Identifying behaviourally homogenous groups in commercial traffic with vehicles under 3.5t total weight by using cluster analysis

Heike Flämig [1] and Kerstin Rosenberger [1]

[1]*Hamburg University of Technology, Institute for Transport Planning and Logistics, Hamburg, Germany*

Abstract. This study focuses on the distinction and elaboration of different mobility patterns. Mobility patterns can be used as a foundation for the conception of design approaches regarding commercial traffic. The work discusses the question if behavioural homogeneous groups can be identified within commercial traffic utilizing vehicles under 3.5t total weight and how they can be clustered into mobility patterns.

1 Introduction

In the transport sector it has so far been impossible to implement the transition to a climate-compatible total for greenhouse gas emissions. Over 80 % of CO_2 emissions caused by traffic arise from motorized road traffic. [1] Efficiency has been substantially boosted since basis year 1990 by technical and organizational measures on road traffic. That has slightly reduced energy-related CO_2 emissions per kilometre covered. Yet growing vehicle mileage led to an increase in climatically relevant gases caused by traffic. The German government sees electrification of the power train as one of the main options to reduce traffic’s negative influence on the climate. Fleet structures and faster renewal cycles have caused the presumption that commercially registered vehicles offer potential for an early e-vehicle market in Germany.

This article considers the question of which areas of commercial traffic, on account of their deployment profile, offer the possibility of successfully integrating electric vehicles into the fleets. The project that furnished the data basis was concerned with the potentials for deploying electric vehicles in companies, and commenced in 2012. At this time there were only vehicles smaller than 3.5 tons, for which an average range of 140km can be assumed. Against this background, this research concentrates on these vehicles and therefore commercial traffic.

2 Data basis and evaluation

The data analysed below were in the “Hamburg-Wirtschaft am Strom “(WaS) research project funded by the German Ministry of Transport. The data were examined by a questionnaire-based online survey of these companies in Hamburg that have acquired publicly-funded battery electric vehicles (BEV). The data cover companies' structural and organizational characteristics, their ranges of activities, procurement behaviour, as well as characteristics of vehicle fleets and mobility patterns. The first describe deployment of the vehicles and were used as variables to form behaviourally homogenous groups. A group consists of similar driving profiles here. Since driving profiles, depending on initial interests, can be described by various parameters, the choice of the latter is not conclusive. Whereas for some questions, only vehicle deployment in traffic is of interest, for others the vehicle’s location between trips is interesting. In the context of this research project, what is primarily of interest is the opportunity of identifying driving profiles where the BEV can be deployed.

To identify the groups, the data was examined with the aid of a two-step cluster analysis using IBM

SPSS Statistics 24 software, since this makes it possible to analyse a data set consisting of differently scaled variables. Judgement of the quality of the cluster solution is based on the interplay between silhouette coefficients and the change of information criterion based on Akaike (AIC). The information criterion indicates how much the cluster solution improves by comparison with the previous one.
The data basis consists of 7153 vehicles from 345 companies. After revision of the data set the random sample comprises 4308 vehicles from 363 companies.
The variables were initially separated into trip-, company- and vehicle-specific variables.
Five variables were used for forming the cluster. The interplay of these gave character to a driving profile indicating a possible potential for using BEV. The two metrically scaled variables “average daily mileage” as well as “number of days when more than 140 km were covered” provide indications of possibilities for substitution. The mileages make it possible to deduce which vehicles on account of their range can be substituted without logistics changes. The categorical variable “location for an idle time of seven hours” serves to deduce interruptions to a trip and hence the options, e.g. for (interim) charging of BEV or for changing vehicles. The variable “Sample trips” as well as “Regularity at destination” allow for the geographical components of the driving profile.
The company-specific and vehicle-specific variables shown in Table 1 were treated as evaluation fields. These evaluation variables were not used in cluster formation, but serve for the deduction of deployment profiles from driving profiles.

