38th International Electric Vehicle Symposium and Exhibition (EVS38) Göteborg, Sweden, June 15-18, 2025

Truck data for the future design of road freight transport

Daniel Speth[1], Steffen Link[1], Patrick Plötz[1]

[1]*Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Str. 48, 76139 Karlsruhe*

daniel.speth@isi.fraunhofer.de

Executive Summary

Compared to conventional drivetrains, the competitiveness of zero emission trucks benefits from an optimal design of vehicles and infrastructure. A good understanding of vehicle use is essential for the design of batteries and the refueling or charging infrastructure. Today, only a few extensive data sets are available regarding the driving behavior of truck fleets. In this paper, we present three data sets that enable insights into European trucking behavior: (1) a dataset on European long-haul traffic, (2) a dataset on European truck stop locations, and (3) a dataset on actual truck trip data. We evaluate them regarding (1) their recency, the (2) regional and (3) temporal resolution, (4) the consecutiveness of the included data, and (5) their representativeness. Please note that this article presents a brief summary of the mentioned data sets. Detailed description of the data sets can be found in the references.

Keywords: Heavy Duty electric Vehicles & Buses, Consumer behaviour, Trends and Forecasting of e-mobility, Optimal charging locations, Modelling & Simulation

1 Motivation

Heavy-duty vehicles (HDVs) (> 12 t) are 5 % of the European vehicle fleet but cause 15 - 22 % of CO_2 emissions from road transport in 2019 [1]. Battery electric trucks (BETs), and possibly also fuel cell electric trucks (FCETs), are promising options to decarbonize road freight transport. Therefore, truck manufacturer expect half of the newly sold HDVs to be BETs und FCETs in Europe by 2030 [2]. However, the range of BETs and FCETs is typically lower than the range of diesel vehicles [2]. Technical process may reduce the difference but will also significantly increase the purchase price of the vehicles [3, 4]. A use-specific design of vehicles and infrastructure will therefore be more important in the future than it has been in the past. Data-driven vehicle and infrastructure design, particularly regarding ranges and refueling and charging infrastructure, is therefore becoming increasingly important. While a range of methodological approaches – agent-based simulation, optimization, as well as big data and artificial intelligence – are available, a corresponding data basis in the context of heavy-duty road freight transport often represents an obstacle. Although GPS or mobile phone data can provide a comprehensive database, in practice their accessibility often prevents them from being used. In this context, accessibility has multiple dimensions, including data protection, data quality, data representativity and the sheer volume of data [5].

When talking about truck data for modeling purposes, spatial and temporal resolution are of utmost importance, with GPS data as the most extensive data source [5]. From a modeler’s perspective, the consecutiveness or connection of single data points is also relevant. A distinction between (1) node-based, (2) path-based, and (3) tour-based data sets can be made. Node-based data sets contain information for a single location, for example traffic count data from a specific road section. Path-based data contains origin-destination paths, for example a truck route from one city to another. Tour-based data includes the sequence of multiple origin-destination paths, for example multiple delivery trips. The more detailed data sets are, the more complex modeling approaches can be. In general, the level of detail increases the required computational effort [6]. However, in practice, even the availability of rather basic data sets often poses a challenge, especially in a commercial context where the route planning is part of the business model.

Given the ongoing changes in freight traffic volumes [7], the recency of the data and its representativeness are also important factors, especially from an scientific perspective.

In the following, we aim to present three publicly available data sets that can provide initial insights into the usage behavior of HDVs: (1) Synthetic data on European long-haul truck traffic, (2) data on European truck stop locations, based on GPS data and routing software, and (3) actual truck trip data, representative for Germany. All datasets have been used in scientific publications in the past and provide a starting point for deeper analysis of heavy-duty road freight transport.

We will briefly introduce the data sets and give a short overview of the underlying methodology. Afterwards, we will evaluate them regarding (1) their recency, the (2) regional and (3) temporal resolution, (4) the consecutiveness of the included data, and (5) their representativeness.

2 Introduction and methodology of the investigated data sets

1.1 Synthetic data on European long-haul truck traffic

The data set describes truck traffic flows between NUTS-3 regions (*Nomenclature des unités territoriales statistiques*. Level 3: small regions for specific diagnoses) in Europe. In total, the dataset considers 1,514,573 directed transport flows between 1630 different origins (NUTS-3) and 1667 destinations (NUTS-3).

