Submitted to the 11th International Conference on Transport Survey Methods ISCTSC, September 24-29, 2017, Québec, Canada

Assessment of Real-world Vehicle Data from Electric Vehicles – Potentials and Challenges

Viktoriya Kolarova[a]*, Tobias Kuhnimhof[a], Stefan Trommer[a]

[a]*German Aerospace Center, Institute of Transport Research, Rutherfordstr. 2, 12489 Berlin, Germany*

Abstract

This paper introduces benefits and challenges related to collecting and analyzing vehicle data, in particular relating to electric vehicles. We use data from a current research project as an example. We address both technical and legal issues related to the data collection. On the technical side, pre-processing steps are needed to enhance data quality because the measured data are not faultless. On the legal side, data-privacy issues arise since precise GPS locations of the vehicles are captured. Two data collection methods are introduced and compared with each other. The advantages of vehicle data collection over other data sources for addressing various research questions are discussed.

Keywords: vehicle data collection, new technology, electric vehicles, assessment methods, longitudinal data, driving patterns

1. Introduction

Electric vehicles have been much discussed as a promising alternative to conventional combustion engine cars and hence as a potential solution for mitigating environmental problems caused by the transportation sector. At the same time, the diffusion rate of electric vehicles is still relatively low. This among other factors raises the question of how to accelerate the spread of electric vehicles by identifying acceptance barriers and examining how electric vehicles meet existing mobility needs. To obtain a better understanding about the potentials and barriers of electro-mobility, data about vehicle uses provide valuable insight. Furthermore, even if electric vehicles have no direct tail-pipe emissions, they still produce indirect emissions associated with the share of renewable energy available in the power grid during the charging of the vehicles. Hence, in order to examine the best way to unfold their potential for solving environmental issues, we need to examine how typical charging patterns affect the indirect emissions of electric vehicles and how to optimize them. Last but not least, we need a standardized method for evaluating different drive train concepts, such as battery electric vehicles (BEV) and plug-in hybrid electric vehicles (PHEV) and their performance under real-world conditions.

Detailed data on the charging and driving patterns of the vehicles is needed in order to analyze their use under real-world conditions. Furthermore, data collected over a longer period of time is needed to address crucial questions related to the effect of different factors, such as weather conditions on energy consumption, or the share of renewables on indirect emissions.

Advanced technologies in the new generation of vehicles, including electric vehicles, use software solutions in the vehicles to collect driving data in order to optimize the driving process or to collect emission relevant data during the life cycle of a vehicle. According to a directive of the EU it is required that all new vehicles with conventional combustion engine since 2001 have to be equipped with an onboard diagnostic system (OBD). The OBD protocols which include vehicle operation information can be retrieved from the system using diagnostic devices (Schaeffer, 2015, Beiter et al., 2012). Due to the continuing development of vehicle connectivity, such data can be sent wirelessly to an IT back-end. This enables new options for collecting driving data which has a huge potential for research. However, because this is a new method of data collection, know-how about the technical and legal requirements of recording the data, about pre-processing steps to ensure data quality, and about interpreting the data have to be gained first. Also, technical and data-privacy standards for vehicle data collection have yet to be established.

First steps for defining a so called standard "minimal data set" for research purposes was introduced in the past few years in the framework of a large scale government initiated program in Germany for funding research

* Corresponding author. Tel.: +49 30-67055-500; fax: +49 30-67055-283.
E-mail address: viktoriya.kolarova@dlr.de

projects in the field of electro-mobility (DDI et al. 2017, IVV, 2016). This “minimal data set” included the definition of required aggregated technical data on vehicle operations used for addressing various research questions on the potential and challenges related to the use of electric vehicles. The established requirements follow the “Data collection and reporting guidelines for European electro-mobility projects” (JRC, 2014). Vehicles which were a part of test fleets in government funded projects had to be able to provide the required data to the researchers. However, although electric vehicle data was measured in recent projects in Germany using the minimal data set, there is to our knowledge no available information on data quality provided by the different data providers or logger device suppliers. Also, there are notable differences in the data sets provided using different data collection methods which we discuss in this paper. Hence, even if there are efforts for standardization of minimal requirements on electric vehicle data collection, there is at the same time a lack of mainly technical standards ensuring the data quality and contents.

In this paper, we introduce the benefits as well as challenges related to collecting and analyzing real-world electric vehicle data using an example of the results of a recently finished project which was part of a German national research program. Within this project, a test fleet of around 150 electric vehicles was set up in the Berlin-Brandenburg region in Germany. The sample includes commercial and private BEVs as well as PHEVs. During the field test, vehicle data was collected over a period of one year. An online questionnaire survey and workshops with electric vehicles drivers provided additional information about the fleet, the usage of the vehicles etc.

The remainder of this paper is structured as follows: Section 2 provides a brief overview of the vehicle data collection. We introduce different data recording methods, the data set structure, data pre-processing steps as well as legal and data-privacy issues related to the data collection. In section 3 we will introduce selected applications of vehicle data collection in transportation and discuss the advantages and challenges related to the method compared to other data sources.

2. Electric vehicle data collection

2.1. Data collection methods and technical requirements

Within the research project introduced in this paper two vehicle data collection methods were used. The first method is equipping the vehicles with a data logging device which communicates in two directions: with the vehicle via the Controller Area Network (CAN) and with an IT back-end via 3G mobile networks. The second method is a direct vehicle data supply by the car manufacturer (OEM). Figure 1 introduces briefly how the two data collection methods work.

