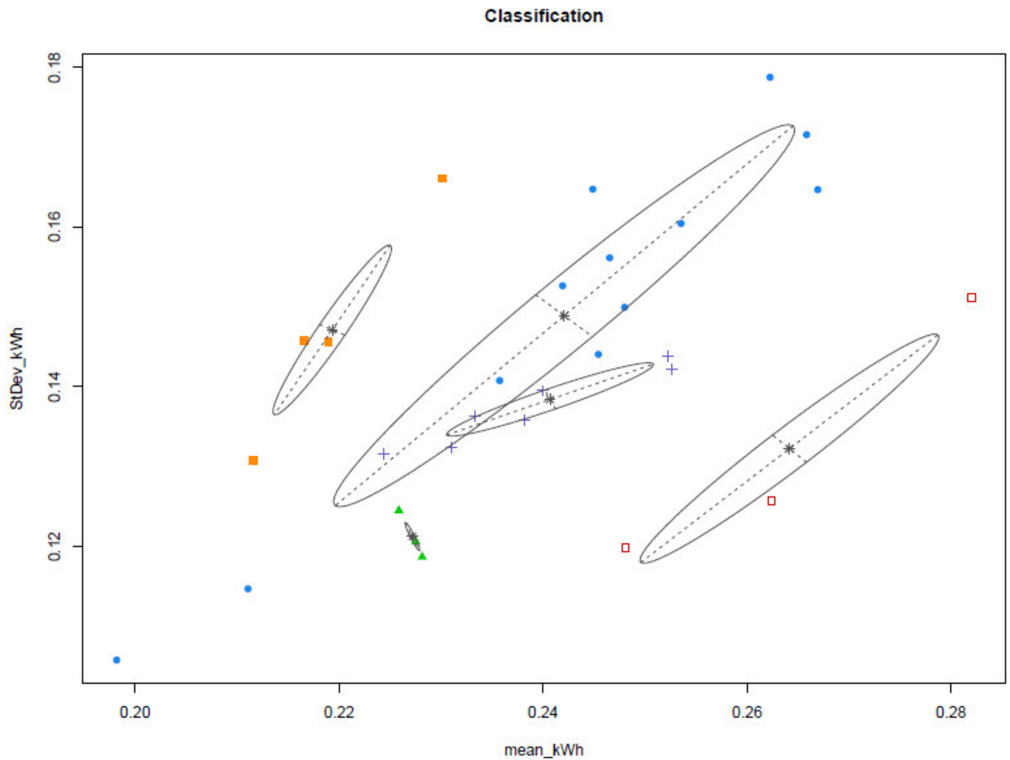
```
> LP_3.1_o <- Mclust(d3.1_o, 3)
fitting ...
  |========================================================================
========================================| 100%
> summary(LP_3.1_o, parameters=TRUE)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust EEE (ellipsoidal, equal volume, shape and orientation) model with 3 components:
```

```
 log.likelihood  n df      BIC      ICL
       162.1819 29 11 287.3235 286.0456

Clustering table:
 1  2  3
20  3  6

Mixing probabilities:
        1         2         3
0.6974969 0.1036055 0.1988976

Means:
               [,1]      [,2]      [,3]
mean_kWh  0.2384349 0.2641326 0.2299986
StDev_kWh 0.1397074 0.1322704 0.1545844

Variances:
[,,1]
             mean_kWh    StDev_kWh
mean_kWh  0.0002712690 0.0002510216
StDev_kWh 0.0002510216 0.0002721135
[,,2]
             mean_kWh    StDev_kWh
mean_kWh  0.0002712690 0.0002510216
StDev_kWh 0.0002510216 0.0002721135
[,,3]
             mean_kWh    StDev_kWh
mean_kWh  0.0002712690 0.0002510216
StDev_kWh 0.0002510216 0.0002721135
```



Appendix 27: LPA 3-component model Dataset 1 without 34.1

```
> summary(LP_5.1_o, parameters=TRUE)
----------------------------------------------------
Gaussian finite mixture model fitted by EM algorithm
----------------------------------------------------

Mclust VEV (ellipsoidal, equal shape) model with 5 components:

 log.likelihood  n df      BIC      ICL
       181.7273 29 25 279.2722 275.3919

Clustering table:
 1  2  3  4  5
 3  3  4 12  7

Mixing probabilities:
        1         2         3         4         5
0.1025729 0.1034480 0.1369193 0.4467520 0.2103078

Means:
               [,1]      [,2]      [,3]      [,4]      [,5]
mean_kWh  0.2271877 0.2641559 0.2193313 0.2420601 0.2406859
StDev_kWh 0.1212504 0.1322585 0.1471300 0.1489012 0.1384169

Variances:
[,,1]
               mean_kWh     StDev_kWh
mean_kWh   5.173992e-07 -1.182688e-06
StDev_kWh -1.182688e-06  3.025530e-06
[,,2]
              mean_kWh    StDev_kWh
mean_kWh  0.0002148284 0.0002045509
StDev_kWh 0.0002045509 0.0002056943
[,,3]
              mean_kWh    StDev_kWh
mean_kWh  3.374377e-05 5.934865e-05
StDev_kWh 5.934865e-05 1.128365e-04
[,,4]
              mean_kWh    StDev_kWh
mean_kWh  0.0005113596 0.0005264828
StDev_kWh 0.0005264828 0.0005725577
[,,5]
              mean_kWh    StDev_kWh
mean_kWh  1.026879e-04 4.429563e-05
StDev_kWh 4.429563e-05 2.108830e-05
```