The ETIS data set from 2010 [8] contains good flows in Europe, transported by trucks, and serves as a basis for the described synthetic truck traffic flow data set. Numerous transport data tables from Eurostat, as well as national databases, were used within the ETIS project to generate a Europe-wide origin-destination matrix for transported goods between NUTS-3 regions. However, the original data set is subject to three shortcomings: (1) freight and traffic volumes have changed since 2010 [7], (2) the data set provides good flows in tons not in vehicles, and (3) there are no routes calculated. A three-step procedure updates the original data set: (1) update road freight volumes, (2) conversion from road freight transport volumes to number of trucks, and (3) routing. Figure 1 shows relevant input parameters and the steps from the original ETIS data set to the updated synthetic European truck traffic flows. A full documentation can be found in [9]. Figure 2 provides an illustration of the resulting traffic flows.

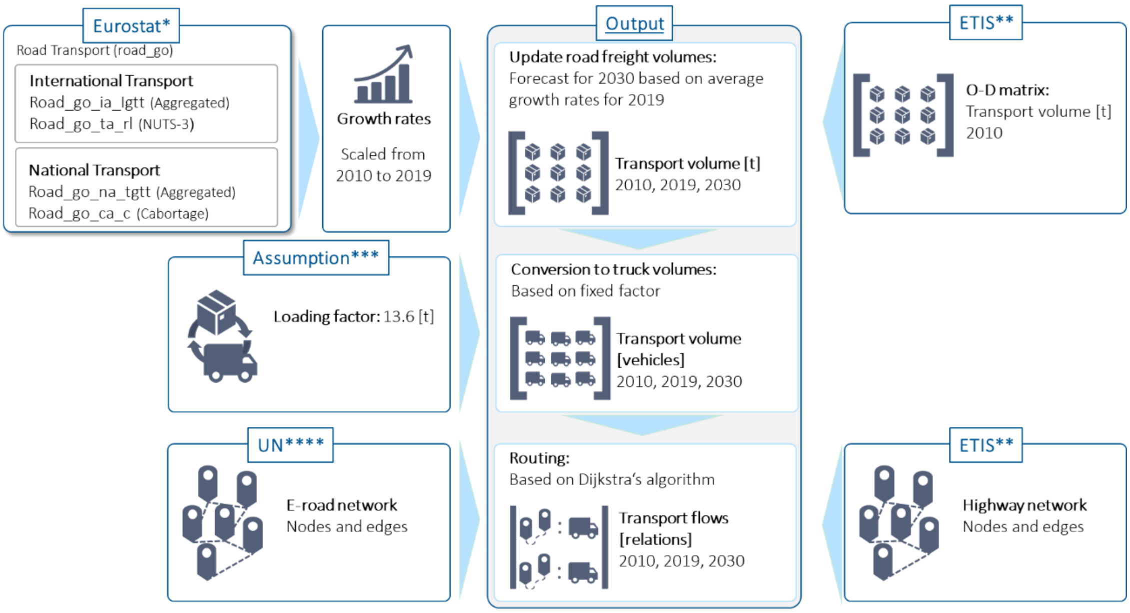


Figure 1: Illustration of the preparation of the ETIS traffic data (*[10], **[11], **[12, 13], ****[14]). Source: [15]