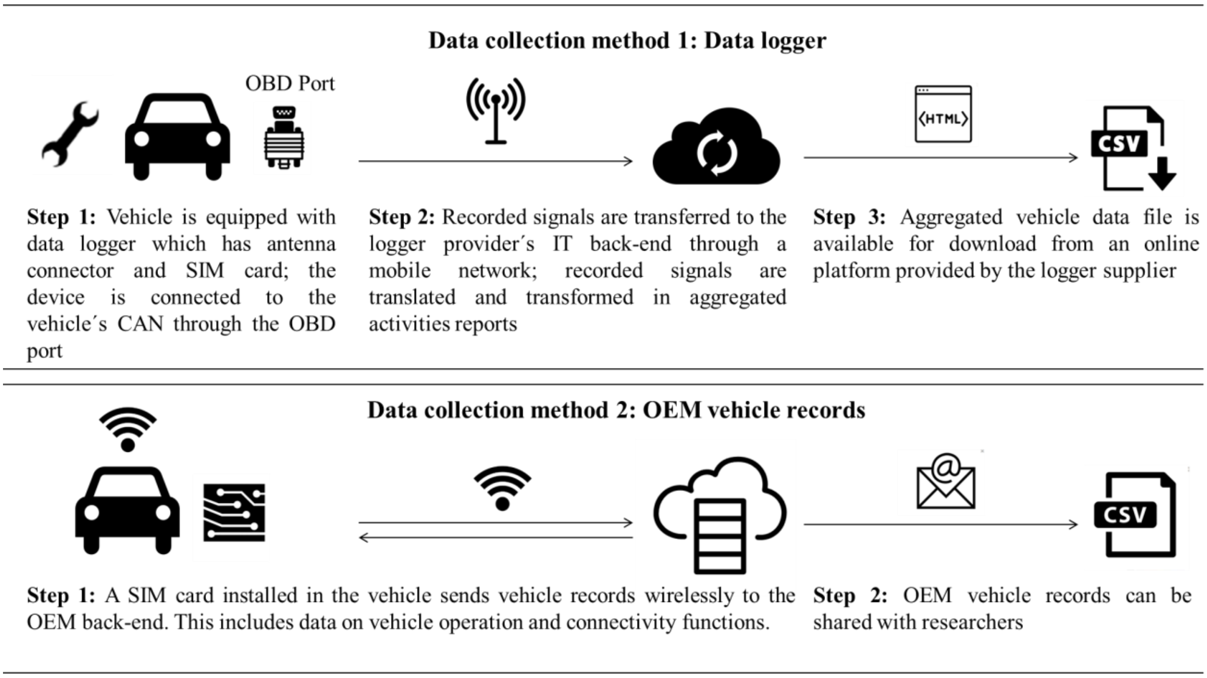

Figure 1. Overview of the two vehicle data collection methods

The first data collection method is technically feasible for all (new) vehicle types since data logging units can be connected to the onboard diagnostics (OBD) port available in every vehicle. However, the ability to interpret and translate CAN messages and signals available as DBC[†] files requires vehicle-specific protocol codes which can only be provided by the car manufacturer (OEM). Hence, the availability of data logging devices able to record vehicle data depends on the OEM´s willingness to cooperate with the data logger provider, i.e. the availability of protocol codes. Another option for translating the signals is applying reverse engineering but the quality is significantly worse and using this technique usually has also to be approved by the OEM.

The second data collection method, i.e. the vehicle records directly supplied by the OEM, became available recently due to vehicle connectivity development but also due to efforts of researchers and car manufacturer to establish cooperations for research purposes. Although the data quality by direct data supply is better than indirect data recording via data loggers, this data collection option is currently only available for certain vehicle models. Table 1 provides an overview of the vehicle models by data collection type available for research purposes. Note that we only summarized the vehicle models used in the research project presented in this paper. A list of OEMs whose electric vehicles data are used in research projects part of the national funding program is available online and is updated continuously (see PTJ, 2016). Also, some previous research projects used own devices for collecting vehicle data developed by project research partners. However, the devices were part of the research process and usually did not reach market maturity.

Table 1. Vehicle data collection methods

Data collection method	Vehicle Models
Data logger (external provider)	Renault ZOE, Renault Kangoo Z.E., VW e-Golf, VW e-Up, Mercedes B Class, Smart electric drive, (BMW i3)
Direct OEM vehicle records supply	BMW i3, Nissan Leaf, Nissan eNV

Using data loggers offered by an external provider is usually more expensive than the OEM vehicle records. This is mainly due to the cost for additional hardware, i.e. for the devices themselves, the related technical equipment (e.g. antenna connector) as well as for installing and removing the devices in/from the vehicles. In addition, since the data from the data loggers are transferred using extern mobile networks, there are additional cost for a SIM card and for the data transfer charged per recording month. On the other hand, there can be additional costs for software development (e.g. further development or adjustment of the devices) or technical support services. Another disadvantage of the data loggers is the necessity of additional appointments with the research participants for installation and removal of the device which require participants' engagement. Nevertheless, as mentioned above, there is no other feasible option for recording vehicle data from vehicles without advanced connectivity functions.

2.2. Data sets structure

The vehicle data sets include aggregated information about charging and driving activities, such as start and end time of each activity, mileage, state of charge, outside temperature, GPS positions at the beginning and end of each activity as well as distance travelled and consumed/charged energy. Figure 2 shows as an example of records for the same vehicle and day collected using the two data collection methods. In both data sets, each row refers to a single activity. The main difference between the two data sets is that the logger data set includes only trips and charging activities, while the OEM vehicle records also include parking events as well as gaps/false records. Also, the OEM vehicle records include a variable which indicates the activity type. In contrast, when using the logging data set each record has to be manually classified first by analyzing whether there are entries for kilometers driven (for trips) or charged energy (for charging event). Some records in the logger data, however, include both driving and charging events because the system failed to separate the two activities correctly. An example for such mixed records is provided in Figure 2. The second record in the data set is characterized by overall positive charged energy (which indicates a charging event) but also by distance travelled (which is a characteristic of a trip). When comparing the recorded activity (trips or charging activities) with the OEM vehicle records data set, it can be observed that this activity refers/corresponds to two trips and a charging event between them. Other recording failures and pre-processing data cleaning steps are introduced more detailed in the next part of the paper. Another logger records characteristic is that the recorded activity´s start time is a

[†] DBC (.dbc) refers to "data base CAN" and is a data/file format which contains definitions of CAN messages and signals

few minutes earlier and the activity´s end time is few minutes later than the time stamps recorded in the OEM data set. This might impact the accuracy of estimation of activities´ duration.