Appendix 28: LPA 4-component model Dataset 1 without 34.1

Influence of acceleration on driving behavior

Apart for the energy consumption clustering, at an earlier state this research tried to understand the influence of acceleration and speed on the energy consumption of electric vehicles based on the eco-driving concept. Therefore, two datasets are compared which collected data on acceleration, speed, and energy consumption of an electric vehicle, for their statistical similarity concerning state of charge consumption. This analysis reveals that both datasets are statistically similar.

In addition, it is proven that the two acceleration measurement methods used for Dataset 1 and Dataset 2 respectively do generate different results. These clusters of drivers are compared according to aggressive and non-aggressive driving behavior as defined in the eco-driving concept

As mentioned the SDR is an area within the friction cycle of wet roads that applies an amount of mental workload on drivers, which is mainly determined by acceleration and speed, that ensures safe driving [38].This area is limited by 2.5m/s^2 for lateral acceleration and positive longitudinal acceleration, and by 3.0m/s^2 in negative longitudinal direction [38]. The upper left and right edges between these straight lines, spanning the boundaries in longitudinal and lateral direction, are described by Equation 5 and Equation 6. The lower left and right edges between the straight lines, spanning the negative acceleration boundaries in longitudinal and lateral direction, are described by Equation 7 and Equation 8. Appendix 29 a shows the SDR as described by these functions.

Equation 5: Equation describing the upper right edge of the Safe Driving Region [38]

$$a(x) = 0.509 * x^2 - 2.351 * x + 2.841$$

Equation 6: Equation describing the upper left edge of the Safe Driving Region [38]

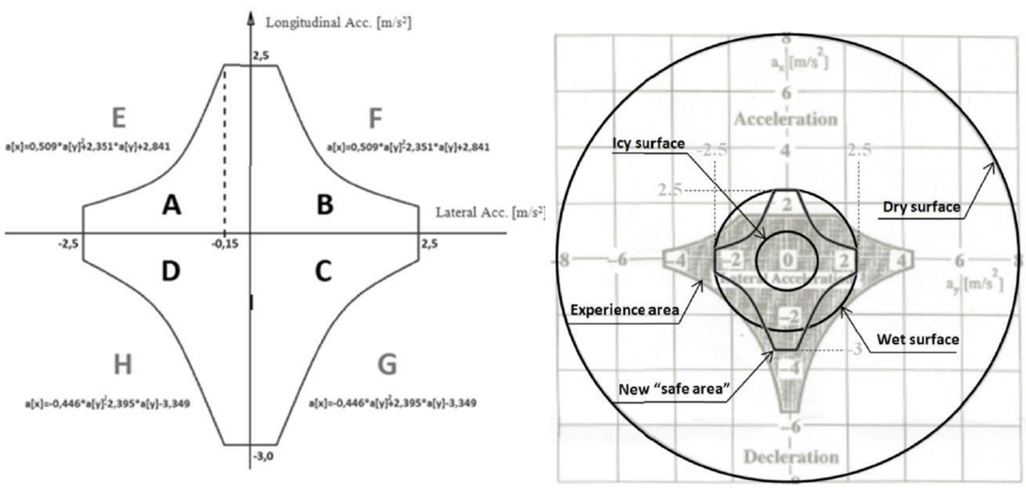
$$a(x) = 0.509 * x^2 + 2.351 * x + 2.841$$

Equation 7: Equation describing the lower right edge of the Safe Driving Region [38]

$$a(x) = -0.446 * x^2 + 2.395 * x - 3.349$$

Equation 8: Equation describing the lower left edge of the Safe Driving Region [38]

$$a(x) = -0.446 * x^2 - 2.395 * x - 3.349$$



a) b)

Appendix 29: Area of acceleration defining the Safe Driving Region [38]

Appendix 29 b displays the friction circles, while the biggest one represents the friction circle for dry surfaces, followed by the next smaller one for wet surfaces, and the smallest one for icy surfaces. The SDR described above is mainly inside the wet surface friction cycle with a little extension in negative longitudinal acceleration.