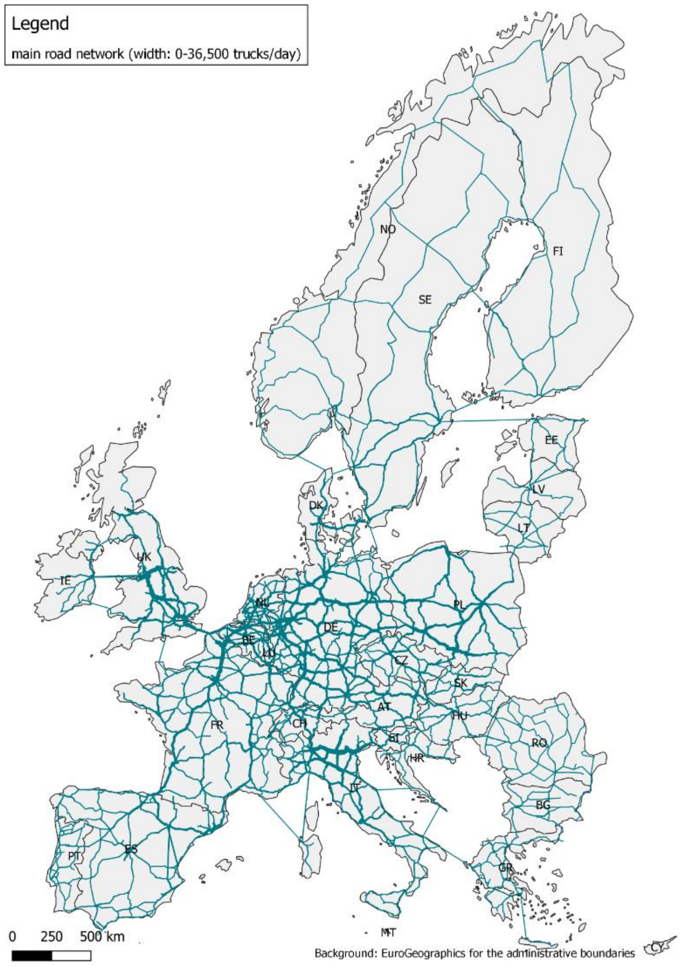


Figure 2: Modeled European truck traffic flow in 2019. Source: [15]

1.2 Data on European truck stop locations

Seven truck OEMs provided GPS data from parking locations of 400,000 trucks (> 7.5 t) in Europe. Parking locations were identified as locations where vehicles park at least 30 minutes. To limit the analysis to locations with many stops, OEMs provided only data with at least 10 stops per year per location. The locations were clustered, using DB Scan algorithm. The maximum distance to add a location to a cluster was set to 200 m. To be part of the final dataset, a cluster needed to have GPS coordinates from at least three different OEMs and at least 100 stops per year. In total, 35,000 locations in Europe were identified. However, the data can be displayed but not used for further processing [16].

Therefore, a new data set was generated, based on publicly available OpenStreetMap data. To generate the data set, a three strep procedure was used: (1) building an extensive data layer, (2) processing available data by clustering locations, and (3) enriching the resulting locations with additional information in a post-processing step. In the first step, we collected OpenStreetMap data on parking areas, rest areas, and fueling stations, using the Overpass API. The results were validated against and supplemented by commercial truck routing software (TomTom, PTV Developer, HERE Developer). After a cleaning process, 142,113 locations were identified. As a second step, the results were clustered, using the mean-shift algorithm. If the resulting clusters were closer than 250 m, they were combined to one cluster. After clustering, 51,964 locations remained. Int the post-processing step, a truck parking confidence (TPC) was applied: truck stops are marked with high TPC, rest areas acquire high or medium TPC, while fueling and parking locations receive medium or low TPC. Finally, locations are evaluated based on their default label, available area information, proximity to the TEN-T network, and area access information to confirm or update the default TPC label. The final data set contains 19,713 locations with a high or medium TPC value. All details are described in [17]. Figure 3 summarizes the most important steps to generate the data set. Figure 4 provides an overview of the identified locations.