Data logger records

location_gid	trip_start	trip_end	trip_time	trip_dist_start	trip_dist_end	trip_distance	trip_energy_cons_pos	trip_energy_cons_neg	trip_energy_cons_sum	trip_soc_min	trip_soc_max	trip_position_start
55555	2015-07-01 08:38	2015-07-01 09:52	73.28	2232	2259	27	4809.99	-1745.94	3064.53	45.88	62.29	52.377994,13.413686
55555	2015-07-01 11:27	2015-07-01 15:26	239.62	2259	2303	44	8140.07	-4371.07	3769.17	37.74	70.55	52.543866,13.470969
55555	2015-07-01 17:05	2015-07-01 21:47	282.13	2303	2335	32	5509.5	-13306.48	-7804.74	21.67	102.11	52.378252,13.066298

OEM vehicle records

vin	zustand	timestamp_start_local	timestamp_end_local	km_stand_start	km_stand_end	soc_customer_start	soc_customer_end	consumed_electric_energy	latitude_start	longitude_start	latitude_end	longitude_end
55555	fahrt	2015-07-01 08:43	2015-07-01 09:49	2233	2259	60.5	45	4.94	52.37817944	13.41362417	52.54386056	13.47099167
55555	ruhe	2015-07-01 09:49	2015-07-01 11:29	2259	2259	45	45	0.02	52.54386056	13.47099167	52.54390389	13.47085472
55555	fahrt	2015-07-01 11:29	2015-07-01 12:07	2259	2272	45	37	2.69	52.54390389	13.47085472	52.50767361	13.32843944
55555	laden	2015-07-01 12:07	2015-07-01 14:27	2272	2272	37	69.5	NA	52.50767361	13.32843944	52.50718611	13.32809417
55555	fahrt	2015-07-01 14:27	2015-07-01 15:13	2272	2303	69.5	46	6.18	52.50718611	13.32809417	NULL	NULL
55555	ruhe	2015-07-01 15:13	2015-07-01 17:07	2303	2304	46	45	0.09	NULL	NULL	52.37886028	13.06627861
55555	fahrt	2015-07-01 17:07	2015-07-01 18:17	2304	2335	45	20	6.09	52.37886028	13.06627861	52.50769333	13.32835944
55555	laden	2015-07-01 18:17	2015-07-01 21:29	2335	2335	20	100	NA	52.50769333	13.32835944	52.50769333	13.32835944
55555	ruhe	2015-07-01 21:29	2015-07-02 14:19	2335	2335	100	100	0.08	52.50769333	13.32835944	52.50752361	13.32812611

Note: „zustand“ = „activity type“; „fahrt“ = „trip“; „ruhe“ = „stop/standby“; „laden“ = „charging activity“

Figure 2. Structure of the two data sets

Although both data sets include almost the same information on vehicle activities, some of the variables as well as their content differ between the data sets or are not included in one of them. For instance, information on velocity and topography of the route is only included in the logger data set. Also, the aggregation of selected variables in the logger data set are based on minimum and maximum values of the indicators during activity, while the same indicators are measured at activity´s beginning and end in the OEM vehicle records. The differences between the aggregation methods can lead to different results of the data analysis. This is especially the case with the recorded state-of-charge (SOC). The SOC is measured as a percentage of the electric vehicle´s battery capacity. The minimum and maximum SOC recorded for a single trip is usually the same as the start and end value of the SOC. However, recuperation[‡] during driving activity can lead to higher maximum SOC value during the trip than the value at the start of the trip. In cases where the differences between both are significant, this in turn leads to false estimation of consumed energy. Hence, a standardization of the variables´ units or techniques for aggregation of information on activity level is important for the data quality and possibility to compare different data sets.

Table 2. Overview of variables and differences between the data sets

Variable [units]	Description	Included in data logger records	Included in OEM vehicle records
ID	Anonymized vehicle ID	yes	yes
Start/end time [JJJJ-MM-TT hh:mm]	Time stamps (date and hour) at activity´s beginning/end	yes; local time	yes; local and UTC time
Start/end km mileage [km]	Km mileage at activity´s beginning/end	yes	yes
Distance travelled [km]	Distance travelled	yes	no; can be computed using km mileage
Duration [min]	Activity´s duration	yes	no; can be computed using time stamps
Customer SOC [% of battery capacity]	Customer state-of-charge	yes; min/max during activity	yes; at activity´s beginning/end
Consumed energy [Wh/kWh]	Total consumed energy; recuperation not included	yes; in Wh	yes; in kWh; recuperation not included
Start/end GPS latitude and longitude [double]	GPS location at activity´s beginning/end	yes	yes
Latitude	Minimum and maximum latitude during activity	yes; min/max during activity	no/not included
Velocity [min/category]	Duration of velocity category/ time spent in a defined velocity category (0-30 km/h; 30-60 km/h;	yes	no/not included

[‡] Recuperation refers to a process by that the electric motor is used as a generator for producing electricity when coasting and breaking.