As mentioned aggressive participants had more than 10% of their acceleration measurement points outside the SDR. For the safe drivers it was less than 8% [38]. To compare the results for Dataset 1 and Dataset 2, this research is following the same approach by evaluating what percentage of acceleration measurement points per driver are outside of SDR. Thus, a threshold for a share of acceleration measurement points inside and outside of SDR, based on the boundaries given in literature, is set up to distinguish aggressive from non-aggressive acceleration. Even though the same method, is applied to analyze the acceleration in Dataset 1, a different algorithm is implemented for Dataset 2 since the one previously stated is not comparable with the data layout from Dataset 2. The algorithm developed for this research was tested on the data from Dataset 1 and could reproduce the results indicating that it worked properly.

A posteriori change of methodology

Originally this research was conducted as an extension to increase the sample size for Dataset 1. The goal was to validate the clusters of energy consumption for drivers, recorded at a higher resolution, as well as the influence of acceleration and speed on energy consumption. For this research it was possible to collect energy consumption at a higher resolution than 0.5% for SOC which drop down around every 800m (0.5 miles) like in [6]. This was done by collecting Amperage and Voltage which was reported and collected at a sampling rate of 1 second by the EV's OBDII. Based on the relationship between electrical power and Amperage and Voltage, electrical power for each recorded operation state was calculated. From looking at the positive or negative algebraic sign of energy consumption it was possible to determine overall energy consumption or recuperation of the EV for each recorded sample.

However, there were problems during the data collection which reduced the sample size that could be used for analysis by almost half. Even though the experiment for Dataset 2 was intended to be as similar as possible to the experiment for Dataset 1, a closer analysis of Dataset 1, in the scope of this research, indicated that differences of the data layout made the two datasets difficult to compare. Some of the minor challenges were the different units for Dataset 1 in mph and Dataset 2 in km/h which could be solved by converting mph to km/h (by applying a factor of 1.61 to km/h). One of the greater challenges was the different measurement methods for collecting acceleration data. For Dataset 1, acceleration was measured with a g-force meter and for Dataset 2 with a linear accelerometer. For the conversion for both measures a procedure is proposed in the section Future Research in this research since the implementation would have exceeded the scope of this research.

As mentioned previously, even though a professional solution was used for data collection there were still some issues with the collection process, meaning that not all data was collected correctly. The collecting device lost GPS signal during some test drives for shorter or longer periods of time, which resulted in a virtual jump of the vehicle along the road. This was a problem since the GPS signal was intended to match the vehicle's position along the route with its operation state parameters. This problem could not be completely solved even after applying an extension cord to place the device closer to the wind shield for a better GPS connection. Another problem that seemed to be related to the malfunction of the GPS was that, for some test drives, the recording of all CAN bus data stopped in the middle of the drive. Even intense search for possible reasons could not detect the problem. However, it must be mentioned that the researcher's ability for defect analysis was limited to an analysis of the experiment process. This revealed the downside of using a holistic solution provided by an external company for data collection, that it was not possible to access the device or the software. Thus, the problem could not be identified which could have provided knowledge to give recommendations for improvement for future research using a similar experimentation layout.

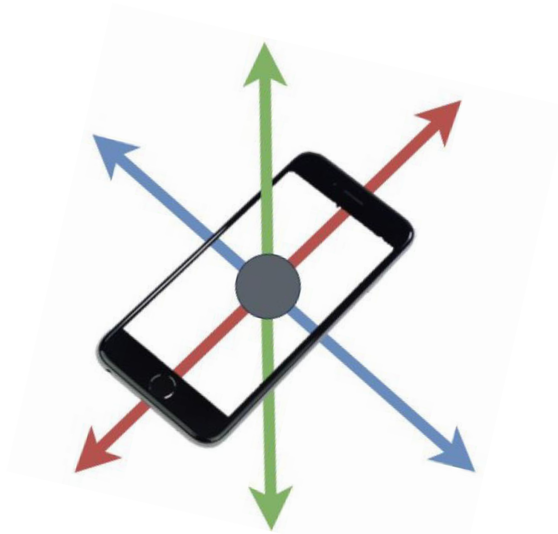
Acceleration is one of the parameters used to describe energy consumption in eco-driving which is used as a reference for this research. Concerning Dataset 1, drives 37 and 38 which were not included in the energy consumption analysis, were control drives for inefficient and efficient drivers respectively. The efficient control drive had 4.08% of their acceleration measurement points outside SDR and the inefficient control drive had 9.53% acceleration measurement points outside of SDR. The 30 samples for Dataset 1 plotted in a boxplot diagram, showed that for efficient drivers the mean percentage of points outside SDR was at around 5% and around 6% for the inefficient group. Both groups were statistically different for percentage of acceleration measurement points outside SDR [6]. Important to note is that this analysis was conducted only for the two groups of energy efficient and energy inefficient drivers but not for the medium energy consuming group, which was found in this research, since this previous research only found two clusters to be significantly different based on the clustering method.

Difference in acceleration measuring

For Dataset 2, regardless of the classification of energy consumption, all test drives, except for one, had zero percent of measuring points outside SDR. Only test drive 12 had around 1% of acceleration measurement points outside of the SDR and belonged to the medium energy consuming group. This is surprising since aggressive acceleration, in accordance to eco-driving, is considered to have a negative effect on energy consumption (i.e. more energy should be consumed). This might be either a lead that aggressive acceleration might not be an important factor for driving BEVs when energy efficient driving is pursued or that the acceleration data has to be analyzed at a more detailed level.