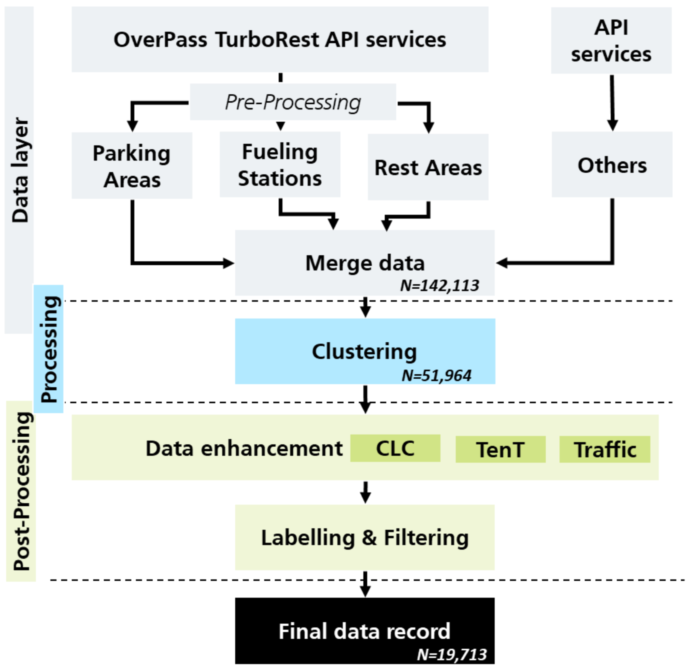


Figure 3: Data compilation process for European truck stop locations. Source: [17]

As shown in [18] for Germany, the previously described GPS data can be used as training data to calculate an attractiveness score for any of the identified European truck stop locations. This can be used, for example, to identify potential infrastructure locations with expected high demand.

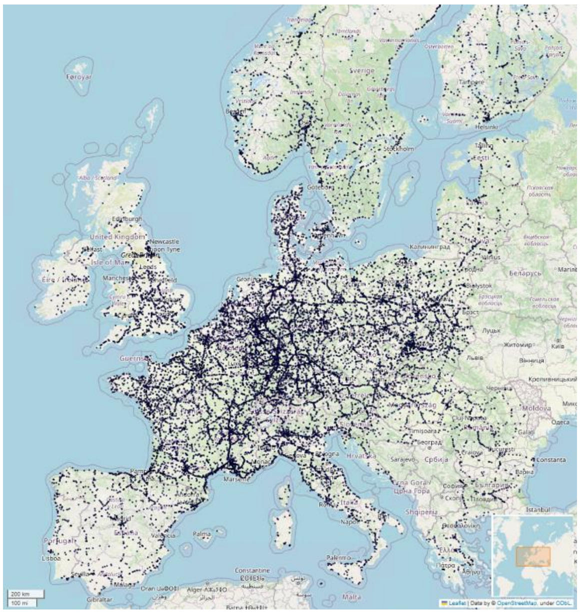
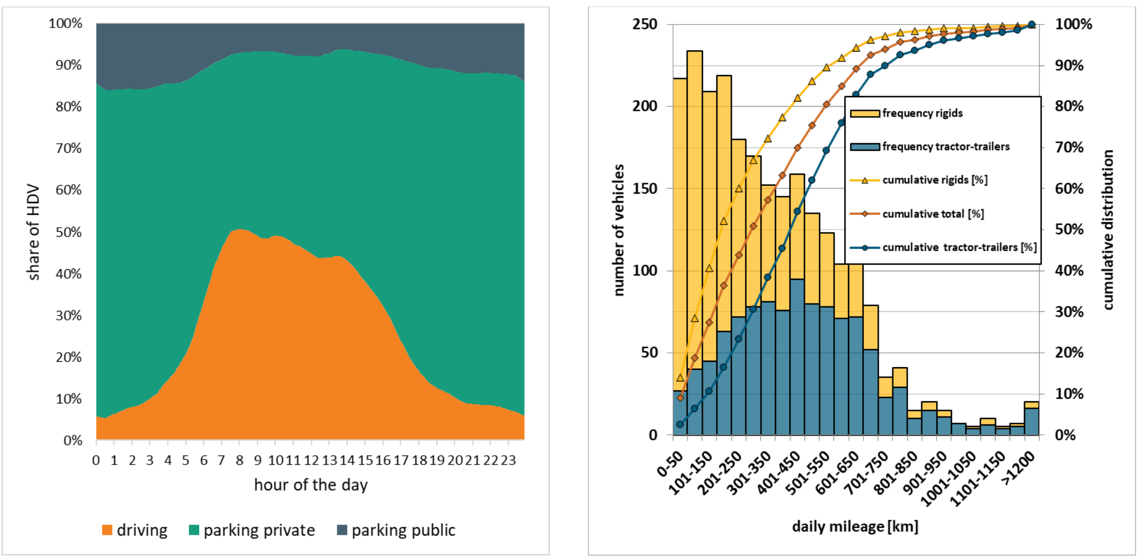