	60-90 km/h; 90-120 km/h)		
Start/end battery temperature [^{0}C]	Battery temperature at activity´s beginning/end	yes	yes
Start/end outside temperature [^{0}C]	Outside temperature at activity´s beginning/end	yes	yes

Another vehicle data´ characteristic is that both data sets include information on the consumed energy. However, it can be reasonable to compute the consumed energy additionally using the recorded start and end SOC. An argument for this is that information on recuperation is not explicit included in the measured "consumed electric energy" so computing it using SOC allows estimation of a more accurate value. At the same time, when computing the consumed energy using the SOC requires considering technical characteristics of the specific vehicle, such as battery capacity. In case of electric vehicles, the usable battery capacity is not identical with the nominal battery capacity reported by the car manufacturer. This is because the "state-of-charge" management systems restrict the battery to be full or empty by 100 percent in order to preserve battery efficiency (Reed, 2011). As a result, only about 85% of the nominal battery capacity is usable depending on the vehicle model. Furthermore, the usable battery capacity reduces further during the time of use. Hence, by estimating the consumed energy using the SOC only the net, i.e. usable, battery capacity has to be taken into account and also vehicle characteristics have to be explicitly considered when interpreting the results of analysis on vehicle´s performance.

2.3. Data pre-processing

Vehicle data collection has advantages, which we will discuss in the next paper section, but there are also challenges. The data sets include gaps and often also implausible records which result from false signal interpretation by the system. This is especially the case by data recorded via logging devices. Thus, various pre-processing steps are required to enhance the data quality and to enable a comparison of different data sets before starting the data analysis. All analyses presented in this paper were performed using the Software R.

To compare and evaluate the two data collection methods described above, data from a single vehicle that was collected using both methods was analyzed. The selected vehicle was used for commercial purposes by a company based in Berlin (Germany). The data was collected over the period of one year. However, the number of days were vehicle data has been recorded varies between the two data sets and is lower than the number of working days per year (see Table 3). Reasons for the lower number of days include less frequent vehicle use (only days on which trip and/or charging activities took place are recorded), but also missing data caused by connection interruptions lasting more than one day. Such interruptions are, however, rather rare events.

Pre-processing steps for both data sets include analysis and preparation of implausible records and artefacts, implausible order of activities or gaps/missing data. When performing the pre-processing steps, differences in the structure of both data sets have to be considered. As mentioned above, OEM vehicle records include the type of each recorded activity; while in the logger records the activity type has to be identified and classified based on other information in the record. Also, parking events and gaps/missing data are explicitly included in the OEM records whereas in the data logger records they are not part of the data set. Hence, it is not clearly defined whether time gaps between activities in the logger data set are parking events or missing data.

First pre-processing steps for both data sets are to clearly identify the type of activities included in the data sets and to mark or remove implausible records. In the OEM vehicle records activities which cannot be clearly identified as a trip or a charging event by the system are classified as "stop/standby" events. These activities have to be classified rule-based using the OEM manual or plausible assumptions. In the logger data set, trips have to be first separated from charging events. An activity is identified as a trip when a distance of at least 1 kilometer was travelled and total (consumed) energy was negative. Activities with positive total energy and no distance travelled are identified as charging events. Activities which cannot be clearly assigned to a certain activity type are defined as "not classified" but were not excluded from the data set since they can be useful when analyzing movement patterns. Not classified activities in the data logger records result from failures in separation of charging and driving activities during data recording.

One common error in both data sets, as shown above, is that there are cases in which multiple records are created for the same activity due to connection interruption or false signal interpretation by the system. Thus, the second pre-processing step is to identify such multiple records and combine them into one activity. Also here, data set specific characteristics have to be considered before combining the records. For instance, there are short records (shorter than 5 minutes) among the OEM vehicle records which take place between two trips but which could not be clearly identified as trips in the first pre-processing steps. In the second data cleaning step, however, such activities can be defined as part of a trip since it can be assumed that these are too short to be parking events. Figure 3 provides two examples of multiple records included in the OEM vehicle records.

zustand	timestamp_start_local	timestamp_end_local	km_stand_start	km_stand_end	soc_customer_start	soc_customer_end	c
fahrt	2015-07-28 20:14	2015-07-28 20:23	2695	2699	78	75	
ruhe	2015-07-28 20:23	2015-07-29 09:21	2699	2699	75	75	
fahrt	2015-07-29 09:21	2015-07-29 09:53	2699	2727	75	53	
ruhe	2015-07-29 09:53	2015-07-29 09:55	2727	2727	53	53	
fahrt	2015-07-29 09:55	2015-07-29 10:29	2727	2743	53	45	
ruhe	2015-07-29 10:29	2015-07-29 11:33	2743	2743	45	45	

zustand	timestamp_start_local	timestamp_end_local	km_stand_start	km_stand_end	soc_customer_start	soc_customer_end	c
ruhe	2015-08-26 09:17	2015-08-26 11:10	3040	3040	15	15	
fahrt	2015-08-26 11:10	2015-08-26 11:16	3040	3041	15	14	
fahrt	2015-08-26 11:16	2015-08-26 11:22	3041	3042	14	13.5	
fahrt	2015-08-26 11:22	2015-08-26 11:57	3042	3071	13.5	6.5	
ruhe	2015-08-26 11:57	2015-08-26 12:05	3071	3071	6.5	5	

Note: „zustand" = „activity type"; „fahrt" = „trip"; „ruhe" = „stop/standby"

Figure 3. Example of multiple records in the raw OEM vehicle records

Similar error patterns also occur for charging activities in both vehicle data sets when charging events are interrupted by the vehicle or by the infrastructure and resumed at a later point in time. If such an interruption was no longer than one hour, no driving activity was recorded during the interruption and the vehicle´s position remains the same, than this interruption can be defined as part of the charging event. As with the treatment of trips, before combining multiple records into one charging activity, gaps between activities have to be analyzed. For instance, during the analysis we found many gaps which can be clearly identified as part of a charging event since they show overall positive total energy and zero distance travelled. About 57% of all gaps in the OEM vehicle records had these characteristics.

After the pre-processing steps described above, the number of activities included in both data sets has changed significantly. Table 3 gives an overview over the type of activities as recorded in the raw data in comparison with the activities identified after data cleaning.

