BERGISCHE
UNIVERSITÄT
WUPPERTAL

Method Report

Truck travel times on freeway facilities in Germany in 2019 based on FCD

mFUND
Das Startkapital für die Mobilität 4.0

0   General information

The following method report describes the data processing to generate dynamic truck travel times on German highways in 2019 based on FCD. The work is done within the mFUND funded research project *ZULANA – Zuverlässigkeit des Lkw-Verkehrs auf Netzabschnitten von Bundesautobahnen* led by Bert Leerkamp[1]. For further information please feel free to contact Marian Schlott[2].

You can find the code under the following link: https://git.uni-wuppertal.de/schlott/truck-travel-times-on-freeway-facilities-and-segments-in-germany-based-on-fcd-from-2019

0.1   How to use the data

1. Network Model (stored as CSV)

The network model provides a spatial assignment of travel times, including two key identifiers:

- Network Segment ID: Differentiates freeway facilities and enables the association of travel times with the network model (a.network_segment_id = b.network_segment_id).
- Segment ID: Refers to individual segments within a freeway facility, which may consist of multiple segments. Travel times can be linked at the segment level using (a.segment_id = b.segment_id).

The network model can be accessed through GIS systems, including all versions of QGIS, as well as standard geospatial libraries such as GeoPandas or sf. It was tested on QGIS 3.38

Table 1 Metadata of the network model

Column	Description
network_segment_id	Integer Identifier for the freeway facilities
segment_id	Integer Identifier for the segments
array_osm_id	Integer[] Array storing all osm_ids on the segment.
geom	MultiLineString (WGS 4326) Stored as Well-Known-Text

2. Truck Travel Times (stored as parquets)

The truck travel times are categorized by freeway facilities and segments. Data is collected during weekday periods (Monday to Friday) in 10-minute intervals between 4:00 AM and 8:00 AM, as well as between 3:00 PM and 8:00 PM. We encourage to acces the data using duckdb and to keep in mind that the timestamps

---

[1] Further Information: https://www.gut.uni-wuppertal.de/en/
[2] schlott@uni-wuppertal.de

are converted to Europe/Berlin. The arquet files can be accessed using libraries such as PyArrow and Pandas in Python, as well as Arrow and tidyverse extensions in R. It was tested on duckdb-version v1.1.3 19864453f7 using the CLI-client on Linux.

The following table describes the Metadata for both parquet-files.

Table 2 Metadata for the parquet-files

Column	Description
network_segment_id	Integer Identifier for the freeway facilities in the parquet-file truck_traveltimes_network_segments.parquet
segment_id	Integer Identifier for the segments in the parquet-fle truck_traveltimes_segments.parquet
timestamp	Timestamp with time zone (Europe/Berlin)
truck_traveltime_minutes	double dynamic travel times of trucks

The following screenshot depicts an example how to access the data using duckdb[3].

Figure 1 Example for accessing the truck travel times

1   Methods

1.1   Network model

The network model, obtained from the Geofabrik for the year 2019, has been segmented to differentiate 3 aggregate levels. The smallest aggregate level is the 100 m-edge which is the basis for the network assignment algorithm; so the 100 m-edge serves as the aggregation level for the FCD to secure a certain

---

[3] https://duckdb.org/

spatial resolution and plays also a crucial role for the interpolation algorithm conducted; The concept is taken from (Holthaus, o. J.).

In general freeway facilities comprise basic, diverge, merge and weaving segments or intersections. (Forschungsgesellschaft für Straßen- und Verkehrswesen, 2015; Lemke, 2016) To enable the analysis of travel times at both segment and freeway facility levels, an algorithm was developed to segment the network model. This process required the classification of various motorway elements, including basic, diverging, merging, and weaving segments, as well as intersections. Based on this classification, 100-meter-edges were aggregated till they intersected with elements classified as basic, diverging, merging, or weaving segments. The freeway facilities were derived from the planning concept of the federal agency of Germany. (Bundesministerium für Verkehr und Digitales, 2018) The following figure depicts the distinction between freeway facilities and segments. (Geistefeldt, 2015)
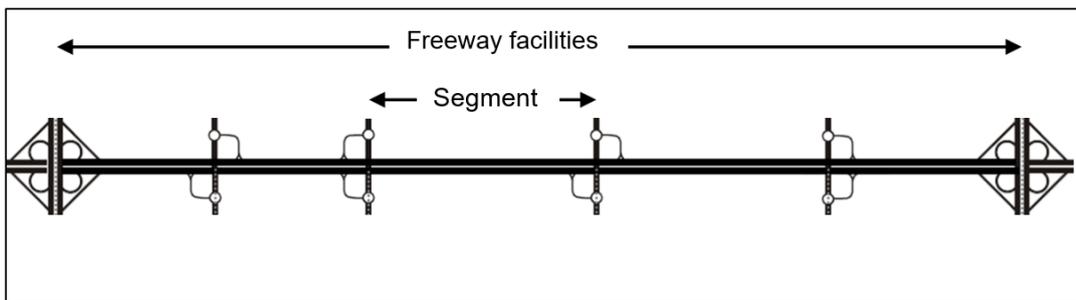