The discrepancy of acceleration measurement points within SDR between Dataset 1 and Dataset 2 could have been due to the fact that different measuring devices and software was used to collect acceleration data in each experiment. As mentioned the acceleration data for Dataset 1 and Dataset 2 differed strongly from each other. Beside the different means in acceleration measurement points outside SDR, which was around 5.7% for Dataset 1 and around 0% for Dataset 2, the values of Dataset 1 were multiplied by a factor of around when compared to the values from Dataset 2. This is potentially due to the different setups for the acceleration data. The first dataset used an iPhone 6 with the application SensorPlay to collect data while Dataset 2 used a Samsung Galaxy S7 Edge with the application Physics Toolbox Suite. Both phones are recent models and their acceleration sensors should be sufficiently accurate. Beside the differences in hardware and software, which should not have a huge effect on the acceleration measurements since both phones and software are considered to be accurate, the main difference between the two setups was the measuring method.

Appendix 30 shows the orientation of both cellphones in the car, with the blue arrows representing the x-axis recording lateral acceleration, the red arrows representing the y-axis recording longitudinal acceleration, and the green arrows representing the z-axis recording translational acceleration.

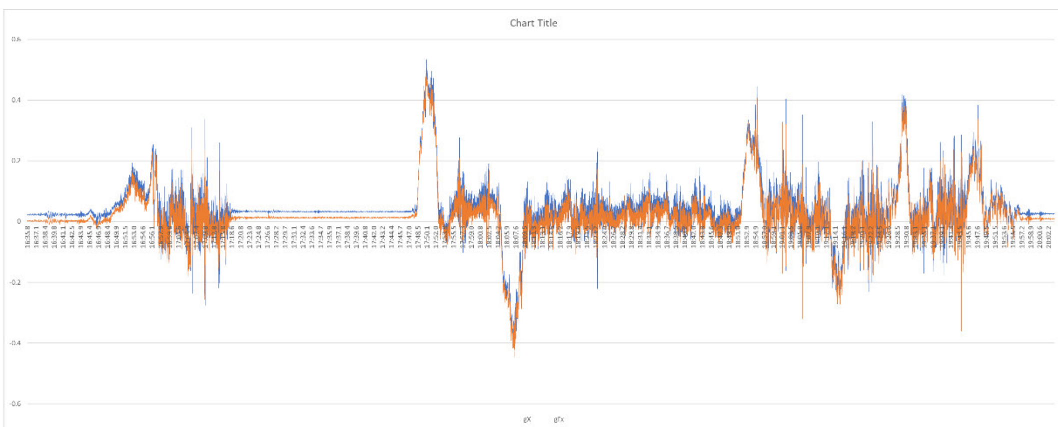


Appendix 30: Cellphone orientation showing z-axis (green), x-axis (blue) and y-axis(red) [6]

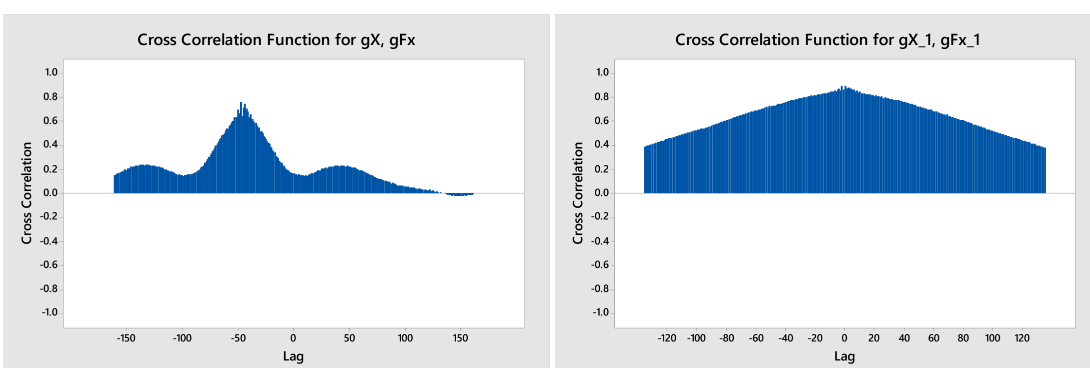
For Dataset 1, the acceleration was measured with a g-force meter which reads direct values from the cellphone accelerometer. Besides the different scale in g (9.81 m/s^2), it would record a constant value of 1g in the z-direction since this is the constant acceleration of earth's gravity. This would not affect the x- and y-direction measurements, which were of interest in previous research and this research if the cellphone's orientation is constantly perpendicular to the earth's center of gravitation, so that the z-axis would be directly through the center of earth's gravitation. However, since the test route includes different elevations, the g-force gets distributed between the z-axis and the other axes which results in changes of acceleration values for both the x- and y-axis without the vehicle being accelerating. This gives a possible explanation for why the acceleration values of Dataset 1 are generally larger in the x- and y-directions than the ones from Dataset 2.