Table 1. Evaluation variables

Vehicle-specific	**Company-specific**
Powertrain	Size of vehicle fleet
Vehicle segment	Company size category
Length of stay	Branch of commerce

3 Findings

After several iterations it became clear that its dichotomous nature gave the variable “destination regularity” a strong influence on cluster formation. It became apparent that all the vehicles in one cluster were serving destinations that were either changing or remaining the same. For further cluster formations, the data block was therefore sub-divided into 755 vehicles with destinations remaining the same, and 3553 with changing ones.
Analysis of the 755 vehicles with destinations remaining the same revealed that 99% of all vehicles were covering less than 140km per day. 60% serve company sites. Basically one can conclude from the case of the identical destinations that destination location is more crucial for BEV deployment than mobility profile. That is also of significance for charging infrastructure.
With the iteration of vehicles with changing destinations, the number of clusters varies. As might be expected, silhouette coefficient rises with an increasing number of clusters, since the groups become more homogenous. This higher cluster quality with very many clusters does not represent a satisfying result, since the number of cases per cluster progressively declines. Taking into account the alteration of AIC, a twelvefold partition produced the best silhouette coefficients. With high or medium silhouette coefficients (>= 0.5), seven of twelve clusters displayed strong homogeneity. These are printed in bold type in Table 2 and will be analysed in more detail below.

Table 2. Silhouette statistics of the twelvefold cluster solution found

Silhouette statistics				
Cluster	Cases	Mean	Min	Max
1	**917**	**.887**	**.663**	**.920**
2	159	.099	-.421	.337
3	84	-.297	-.605	.103
4	**926**	**.911**	**.735**	**.938**
5	190	.435	.318	.493
6	**181**	**.756**	**-.485**	**.871**
7	174	.294	-.105	.468
8	154	.375	-.185	.557
9	**133**	**.877**	**.160**	**.914**
10	**157**	**.850**	**.079**	**.919**
11	**153**	**.823**	**.607**	**.877**
12	**325**	**.901**	**.684**	**.925**
Total data set	3553	.744	-.605	.938

Table 3 below provides an overview of the expression of the characteristic values in cluster variables in the clusters identified. That all clusters are clearly characterized by a trip pattern is the most striking finding. Yet the same trip pattern is to be found in part in several clusters. Vehicles in clusters 1 and 12 exclusively serve trip pattern 2, yet they differ markedly in the other cluster characteristics. Bacher et al. [2] see the reason for that as the greater influence of categorical variables than metrically scaled variables in cluster formation.
In the present state of the technology, cluster 6 is not suitable for the deployment of BEV, since both the average daily mileage as the number of days (130) involving over 140km would only be attainable through major procedural alterations in logistics at companies. Even allowing for the standard deviation, with all the other clusters the average daily mileage and the number of days involving one of 140km are in the area where deployment of BEV would basically be feasible. For the remaining clusters and taking cluster 1 as the example, a possible deployment of BEV is investigated below using an analysis of the correlation with the evaluation variables (mobility profile).

Table 3. Overview of driving profile of seven selected clusters with changing destinations

Cluster No.	Number of vehicles (percent)	Average daily mileage in km (Standard deviation)	No. of days >140 km (Standard deviation)	Exclusive sample trip	Most frequent location for an idle time of > 7h
1	917 (25.8%)	48.36 (26.9)	15.50 (22.0)	2	Own company site
4	926 (26.1%)	57.35 (27.3)	9.76 (13.2)	3	Own company site
6	181 (5.1%)	122.06 (46.7)	132.70 (55.2)	4	Own private land
9	133 (3.7%)	57.56 (44.6)	11.44 (15.0)	4	Own company site
10	157 (4.4%)	54.43 (21.4)	84.80 (41.4)	4	Public road
11	153 (4.3%)	72.07 (21.0)	40.72 (54.4)	3	Public road
12	325 (9.1%)	88.37 (30.5)	15.48 (13.8)	2	Own private land