Figure 4: Identified truck parking locations in Europe. Source: [17]

1.3 German truck trip data

Representative samples of truck driving data from single vehicles are hard to obtain, due to privacy reasons. For Germany, there is a single data set (Kraftfahrzeugverkehr in Deutschland (KiD)) on the driving behavior of 2,810 diesel trucks (> 12 t) throughout a single day [19]. The data set is based on manual recording. The vehicles were selected on a representative basis from all vehicles registered in Germany. However, as some vehicles did not travel at the selected date or as some questionnaires were filled out incorrectly, 2,410 driving profiles can be used. The data set contains information on the start, the end, and the distance of the single trips of each vehicle. Additionally, it is known whether the vehicles park at public locations or at private locations. Unfortunately, there is no geospatial information. However, the dataset can be used to gain an understanding of the daily driving behavior of the fleet. Figure 5a provides an overview of the driving and parking behavior of the fleet, Figure 5b presents the distribution of the daily km traveled of the vehicles. A more detailed description can be found in [15].



(a) (b)

Figure 3: (5) Share of HDVs driving and parking. Source: [15] with data from [19]. (b) Distribution of daily km traveled from N = 2410 HDVs in the KiD sample. Source: [20] with data from [19].

3 Evaluation of the data sets

In the following, a comparison regarding (1) the recency, the (2) the regional and (3) temporal resolution, (4) the consecutiveness of the included data, and (5) the representativeness is presented.

Recency: As truck traffic constantly changes [7], up-to-date data is of utmost importance.
The synthetic European long-haul data set represents a good example in this context. While the original data set stems from 2010, the presented version in this paper contains an update, based on truck data from 2019, and a forecast for 2030.
The truck stop locations identified in our second data set are based on current research from 2023.
The third data set on German truck trips is based on a survey from 2010. Even though basic regulations, such as mandatory driving breaks, have barely changed over time, driving behavior may still have changed over time, for example due to route optimization. Therefore, critical evaluation of representativeness is highly important for this data set.

Temporal resolution: The higher the temporal resolution and the more extensive the timeframe, the more complex analyses can be. Simultaneously, with the level of detail, the computational effort also increases [5].
The European long-haul traffic data set provides annual values for truck traffic along the origin-destination paths. To obtain daily or even hourly traffic additional assumptions are needed.
The European truck stop locations themselves do not have a temporal dimension. Additional data, especially the passing traffic flow, is based on the European long-haul traffic data set and therefore also in annual resolution.
The German truck trip data contains time-discrete arrival and departure information with resolution per minute. However, the data set only covers a single day.

Regional resolution: A high regional resolution allows for a high level of detail, for example for charging infrastructure planning or driving simulation [5].
The origins and destinations in the European long-haul traffic data set are available on NUTS-3 level. Traffic within one NUTS-3 region is not part of the data set. The traffic between different NUTS-3 regions is modeled on a slightly simplified European highway network. A single road segment is typically shorter than 10 km.
The European truck stop locations contain geographical coordinates per location. However, different locations within several hundred meters are combined to single location.
The German truck trip data does not contain any regional information, beside the information whether a parking event occurs at public or private ground.

Consecutiveness: As mentioned earlier, the connection or interplay of single datapoints influences the modeling approach [6].
The European long-haul traffic data set contains path-based information. This means that origin-destination trips are known, but there is no information on the consecutiveness of single trips.
The European truck stop locations provide information for single locations – nodes – but do not interact with each other. In this sense, the data set is similar to road count data where traffic volumes are counted at specific locations in the road network.
Finally, the truck trip data set contains multiple trips of all the vehicles considered. This means one can follow the tour of a specific vehicle throughout the day.

Representativeness: To assess the transferability or scalability of results, the data set used must be classified in terms of its representativeness.
The European long-haul traffic data set was built on European statistics and should therefore be representative of European long-haul traffic. To gain a better understanding of the representativeness, the modeled data was exemplarily compared to the traffic count data in Germany. Figure 6 shows the variation. If all tours were perfectly assigned to the road network, all points would lay on the angle bisector. Deeper analyses show that deviation mainly occurs in two situations: (1) roads are parallel to each other and the algorithm strictly selects the (minimal) shorter route, while drivers choose the one with less traffic, and (2) a high share of inner-regional traffic (not part of the European ling-haul traffic data set) occurs.