For Dataset 2, a linear accelerometer is used that derives linear acceleration from the internal cellphone accelerometer, only considering changes in acceleration for all axes and eliminating the constant earth acceleration on the z-axis of 1*g using further cellphone sensor like a gyroscope. The values are recorded in the physical unit for acceleration (m/s^2). So, these values represent the relative change in acceleration of the vehicle without being influenced by the earth's gravitation during changing elevations along the test route.

To understand the differences between these two measurement methods, g-force meter and linear accelerometer, further test drives were performed with both devices recording acceleration data. First, to test the difference in sensors for each cellphone and the difference in recording software for each app, 2 test drives were conducted over a distance of less than 1 mile with both cellphones recording acceleration with a g-force meter. Plotting and comparing the results from both recordings, visually they



Appendix 32: Data for test drives 1.2 for Calibration test for iPhone S6 and Android Samsung Galaxy S7 Edge in x-direction



Appendix 31: Cross-correlation test for test drives 1.1 and 1.2 in x-direction

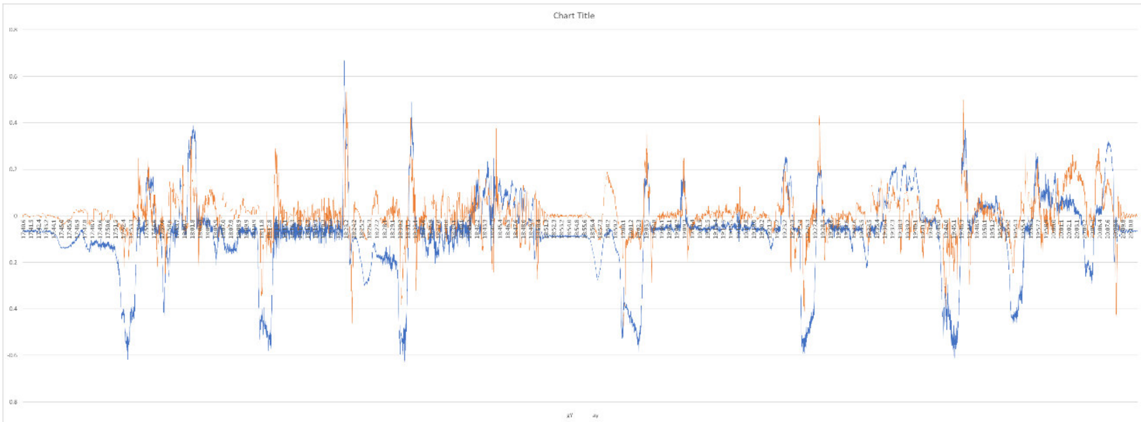
look very similar. A cross correlation test revealed a mean correlation value of around 0.85 in x direction and around 0.95 in y direction, indicating that they are highly correlated for the respected directions.

Appendix 32 exemplary shows the recordings in the x-direction for the first test drive using a g-force meter on both devices, with the blue line representing the values measured by the iPhone and the orange line representing the values measured by the Android Phone. Appendix 31 show the cross-correlation between both signals.

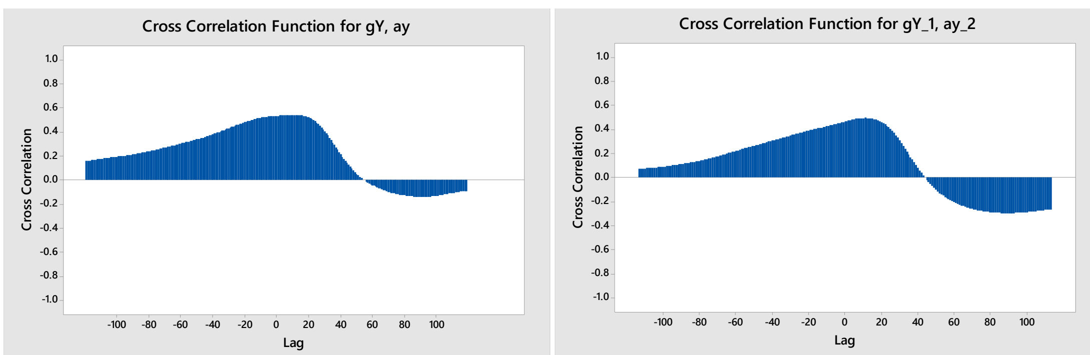
In a second test, 2 further test drives were conducted using their cellphones' original measurement methods i.e. g-force meter for the iPhone S6 using SensorPlay and linear accelerometer for the Android Samsung Galaxy S7 Edge using Physics Toolbox Suite. From a qualitative, visual analysis of both datasets, even though the two signals show a similar pattern they look different. The cross-correlation reveals a correlation value of around 0.58 for signals in the x-direction and a value of 0.5 for signals in the y-direction, indicating that they are not highly correlated.