Table 3. Overview of activity types – comparison between the two data collection methods

	Data collection method 1: Data logger records		Data collection method 2: OEM vehicle records	
Recorded days [n] *days on which trip or charging activity took place*	168		158	
	Before data cleaning	After data cleaning	Before data cleaning	After data cleaning
Recorded activities	497 (100%)	454 (100%)	1243 (100%)	923 (100%)
(only trips/charging activities or mixed records)	*(497)*	*(454)*	*(600)*	*(498)*
Trips	301 (61%)	305 (67%)	441 (35%)	393 (43%)
Charging activities	137(28%)	94 (21%)	159 (13%)	105 (11%)
Stops/standby	-	-	560[1] (45%)	399 (43%)
Not classified/ mixed records	59[2] (12%)	55[2] (12%)	n.a.[1]	-
Gaps/false data	n.a.[3]	n.a.[3]	83 (7%)	26 (3%)

[1] *stops/standby events in the OEM vehicle records can be a parking event but also a trip or charging events which could not be classified as such by the system; hence, this classification has to be done during the pre-processing phase*

[2] *the logger records includes only driving and charging events; however, there are cases where due to failure in separation during data records driving and charging activities are merged together; mixed records can be used for analyzing movement patterns (e.g. kilometer driven)*

[3] *in the logger records gaps/missing data are not explicitly included in the data as it is the case in the OEM vehicle records*

A comparison of both data sets shows that after the data cleaning process, there are still differences between the information covered in the data sets and therefore also in the data quality. The main reason for the different number of trips and charging activities are the mixed records, i.e. records combining both a charging event and a trip, in the logger records. However, even if mixed records cannot be used for certain analysis, such as estimation of duration of trips or charging events, they can still be used for analyzing movement patterns. For instance, these records can be used for computing daily driving distances. Overall, when comparing only the driving and charging activities (including the mixed records) in both data sets, the OEM vehicle records include about 9% more activities than the logger data set. In the next paper section, we discuss how these differences affect simple descriptive statistics on vehicle trips. Further pre-processing steps required only for certain research questions will be also introduced and discussed in section 3 of this paper.

2.4. Legal issues related to vehicle data collection/data privacy issues

Beside technical issues, there are also some legal and data-privacy issues related to the vehicle data collection. Data-privacy issues arise since precise GPS location of the vehicles is captured and depends mainly on the aggregation level and content of the data. Furthermore, exact location information is more confidential for certain companies that participated in our survey with their commercial vehicles than for others, for example security companies. Moreover, since personal data are collected, this requires the explicit agreement of the study participants, i.e. employees which use the test vehicle. From an organizational point of view, this can be a challenging issue, especially when collecting data from fleet vehicles available for use to many employees. Also, vehicle data collection in commercial fleets is in some cases directly in conflict with company´s work council regulations. Hence, data-privacy regulations regarding the collection of vehicle data for research purposes, especially when investigating commercial fleets, needs to be established.

Another point to be considered is that commercial users which participate at research projects are usually also interested in detailed data collected during field tests in order to evaluate the performance of the vehicle on company level. Here again, collecting private data of employees and providing it to the employer might be against the data privacy rights of the users of the fleet vehicle. In this context, in order to provide additional incentives to companies to support research projects without violating privacy rights, standardization of aggregated data that comply with data privacy regulations can be a considerable solution.

3. Transportation research applications

3.1. Analysis of driving and charging patterns

One of the most important applications of electric vehicle data collection is a detailed analysis of vehicle driving and charging patterns. Driving patterns can include time, duration and frequency of trips, number and distance of trips per day as well as daily, weekly, monthly or yearly kilometers driven. Charging patterns include time, duration and frequency of charging events as well as how much energy was charged and which charging speed was used. Also, charging needs with regard to vehicle-specific driving patterns can be examined. In this section we provide some examples for driving and charging patterns analysis. Moreover, longitudinal vehicle data provide detailed information on the regularity of trips, i.e. it allows for separating routine and rare driving events.

Figure 4 shows an example of descriptive statistics of trips including single-trip distance travelled, daily distance travelled for certain weekdays as well as monthly and yearly distance travelled. In order to compare the data quality of the two data sets, we again show results from both data collection methods for the same vehicle.

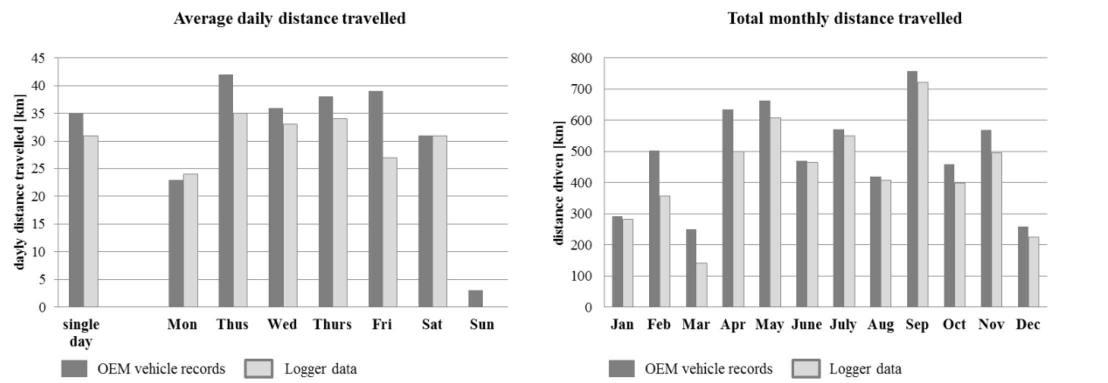

Figure 4. Descriptive statistics of vehicle kilometers travelled

Table 4 provides statistics on the daily distance travelled recorded in the OEM vehicle records compared to the logger data for the same period of time which were collected from the single vehicle introduced above. The analysis shows that in most of the cases (64%) an identical number of kilometers were recorded in both data sets. In many of the cases (28%), however, the OEM vehicle records include more kilometer than the logger records. This indicates that the information in the OEM vehicle records is more complete than that from the data loggers.