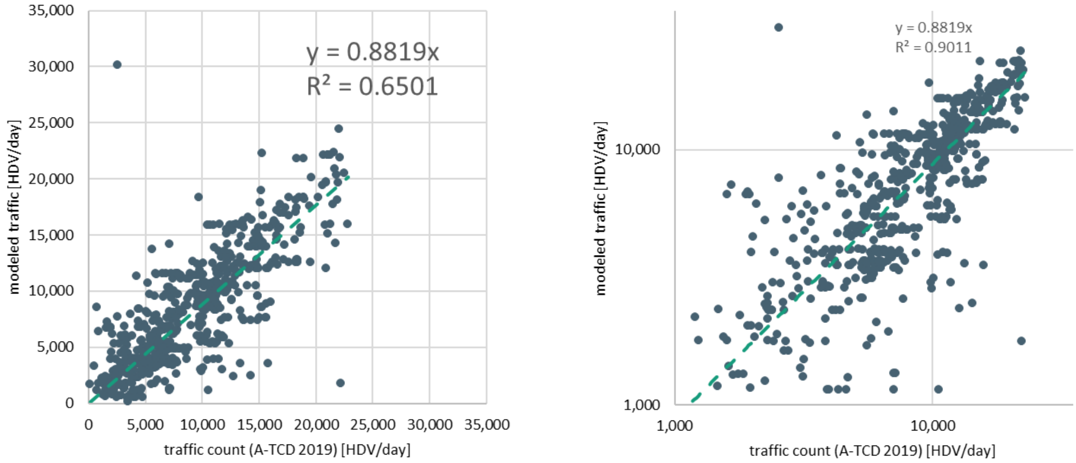


Figure 6: Comparison of modeled traffic flow volumes in the European long-haul traffic data set with actual traffic count data. Source: [15]

The European truck stop locations are based on OpenStreetMap and additional sources and should therefore cover all European countries almost equally. Additionally, locations are only part of the final dataset, if multiple sources name them as truck parking locations. In that sense, the data set represents a minimum dataset of truck parking locations.
The vehicles selected for the German truck trip data set are a representative sample of the German truck fleet in 2010. To gain a better understanding of possible deviations from today's fleet, we compared the driving and parking behavior in the data set with traffic count data from Germany. As shown in Figure 7, the data set slightly underestimates the duration of traffic on highways and rather follows the structure of traffic on federal roads, but with the typical peak of highway traffic in the early morning.

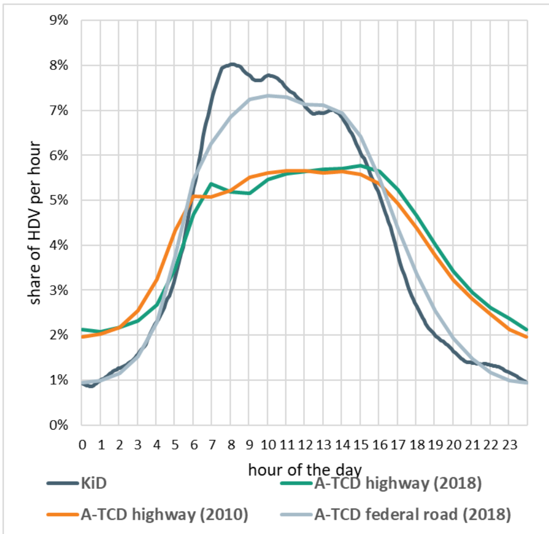


Figure 7: Proportion of hourly HDV traffic to daily HDV traffic with A-TCD as automated traffic count data on highways and federal roads and the German truck trip data as KiD. Source: [15]

All in all, the representativeness of the data sets presented seems to be sufficient, but users should be aware of the identified limitations. Table 1 summarizes the most important aspects regarding the evaluation of the three data sets.