A visual analysis, especially in the y-direction, shows that there are negative acceleration peaks with a plateau at the peak point for the iPhone data. This could be a result of the g-force getting distributed to the y-axis due to inclination of the car when applying the brakes.

Appendix 34 exemplary shows the acceleration measurements for the iPhone using a g-force meter and the Android Phone using a linear accelerometer in the y-direction, with the blue line representing the iPhone measurements and the orange line representing the Android Phone measurements. Appendix 33 shows the cross-correlation values for both test drives conducted for this analysis.



Appendix 34: Test drive 2.1 for using g-force meter using at iPhone and linear accelerometer at Android Phone in y-direction



Appendix 33: Cross-correlation for test drives 2.1 and 2.2 in y direction

Results from acceleration measurement comparison

The results from these test drives show that both setups collect similar acceleration data when using different phones and different applications as long as the measurement method is the same, here g-force meter. Furthermore, it gives a possible explanation as to why the values for the acceleration measurements with the g-force meter are significantly higher than the ones from the linear accelerometer. Since these plateaus of negative acceleration for the g-force meter from the iPhone are clearly outside the SDR of 2.5 m/s^2 and aggressive driving is determined by acceleration measurement points outside SDR, the plateaus increase this number considerably compared to a single peak which is mostly seen when using the linear accelerometer measurements form the Android Phone.

Conducting the SDR analysis for these two test drives reveals a percentage of 0 acceleration measurement points outside of the SDR for both test drives concerning the Android Phone data. For the iPhone data a percentage of 25.5 for test drive one and a percentage of 26.7 for test drive two is obtained. For the data collected with the iPhone, the values are considerably higher than the ones from Dataset 1. This is due the fact that it was deliberately driven very aggressively to generate a high amplitude for the measuring devices. The results show that the measurement method is relevant for the SDR analysis. G-force meter generates considerably more acceleration measurement points outside SDR than the linear accelerometer and supports the claim that the plateaus of measuring point around the peaks have a strong influence on the results of the SDR analysis.

A comparison of acceleration and speed data would have given insights about the driving behavior in both datasets. However, since the acceleration data of the first dataset was collected with a different method than the data for the second dataset, this biased the results for the SDR analysis so that they could not be used for comparison. Even though it was not possible to make both datasets comparable, this research can prove that the influence of the different data collection methods is not negligible. In addition, it showed that the g-force meter used for the first dataset generated considerably higher values in SDR analysis in test drives than the linear accelerometer used for Dataset 2. This is a valuable contribution for any future research that intends to collect acceleration data, especially in the context of understanding driving behavior.

Conclusion and future research

During the investigations of this research, major differences between the acceleration data from Dataset 1 and Dataset 2 was revealed. This was due to the fact that the acceleration data for both datasets was measured with different methods (a g-force meter for Dataset 1 and a linear accelerometer for Dataset 2). This research showed that there are considerable differences in the results for these measuring methods on the same test drive. Also, it shows that the values from the SDR analysis of Dataset 1 are by a factor of 10 larger than the ones from Dataset 2. An assumption is that this is due to the contribution of earth's gravity to acceleration in the x- and y-direction through inclination of the vehicle while either accelerating or braking or going uphill or downhill along the test route.

However, Dataset 1 consists of a large number of test drives which, apart from the acceleration measurement, are accurate. To make Dataset 1 comparable for future evaluation the biasing effect of earth's gravitation would need to be stripped from the dataset. Earth's gravitation has a strong effect on the acceleration in the y-direction when the EV goes uphill or downhill so, to erase the acceleration of gravity in the y-direction, degrees of elevation could be derived from the altitude data from Dataset 1 and, based on this, the contribution of gravity in the y-direction could be determined. This might give more accurate results for the acceleration values in the y-direction. The limits of this approach, however, is that inclination of the vehicle, which seemed to have a great impact on the results conducted in this research for comparing the measurement methods, is not recorded and hence cannot be excluded from the measurements. This problem also arises in the y-direction when braking or accelerating as well as in the x-direction when going around a corner at sufficiently high speed.

BIBLIOGRAPHY

[1] K. Dudenhöffer, *Akzeptanz von Elektroautos in Deutschland und China*. Wiesbaden: Springer Fachmedien Wiesbaden, 2015.

[2] M. Wietschel, P. Plötz, A. Kühn, and T. Gnann, "Fraunhofer-ISI-Markthochlaufszenarien-Elektrofahrzeuge-Zusammenfassung," 2013.

[3] *GEVO 2018.* [Online] Available: https://www.iea.org/gevo2018/?utm_content=buffer8f643&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer. Accessed on: Jul. 08 2018.

[4] L. Grubwinkler, "A modular and dynamic approach to predict the energy consumption of electric vehicles," *https://www.researchgate.net*, 2013.

[5] R. Zhang and E. Yao, "Electric vehicles' energy consumption estimation with real driving condition data," *Transportation Research Part D: Transport and Environment*, vol. 41, pp. 177–187, 2015.