Table 4. Comparison between daily distances travelled recorded in the two data sets

Comparison between the data sets	Differences between data sets [km]	Frequency (base: 172 days)	Relative share [%]
More kilometer recorded using the data logger compared to using OEM vehicle records	More than 10 km *(max: 31 km, outlier: 153 km)*	7	5%
	Between 0 and 10 km	6	3%
No differences	0 km	110	64%
More kilometer recorded using the OEM vehicle records than using data logger	Between 0 and 10 km	29	17%
	More than 10 km *(max: 51 km, outlier: 167 km)*	18	11%

Analyzing electric vehicle charging events collected over a longer period of time can provide insights on typical charging patterns as well as important insights on user needs related to the charging infrastructure under real-world conditions. This can be an important output since charging infrastructure and demand play a crucial role for the acceptance of electric vehicles. This is because long recharging time and lack of public infrastructure are found to negatively influence the willingness to use such vehicles.

One possible application of electric vehicle data in this context is an analysis of average duration and average energy charged during recharging. Also, daytime-specific energy demand can be examined. Figure 5 provides an example for different charging patterns identified analyzing longitudinal data from the vehicle test fleet. The charging patterns were identified by cluster analysis using the total hourly charged energy for each vehicle collected over one year period of time. One example for interpreting the results of the cluster analysis is the following: looking at the results for vehicles belonging to cluster 4 shows that these vehicles are typically charged at the morning or at the noon (between 9 am and 2 pm).

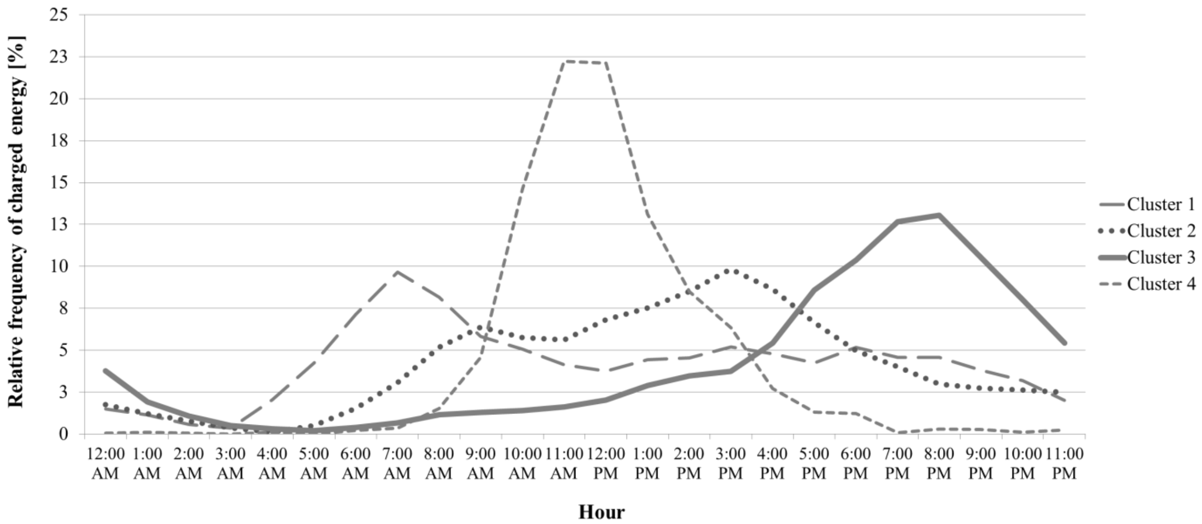

Figure 5. An example of different charging patterns

Another focus of research on driving and charging patterns can be spatial analysis of activities. The GPS location recorded at the beginning and at the end of each event allows detailed analysis of points of interest, such as trip starting points, destinations and charging locations. This is, however, only possible if additional information on the points of interest is available, such as the address of place of work or residence. Also, data mining techniques, such as clustering methods, have to be applied in order to cope with inaccuracies of the location measurement. One clustering method we used in the recent project is the DBSCAN clustering algorithm proposed by Ester et al. (1996) and successfully applied in another research project that used vehicle data (Hardinghaus, Lehne & Kreyenberg, 2016). The goal of the clustering is to assign certain GPS locations to a point of interest. Using the algorithm allows to assign different points to a certain location by defining a range in which the points have to be located when belonging to a pre-defined location. This method enables, for instance, the analysis of frequently selected destinations or charging locations.

Alternative established methods for collecting travel behavior data are travel surveys based on travel diaries/logbooks or additionally including GPS tracking devices. While such travel surveys are without doubt the most suitable method to obtain a comprehensive and multimodal picture of person travel, vehicle data that can be retrieved automatically emerges as a viable alternative to collect information on vehicle usage patterns. One of the advantages of passively collected vehicle data is an automatic recording of longitudinal data that goes along with low respondent burden and a lower need for respondent engagement. Thus longitudinal data recording even over longer periods of time becomes easily feasible.

Disadvantages of the vehicle data collection are, however, some technical issues influencing the data quality and the lack of information on trips purpose or vehicle use motives. A combination of questionnaires and vehicle data might therefore be advisable when such information is needed.

3.2. Analysis of energy consumption and emissions

Another application of the vehicle data collection can be detailed analysis of energy consumption and emissions of electric vehicles. There are different factors influencing energy consumption of electric vehicle including outside temperature, street topography, velocity, driving behavior as well as urban- or rural-specific conditions. Analyzing vehicle data that cover longer periods of time can provide important insights on the effect of such factors and hence on the performance of the vehicles under real-world conditions. Furthermore, more complex interrelation between the factors and also potentials for optimization of the vehicle use can be examined.