Table 1: Overview of the evaluation of the three data sets presented

	European long-haul traffic	**European truck stop locations**	**German truck trip data**
Recency	2019 (forecast: 2030)	2023	2010
Temporal resolution	annual values	annual values	single day - minutes
Regional resolution	NUTS3, road segments	geographical coordinates	private or public locations
Consecutiveness	path-based	node-based	tour-based
Representativeness	validated against traffic count data - sufficient	intersection of multiple data sources - sufficient	validated against traffic count data - sufficient

4 Discussion and Conclusions

The data sets presented show that there is a wide range of different data available to describe truck traffic. The range includes truck traffic flow data (European long-haul traffic), truck stop location data (European truck stop locations), and truck trip data. We selected those three data sets as they are publicly available, covering good information for multiple types of research questions. However, we are also aware of other data sets that might offer similar opportunities. Nevertheless, no data set is known that covers all the topics mentioned in the evaluation section at the time. Therefore, users should always be aware of the specific strengths and weaknesses of the selected data set.

In future, it might be helpful to have a (synthetic) data set that combines both temporal resolution (e.g., based on quarter-hourly travel profiles) and spatial resolution (at least NUTS3) with consecutive trips (consecutive journeys over several days). However, it is also clear that such a data set would require considerable modeling effort.

Acknowledgments

This publication was written in the framework of the KAMO High Performance Center Profilregion, which is funded as a national High Performance Center by the Fraunhofer-Gesellschaft. The authors received funding in the Horizon Europe project MACBETH (101192466).

References

[1] Eurostat, *Data Browser: Greenhouse gas emissions by source sector (source: EEA)*. Online data code: ENV_AIR_GGE. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/ENV_AIR_GGE__custom_1533603/default/table?lang=en (accessed: Oct. 31 2022).

[2] NOW, "Market development of climate-friendly technologies in heavy-duty road freight transport in Germany and Europe: Evaluation of the 2022 cleanroom talks with truck manufacturers," Im Auftrag des Bundesministeriums für Digitales und Verkehr (BMDV), NOW GmbH, Berlin, 2023.

[3] S. Link, A. Stephan, D. Speth, and P. Plötz, "Rapidly declining costs of truck batteries and fuel cells enable large-scale road freight electrification," *Nat Energy*, vol. 9, no. 8, pp. 1032–1039, 2024, doi:

10.1038/s41560-024-01531-9.

[4] B. Nykvist and O. Olsson, "The feasibility of heavy battery electric trucks," *Joule*, vol. 5, no. 4, pp. 901–913, 2021, doi: 10.1016/j.joule.2021.03.007.

[5] W. Shoman *et al.*, "A Review of Big Data in Road Freight Transport Modeling: Gaps and Potentials," *Data Sci. Transp.*, vol. 5, no. 1, 2023, doi: 10.1007/s42421-023-00065-y.

[6] M. O. Metais, O. Jouini, Y. Perez, J. Berrada, and E. Suomalainen, "Too much or not enough? Planning electric vehicle charging infrastructure: A review of modeling options," *Renewable and Sustainable Energy Reviews*, vol. 153, p. 111719, 2022, doi: 10.1016/j.rser.2021.111719.

[7] Eurostat, *Data Browser: Summary of annual road freight transport by type of operation and type of transport (1 000 t, Mio Tkm, Mio Veh-km)*. Online data code: ROAD_GO_TA_TOTT. [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/ROAD_GO_TA_TOTT/default/table?lang=en&category=road.road_go.road_go_tot (accessed: Feb. 14 2023).

[8] E. Szimba *et al.*, "ETISplus Database Content and Methodology: ETISplus Deliverable D6," Project co-funded by the European Commission under the 7th Framework Programme, Zoetermeer., 2012.

[9] D. Speth, V. Sauter, P. Plötz, and T. Signer, "Synthetic European road freight transport flow data," *Data in brief*, vol. 40, 2022, doi: 10.1016/j.dib.2021.107786.

[10] Eurostat, *Transport Database.* [Online]. Available: https://ec.europa.eu/eurostat/web/transport/data/database (accessed: Jan. 26 2023).

[11] ETIS, *etis_2010_modelled.zip.002.* [Online]. Available: https://ftp.demis.nl/outgoing/etisplus/datadeliverables/TextFiles (accessed: Jan. 26 2023).