[6] D. Kowalsky, "Quantifying behavioral impacts on electric vehicle efficiency," University of Rhode Island, Kingston, 2017.

[7] "statistic_id425466_annual-sales-of-new-energy-vehicles-in-china-2011-2016-by-type,"

[8] "statistic_id642799_sales-figures-for-electric-vehicles-in-the-eu-from-2010-2015,"

[9] "statistic_id698294_number-of-electric-passenger-cars-in-use-in-belgium-2007-2017,"

[10] "statistic_id788265_sales-of-electric-bicycles-in-sweden-2014-2017,"

[11] “study_id50929_electric-vehicles-in-the-us,”

[12] *U.S. electric vehicle sales by brand 2017 | Statistic.* [Online] Available: https://www.statista.com/statistics/665958/sales-of-electric-vehicles-in-the-us-by-brand/. Accessed on: Jun. 12 2018.

[13] F. Wefering, S. Rupprecht, S. Bührmann, and S. Böhler-Baedeker, “Guidelines. Developing and Implementing a Sustainable Urban Mobility Plan,” *European Commission - Directorate-General for Mobility and Transport*, 2014.

[14] S. Shaheen, C. Adam, and Z. Ismail, “Shared Mobility Resources: Helping to Understand Emerging Shifts in Transportation,” https://escholarship.org/uc/item/73t0405d, 2017.

[15] B. Jäger, M. Wittmann, M. Lienkamp, “Agent-based Modeling and Simulation of Electric Taxi Fleets,” *Sustainable Mobility in Metropolitan Regions*, 2017.

[16] M. Lienkamp, *Status electromobility 2016 or how tesla will not win*: https://www.researchgate.net, 2016.

[17] M. Masih-Tehrani, M.-R. Ha'iri-Yazdi, V. Esfahanian, and A. Safaei, “Optimum sizing and optimum energy management of a hybrid energy storage system for lithium battery life improvement,” *Journal of Power Sources*, vol. 244, pp. 2–10, 2013.

[18] G. Wager, M. P. McHenry, J. Whale, and T. Bräunl, “Testing energy efficiency and driving range of electric vehicles in relation to gear selection,” *Renewable Energy*, vol. 62, pp. 303–312, 2014.

[19] P. Clarke, T. Muneer, and K. Cullinane, “Cutting vehicle emissions with regenerative braking,” *Transportation Research Part D: Transport and Environment*, vol. 15, no. 3, pp. 160–167, 2010.

[20] T. Chen, K. Kockelman, and M. Khan, “Locating Electric Vehicle Charging Stations,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2385, pp. 28–36, 2013.

[21] M. A. Rahman, Q. Duan, and E. Al-Shaer, “Energy efficient navigation management for hybrid electric vehicles on highways,” in *Proceedings of the ACMIEEE 4th International Conference on Cyber-Physical Systems*, Philadelphia, Pennsylvania, 2013, p. 21.

[22] Preston, “Integration for seamless transport,” *International Transport Forum Discussion*, 2012.

[23] D. Goeke and M. Schneider, “Routing a mixed fleet of electric and conventional vehicles,” *European Journal of Operational Research*, vol. 245, no. 1, pp. 81–99, 2015.

[24] K. Kraschl-Hirschmann and M. Fellendorf, “Estimating energy consumption for routing algorithms,” in *IEEE Intelligent Vehicles Symposium, Proceedings*, 2012, pp. 258–263.

[25] A. Bolovinou, I. Bakas, A. Amditis, F. Mastrandrea, and W. Vinciotti, “Online prediction of an electric vehicle remaining range based on regression analysis,” in *2014 IEEE International Electric Vehicle Conference (IEVC): 17 - 19 Dec. 2014, Florence, Italy*, Florence, Italy, 2014, pp. 1–8.

[26] J. C. Ferreira, V. Monteiro, and J. L. Afonso, “Dynamic range prediction for an electric vehicle,” in *2013 World Electric Vehicle Symposium and Exhibition (EVS27): Barcelona, Spain : November 17-20, 2013*, Barcelona, Spain, 2013, pp. 1–11.

[27] M. Baum *et al., Eds., Speed-Consumption Tradeoff for Electric Vehicle Route Planning*. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany, 2014.

[28] J. Asamer, A. Graser, B. Heilmann, and M. Ruthmair, “Sensitivity analysis for energy demand estimation of electric vehicles,” *Transportation Research Part D: Transport and Environment*, vol. 46, pp. 182–199, 2016.

[29] T. Franke, F. Schmalfuß, and N. Rauh, “Human Factors and Ergonomics in the Individual Adoption and Use of Electric Vehicles,” in *Ergonomics and Human Factors for a Sustainable Future: Current Research and Future Possibilities*, A. Thatcher and P. H.P. Yeow, Eds., Singapore: Springer Singapore, 2018, pp. 135–160.