Figure 6 provides an example for simple analysis of the relationship between outside temperature and energy consumption using longitudinal data from a single vehicle. As shown on the figure, optimal energy consumption is at temperatures between 20 and 24^{0}C. When it is colder or warmer energy consumption increases for heating or cooling the vehicle. Here again, when interpreting the estimated energy consumption, the differences in direct estimated and SOC computed energy consumption have to be considered.

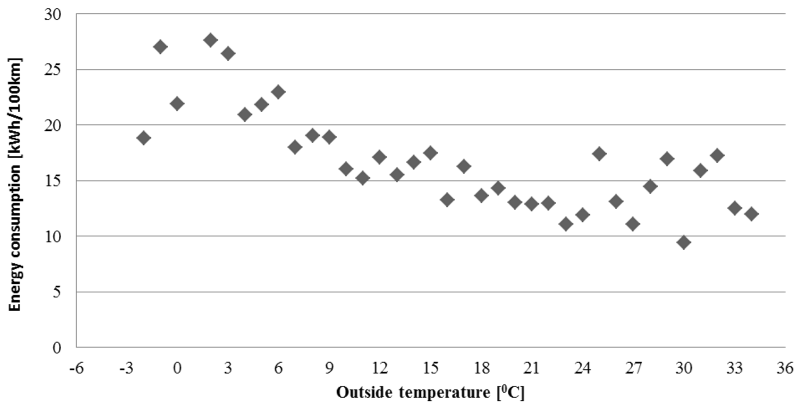

Figure 6. Relationship between energy consumption and outside temperature

Combing energy consumption of electric vehicle and some additional data allows for the estimation of vehicle emissions. The emissions from electric vehicles result, other than for conventional cars, not directly from the vehicle operation but rather from the production of the electric energy used for charging the vehicle (Holdway et al., 2010). Thus, electric vehicles have indirect emissions which depend on the sources used for electricity production (electricity mix). However, the electricity mix in the grid varies by times of day and day of the year and also depends on weather conditions. Indirect electric vehicle emissions as well as the factors which influence them under real-world conditions can be analyzed by combining vehicle data with data on the electricity mix. In the following, we briefly describe a method for analyzing indirect electric vehicle emissions applied in the presented study.

In the first step of the analysis, hourly charged energy during each charging event (in kWh) as well as energy consumption during trips (in kWh/km) are established using the vehicle records. In the second step, the vehicle records are combined with detailed data on power consumption for Germany for the same period of time. The power consumption data set includes information on the hourly relative share of energy from conventional (e.g. hard coal, lignite) and renewable (e.g. solar, wind) sources. Additionally, each of the sources is assigned a CO_2-equivalent emission factor. Finally, the charged gCO_2/kWh during one charging event is multiplied with the energy consumption during trips. As a result, indirect emissions in gCO_2/km are computed taking into account vehicle-specific charging times and energy consumption. Figure 6 provides an example of descriptive analysis of the indirect CO_2-emissions from one electric vehicle.

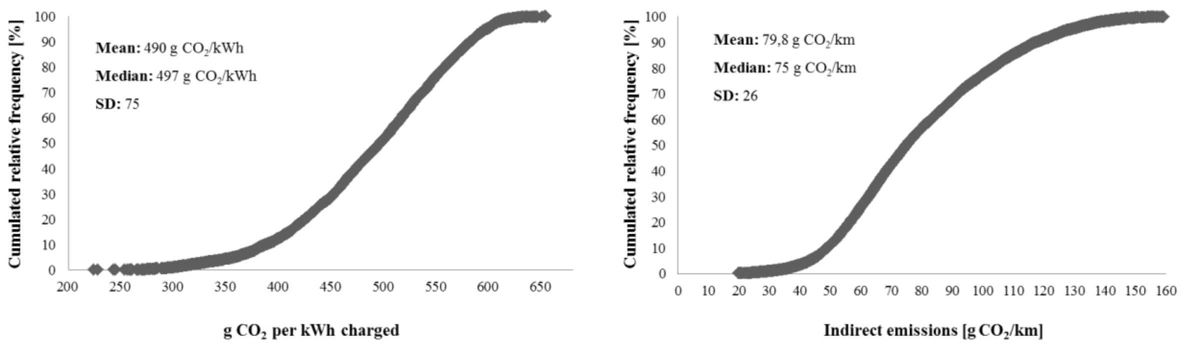

Figure 6. Estimated indirect CO_2-emissions of electric vehicles

3.3. Evaluation of different vehicle models and driving trains

Last but not least, analyzing data from different vehicle models or from vehicles with different driving trains can provide valuable insights on vehicle performance or how suitable different vehicle types are for certain mobility needs. However, comparing data sets from different vehicle types can be as challenging as comparing data sets from different data sources described above.

For instance, the minimal data set used in electro-mobility research projects does not include any variable that records fuel consumption of plug-in-hybrid vehicles or vehicles with range extenders (Deutsches Dialog Institut et al., 2017). This makes it impossible to estimate the energy consumption of these vehicle types under real-world conditions. Including information on fuel consumption in the required minimal data set is therefore important in order to compare the performance of battery electric vehicles with electric vehicles with additional internal combustion engine (ICE). Furthermore, data set quality and structure differ across vehicles from different manufacturers and vehicle models. This requires the development of methods for combining the different data sets in order to compare the different types of vehicles.

Also when comparing the performance of electric vehicles with conventional vehicles using vehicle data, specific characteristics of each of the driving trains have to be considered. One simple example is that the value of state-of-charge of electric vehicles depends on net battery capacity which is not the same as nominal battery capacity. In contrast to this, in vehicles with ICE the fuel tank capacity is always the same as the nominal one reported by the OEM.