3.1 Driving profile of Cluster 1

Cluster 1 constitutes 25% of the sample (917 vehicles) that are suitable for deployment of BEV. The vehicles may serve changing destinations, but with their daily mileage of 48.36km and a standard deviation of 26.86km are far below the technical range of e-vehicles. These daily mileages are moreover distributed between several separate journeys during the day (trip pattern 2). With the vehicles in Cluster 1, on an average of 15.5 days per year, range conflicts will arise, with the vehicles covering over 140km per day here. Yet since almost 96% of them belong to medium or large vehicle fleets, appropriate organisation of the logistics will enable a substitute to be found from their own vehicle fleet. All the vehicles grouped in Cluster 1 park for idle times of over seven hours on the company's own site, which renders battery charging perfectly possible.

3.2 Mobility profile of Cluster 1

76.3% of the vehicles in Cluster 1 have a conventional powertrain. The proportion of electrically powered vehicles is 17%, with 1.6% having hybrid drive, and 5% having other powertrains. As shown in the diagram, all vehicle segments except the luxury class are represented in Cluster 1. The small/compact class with 410 vehicles (approx. 44%) and small and other transporters of up to 3.5t, with 345 vehicles (38%), constitute the highest proportion.

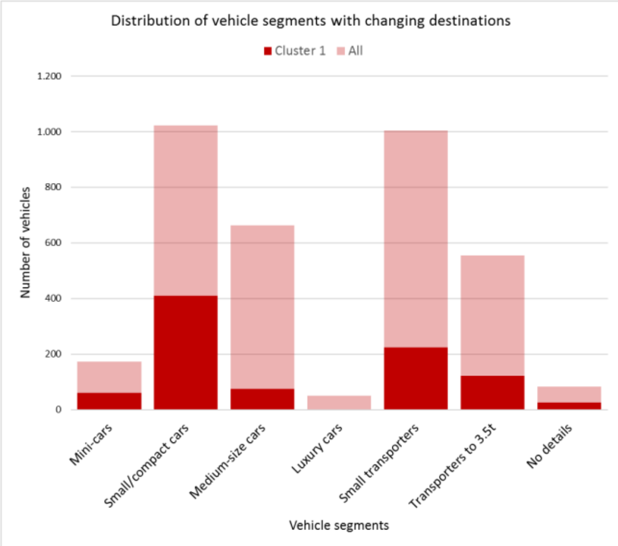

Figure 1. Vehicle segments

43.7% of vehicles do trips to provide professional services. The second highest proportion (32.4%) consists of those purely for carrying passengers. Yet another questionnaire as part of this project reveals that trips are often mentioned as purely for conveying passengers that, depending on the selected categorization, could also have been reckoned as being for providing professional services.

65% of the vehicles in Cluster 1 belong to large companies with over 250 staff. 17% belong to medium-sized companies with 50 to 249 staff. 70% of vehicles are in large vehicle fleets of over 50 vehicles. Company-specific variables suggest that in terms of logistics, there would be a reaction to BEV-specific properties.

Cluster 1 does not reveal any majority attribution of vehicles to one or more sectors of the economy. Cluster 1 contains vehicles from each of the 18 economic sectors investigated. This permits the statement that deducing the mobility behaviour of companies on the basis of the sector to which they belong serves no purpose here.

4 Conclusion for deployment potential of BEV

The average daily mileage of all vehicles in the data block is almost 70km, with a standard deviation of 50.8km, so that deployment of BEV would be basically feasible. Yet for BEV to be used, some driving profiles – for example clusters 6 and 10 – would require changes to logistics procedures. To calculate this potential, usage-based variables need to be taken into account. Here vehicle fleet size is of importance.

References

1. TREMOD; Transport Emission Model (2014): Version 5.53, Retrieved 15 November 2014
2. Johann Bacher, Knut Wenzig, Melanie Vogler: SPSS TwoStep Cluster – a first evaluation. Nürnberg, (2004)