[12] EC, "Road Freight Transport Vademecum 2010 Report: Market trends and structure of the road haulage sector in the EU in 2010," European Commission, Brussels, 2011.

[13] Eurostat, *Empty road journeys by type of operation.* [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Empty_road_journeys_by_type_of_operation,_2018_(%25_share_in_vehicle-kilometres).png (accessed: Feb. 1 2021).

[14] UN, *European Agreement on Main International Traffic Arteries (AGR)*. Consolidated Version. [Online]. Available: https://unece.org/fileadmin/DAM/trans/doc/2016/sc1/ECE-TRANS-SC1-2016-03-Rev1e.pdf (accessed: Jul. 19 2021).

[15] D. Speth, *Electrification of road freight transport – public fast charging infrastructure and the market diffusion of battery electric trucks*: Karlsruher Institut für Technologie (KIT), 2024.

[16] P. Plötz and D. Speth, "Truck Stop Locations in Europe: Final report," Client: European Automobiles Manufacturers Association (ACEA), Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe, 2021. Accessed: Jan. 24 2023.

[17] S. Link and P. Plötz, "Geospatial truck parking locations data for Europe," *Data in brief*, vol. 54, p. 110277, 2024, doi: 10.1016/j.dib.2024.110277.

[18] S. Link, J. Auer, and P. Plötz, "Where to charge battery-electric trucks in Germany: A GIS-based statistical analysis using real-world truck data: 36th International Electric Vehicle Symposium and Exhibition (EVS36) Sacramento, California, USA, June 11-14, 2023," 2023. Accessed: Apr. 28 2025. [Online]. Available: https://publica-rest.fraunhofer.de/server/api/core/bitstreams/6a2aa09a-bc79-49e6-95f9-c9b9cc8aad49/content

[19] WVI, IVT, DLR, and KBA, "Kraftfahrzeugverkehr in Deutschland 2010 (KiD 2010): Projekt-Nr. 70.0829/2008," Datensatzbeschreibung, WVI Prof. Dr. Wermuth Verkehrsforschung und Infrastrukturplanung GmbH; Institut für angewandte Verkehrs- und Tourismusforschung e.V.; Deutsches Zentrum für Luft- und Raumfahrt - Institut für Verkehrsforschung; Kraftfahrt-Bundesamt, Braunschweig, 2012. Accessed: Mar. 10 2023.

[20] D. Speth and P. Plötz, "Depot slow charging is sufficient for most electric trucks in Germany," *Transportation Research Part D: Transport and Environment*, vol. 128, p. 104078, 2024, doi: 10.1016/j.trd.2024.104078.

Presenter Biography

Dr. Daniel Speth studied Industrial Engineering and Management at the Karlsruhe Institute of Technology (KIT). His master thesis dealt with European CO_2-legislation for passenger cars and its implications on market diffusion of alternative fuel vehicles. Since 2019, he is a researcher at the Fraunhofer Institute for Systems and Innovation Research ISI in Karlsruhe, Germany. In 2024, he completed his doctoral degree at KIT on electrification of road freight transport and the necessary public charging infrastructure for trucks. Areas of work are the modelling of market diffusion of electric vehicles with a special focus on heavy-duty vehicles, their infrastructure, and the implications on the energy system.

Steffen Link studied Industrial Engineering and Management at the Karlsruhe Institute of Technology (KIT). His master thesis dealt with European CO_2-legislation for passenger cars and its implications on market diffusion of alternative fuel vehicles. Since 2019, he is a researcher at the Fraunhofer Institute for Systems and Innovation Research ISI in Karlsruhe, Germany. Areas of work are the modelling of market diffusion for electric vehicles with a special focus on heavy-duty vehicles and its implications on the energy system.

Prof. Dr. Patrick Plötz studied Physics in Greifswald, St. Petersburg and Göttingen. Doctorate degree in Theoretical Physics from the University of Heidelberg. Since 2011 researcher in the Department Energy Technology and Energy Systems at the Fraunhofer Institute for Systems and Innovation Research ISI and since March 2020 he is coordinator of Business Unit Energy Economy. Since 2020 private lecturer at the Karlsruhe Institute of Technology (KIT) and adjunct professor for Sustainable Energy and Transport at Chalmers University of Technology.