Figure 2 Distinction between freeway facilities and segments based on (Geistefeldt, 2015)

1.2   Floating Car Data

In general, Floating Car Data (FCD) is generated by on-board units or navigation systems in vehicles and provides information on vehicle velocity over time (timestamp) and space (geocoordinates). Data providers typically offer either 'raw' FCD or aggregated FCD, such as percentiles of velocities over specific network elements. 'Raw' FCD, in this context, refers to data that can either track an individual vehicle's movement over time or be aggregated over specific network elements to provide an overview of traffic conditions. We utilize 'raw' FCD data for the year 2019, provided by ADAC . The ADAC collects FCD from various providers to monitor traffic conditions through their own systems. We do not have specific information on the providers from which the ADAC collects the data. Our sample comprises more than 25 billion data points. The number of FCD hits varies between months; approximately 2 billion FCD hits are recorded each month, with notably higher numbers in the summer months, which can be attributed to holiday traffic. The hourly distribution of FCD hits is relatively consistent across months and days, suggesting that the composition of the FCD sample is primarily influenced by fleets. Checking the daily distribution of the sample we can observe the formation of a plateau which is more common in truck traffic unlike to the observed daily patterns in passenger traffic, with two peaks during the morning and evening rush hours. In addition to the number of FCD hits and their hourly distribution, the examination of individual vehicles is of interest. For anonymization reasons, a vehicle can only be tracked in the FCD sample for 24 hours. Even in this context, the distribution of the number of vehicle IDs remains relatively constant, with approximately 15 million vehicle IDs. Given that we are using a raw FCD sample, it is essential to perform a network assignment and vehicle classification.

1.2.1 Network Assignment

To compute travel times, it is necessary to assign the FCD to the network model. For this study, the map matching service of the OpenSourceRoutingMachine (OSRM) is used.(Luxen and Vetter, 2011) OSRM is an open-source router offering various services based on OSM data, utilizing a Hidden Markov Model (HMM) for map matching. HMMs determine the probability of reaching a state B from an initial state A. In map matching, HMMs evaluate the likelihood that a sequence of road elements forms the actual route of a vehicle trajectory. The states in the HMM are network elements, while the state observations are FCD points dispersed spatially and temporally. The goal of the HMM is to assign the FCD to the correct network element with high probability. (plausibility).(Newson and Krumm, 2009)

Due to the high cost of map matching services, an OSRM instance is initiated in a Docker environment. The required inputs include GPS coordinates, timestamps, direction, and a maximum search radius per vehicle, set at 25 meters due to the computational intensity of network assignment. OSRM returns the corresponding OSM node, and its methodology prevents "noisy" FCD hits from being assigned to a network edge, reducing the need for extensive plausibility checks. For instance, the algorithm verifies if the transmitted velocity is realistic for the specific road class. To facilitate small-scale assignment and data aggregation, each plausible FCD data point is assigned to a 100m edge. Following the network assignment, there are approximately 19,995 FCD hits per kilometre of freeway facility per month.

1.2.2 Vehicle classification

Based on cluster analysis, FCD is classified into vehicle types. A vehicle database is prepared, capturing distribution parameters (minimum, maximum, median, and average velocity) for each vehicle on different road types (e.g., highways). A k-means clustering algorithm is applied, which partitions the data into a predefined number of clusters, k. Each cluster contains at least one object, and each object belongs to exactly one cluster. The goal is to minimize the sum of squared deviations from the cluster centroids. Initially, k centroids are randomly placed within the data points. The centroids are then recalculated as the mean of all members within each cluster, and data points are reassigned. This process repeats until no changes occur, reaching an optimal solution.(Jin and Han)

We used three initial clusters for the k-means algorithm. The distribution of mean, median, and maximum velocities on freeways is depicted in the next figure. Figure 2 illustrates the velocity distributions of clustered vehicles (cars and trucks) on German freeways. The car cluster exhibits a broader distribution with higher average and maximum velocities, reflecting typical passenger vehicle behaviour, while the truck cluster is tightly centred around lower velocities (approximately 80–100 km/h), consistent with speed limits and operational constraints. Minimal overlap between clusters suggests effective differentiation, though some ambiguity may exist in the overlapping range. The plotted mean, median, and maximum velocities align well with observed distributions, supporting the clustering approach's plausibility.