4. Conclusions and outlook

The aim of this paper was to discuss the benefits and the limits of vehicle data collection as a data source for addressing research questions. Moreover, it aimed to address the topic in light of increasing connectivity and digitalization in today’s vehicles. For this purpose, two different methods of vehicle data collection were introduced along with selected research questions that can be addressed using such vehicle data. The first data collection method is using third party logging devices installed in the vehicle able to record and send information on vehicle activities to an IT back-end. The second one is using OEM vehicle records sent using connectivity functions of the vehicle.

The automatic vehicle data collection has the benefit that data is passively recorded and thus independent from research participants´ accurate and truthful response. When comparing both data collection methods, it can be concluded that the data quality from both methods is satisfactory. However, the OEM vehicle records include clearly more information and less implausible records than the third party data logger records. Also, with increasing vehicle connectivity collecting vehicle data automatically becomes easier without requiring extra time from the respondents for appointments to install or remove the data logging devices. This might additionally lower barriers for respondents to participate in research projects. At the same time, vehicle data collection using third party data loggers is still solution to be considered as many new vehicles do not have advanced connectivity functions or OEMs are not providing them for research purposes.

In this paper, we discussed the characteristics of vehicle data collection in the light of other data sources used in transportation research. Using detailed vehicle data on activities level over a longer period of time provides more detailed information about typical travel and charging patterns. Furthermore, routine as well as rather rare driving or charging events can be examined. This enables evaluation of vehicle performance and vehicle compatibility to user-specific mobility needs.

Despite the advantages of vehicle data, there are still some challenges related to data collection and interpretation. The data are not faultless and requires elaborative pre-processing steps to cope with gaps or implausible records. Also, data sets gathered using different methods of data collection differs in their structure, content or variable units. Hence, standardization of the variables in data vehicle records should be considered as a solution in order to improve comparability of data sets gathered using different methods but also for improving data quality. Know-how and experience exchange in discussions between researchers and data providers on possible solutions for challenges related to vehicle data collection might support the standardization process. Another option to facilitate the process of data analysis and to ensure sufficient data quality for research is defining clear process steps for dealing with implausible records or missing data. This in turn is only possible in a dialogue with data providers and OEMs since in-depth knowledge about the vehicle system is required.

In summary, advanced vehicle technologies including vehicle connectivity enables new methods for travel data collection which have a huge potential for research work in transportation. Unfolding the full potential of the methods for research purposes requires establishing technical and data privacy standards for vehicle data collection but also supporting a dialog between researchers and data providers as well as information transfer of lessons learned from recent research projects.

Acknowledgements

The research project “InitiativE BB” introduced in this paper was funded by the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety in Germany (grant number: 16EM2066-2). We would like to thank to the project partner, all co-workers at the Institute of Transport Research involved in the project for their work and to the study participant for their cooperation. We also want to thank to the co-workers at the project “CCS-Schnellladen Berlin” funded by the Federal Ministry of Economic Affairs and Energy in Germany (grant number: 16SBB017H) for exchanging experience related data collection using data loggers.

5. References

Beiter, K.,, Raetz, Ch., Garnatz, O. (2012) Gesetzliche On-Board-Diagnose und ODX. Vector Informatik. Retrieved from: https://vector.com/portal/medien/diagnostics/odx/Gesetzliche_OnBoard_Diagnose_und_ODX.pdf [14.08.2017]

DDI - Deutsches Dialog Institut GmbH, Ingenieurengruppe IVV GmbH & Co. KG, and NOW GmbH Nationale Organisation Wasserstoff- und Brennstofftechnologie (2017). Minimaldatensets zu Erhebung von Forschungsdaten in der Elektromobilität. Ergebnisse aus den regionalen Demonstrationsvorhaben der Bundesregierung. Retrieved from: https://www.now-gmbh.de/content/7-service/4-publikationen/3-modellregionen-elektromobilitaet/minimaldatensets-zu-erhebung-von-forschungsdaten-in-der-elektromobilitaet.pdf [01.08.2017]

Ester, M., Krieger, H. P., Sander, J., and Xu, X. (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise," Institute for Computer Science, University of Munich, vol. Published in Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)

JRC – Joint Research Center (2014). Data collection and reporting guidelines for European electro-mobility projects. Report EUR 26969 EN.

Hardinghaus, M., Lehne, M., and Kreyenberg, D. (2016). Charging behaviour and energy consumption of various electric vehicles in Berlin using different charging infrastructure (slow and fast) over one year. In: proceedings of EVS29 Symposium, Montreal, Quebec, Canada, June 19-22, 2016.

Holdway, A. R., Williams, A., Inderwildi, O., and Sir King, D. (2010). Indirect emissions from electric vehicles: Emissions from electricity generation. In: Energy & Environmental Science, 3 (12).

IVV (2016). Zentrales Datenmonitoring für regionale Demonstrationsvorhaben zur Elektromobilität in Deutschland. Retrieved from: http://www.ivv-aachen.de/aktuelles/70/f00c58fe09968394e1faece19487eb0e.html [10.08.2017]

PTJ – Projetträger Jülich (2016). Liste der PKW Fahrzeughersteller zur Förderrichtlinie Elektromobilität des BMVI vom 9. Juni 2015. Stand: 22. November 2016. Retrieved from: https://www.ptj.de/lw_resource/datapool/_items/item_7476/fahrzeugherstellerliste_stand_22.11.2016.pdf [13.08.2017]

Reed, P. (2011). Electric car battery basics: capacity, charging and range. Fuel tank analogies don´t always apply. Published: 25.01.2011, Retrieved from: https://www.edmunds.com/car-technology/electric-car-battery-basics-capacity-charging-and-range.html [12.08.2017]

Schaefer, F. (2015). Fahrzeugdiagnose mit OBD.3. Auflage. Elektor-Verlag GmbH, Aachen. Retrieved from: https://www.elektor.de/fahrzeugdiagnose-mit-obd#preview [10.08.2017]