[30] J. N. Barkenbus, “Eco-driving: An overlooked climate change initiative,” *Energy Policy*, vol. 38, no. 2, pp. 762–769, 2010.

[31] G. Wager, J. Whale, and T. Braunl, “Driving electric vehicles at highway speeds: The effect of higher driving speeds on energy consumption and driving range for electric vehicles in Australia,” *Renewable and Sustainable Energy Reviews*, vol. 63, pp. 158–165, 2016.

[32] “statistic_id572085_number-of-publicly-available-fast-evse-chargers-in-the-us-2007-2017,”

[33] *2017 Update: Geographic Distribution of DC Fast Chargers | PlugInSites.* [Online] Available: http://pluginsites.org/2017-update-geographic-distribution-of-dc-fast-chargers/. Accessed on: Jun. 14 2018.

[34] X. Wu, D. Freese, A. Cabrera, and W. A. Kitch, “Electric vehicles’ energy consumption measurement and estimation,” *Transportation Research Part D: Transport and Environment*, vol. 34, pp. 52–67, 2015.

[35] C. Andrieu and G. S. Pierre, “Comparing Effects of Eco-driving Training and Simple Advices on Driving Behavior,” *Procedia - Social and Behavioral Sciences*, vol. 54, pp. 211–220, 2012.

[36] L. Eboli, G. Mazzulla, and G. Pungillo, “Combining speed and acceleration to define car users’ safe or unsafe driving behaviour,” *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 113–125, 2016.

[37] A. Alessandrini, A. Cattivera, F. Filippi, and F. Ortenzi, *Driving style influence on car CO2 emissions*, 2012.

[38] R. Vaiana *et al.,* “Driving Behavior and Traffic Safety: An Acceleration-Based Safety Evaluation Procedure for Smartphones,” *MAS*, vol. 8, no. 1, 2013.

[39] M. Yamakado, J. Takahashi, S. Saito, A. Yokoyama, and M. Abe, “Improvement in vehicle agility and stability by G-Vectoring control,” vol. 48, pp. 231–254, 2010.

[40] A. Shaout and A. Bodenmiller, *A Mobile Application for Monitoring Inefficient and Unsafe Driving Behaviour*, 2011.

[41] A. Diaz Alvarez, F. Serradilla Garcia, J. E. Naranjo, J. J. Anaya, and F. Jimenez, “Modeling the Driving Behavior of Electric Vehicles Using Smartphones and

Neural Networks," *IEEE Intell. Transport. Syst. Mag.*, vol. 6, no. 3, pp. 44–53, 2014.

[42] W. Weißgerber, *Elektrotechnik für Ingenieure - Formelsammlung: Elektrotechnik kompakt,* 6th ed. Wiesbaden: Springer Vieweg, 2018.

[43] C. S.S. Electronics, *CAN Bus Explained - A Simple Intro (2018).* [Online] Available: https://www.csselectronics.com/screen/page/simple-intro-to-can-bus/language/en. Accessed on: Jul. 03 2018.

[44] S. V. Liu, F.-L. Chen, and J. Xue, "Evaluation of Traffic Density Parameters as an Indicator of Vehicle Emission-Related Near-Road Air Pollution: A Case Study with NEXUS Measurement Data on Black Carbon," (eng), *International journal of environmental research and public health*, vol. 14, no. 12, 2017.

[45] *Highway Functional Classification Concepts, Criteria and Procedures*, 2013.

[46] D. S. Wilks, "Cluster Analysis," in *International geophysics series*, vol. 100, *Statistical Methods in the Atmospheric Sciences*, D. S. Wilks, Ed., 3rd ed., Amsterdam: Elsevier/Academic Press, 2011, pp. 603–616.

[47] J. D. Banfield and A. E. Raftery, "Model-Based Gaussian and Non-Gaussian Clustering," *Biometrics*, vol. 49, no. 3, p. 803, 1993.

[48] V. Melnykov and R. Maitra, "Finite mixture models and model-based clustering," *Statist. Surv.*, vol. 4, no. 0, pp. 80–116, 2010.

[49] J. Hedderich and L. Sachs, *Angewandte Statistik: Methodensammlung mit R,* 16th ed. Berlin, Germany: Springer Spektrum, 2018.

[50] Comprehensive R Archive Network, "Gaussian Mixture Modelling for Model-Based Clustering, Classification, and Density Estimation: Package 'mclust'," 2018.

[51] C. Fraley and A. Raftery, "Model-based Methods of Classification: Using the mclust Software in Chemometrics," *J. Stat. Soft.*, vol. 18, no. 6, 2007.

[52] US Departmen of Transportation, *Average Annual Miles per Driver by Age Group.* [Online] Available: https://www.fhwa.dot.gov/ohim/onh00/bar8.htm. Accessed on: Jul. 26 2018.

[53] Cyrus Moulton Telegram & Gazette Staff, *National Grid proposes rate increase for winter, but not as much as last year.* [Online] Available: http://www.telegram.com/article/20150915/NEWS/150919445. Accessed on: Jul. 26 2018.