Velocities of clustered vehicles on highways in Germany
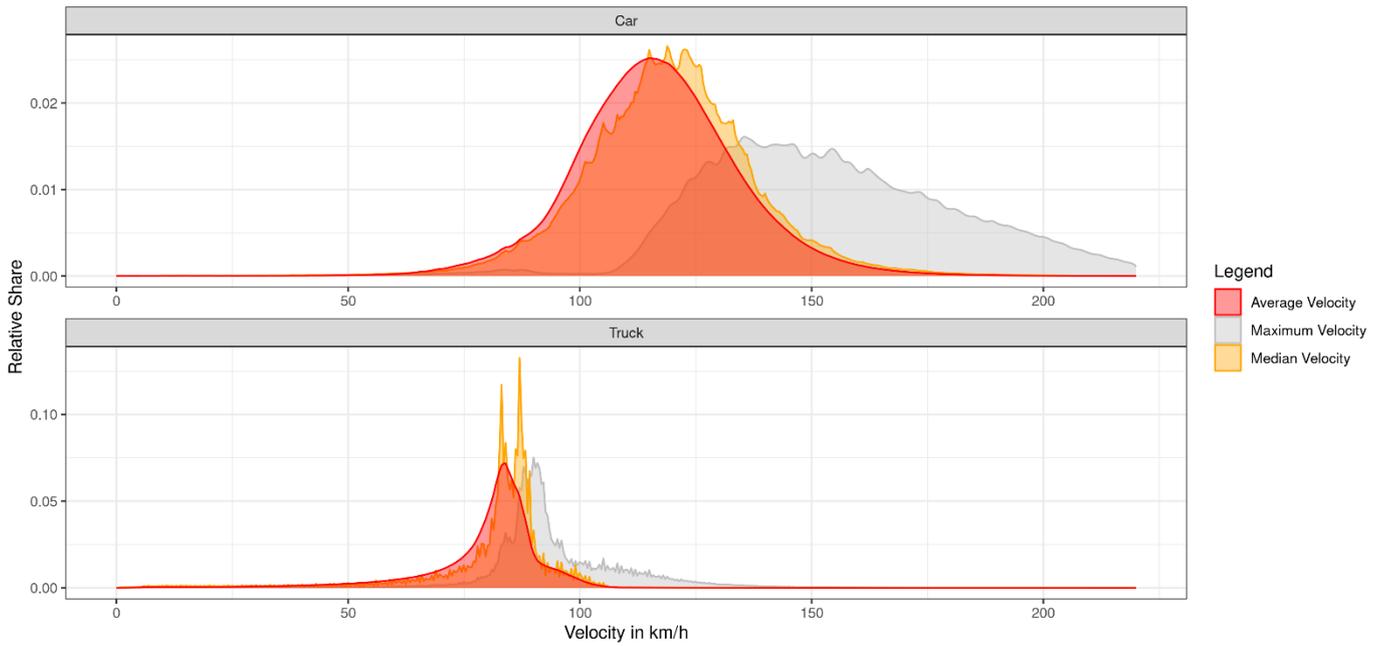Derived using a k-means approach based on FCD
X-values are clipped at 220

Figure 3 Velocities of clustered vehicles on highways in Germany

1.3   Dynamic travel times

To compute travel times on segments and freeway facilities that realistically reflect actual velocities (ground truth), it is essential to determine the number of FCD hits required within a specific time interval to reliably estimate the velocity. The first step in this process involves evaluating the penetration rate of the FCD sample at counting stations operated by the Federal Highway Research Institute (BASt) over an entire year. The counting stations of the federal highway network are semi-automatically assigned to the road network model. Highway directionality is used to distinguish outbound and return directions, enabling accurate assignment of counting stations to the network. On average, penetration rates of approximately 5.7 % for trucks and 2.3 % for passenger cars were derived by comparing the hourly traffic volumes recorded at the counting stations with the FCD-based traffic data. For hourly evaluations, the penetration rate is sufficient. A penetration rate of 2 % is adequate to reliably provide information about traffic conditions. (Vortisch and Friedrich, 2022)

However, hourly travel time distributions are too "coarse" to accurately capture traffic conditions. In the literature it is recommended to aggregate travel times or velocities within time intervals of no more than 15 minutes for traffic condition estimation using FCD. (Peter et al., 2021) To evaluate penetration rates for different time intervals, we analyzed data from two counting stations in North Rhine-Westphalia (NRW) for which traffic data (vehicle volumes and velocities for cars and trucks) were available at 1-minute intervals over an entire year. For the traffic station MQ_46.14a_HFB_SW, the average monthly penetration rate ranges between 1.9% and 2.5%. In contrast, at the traffic station MQ_46.16a_HFB_NO, the average monthly penetration rate spans between 4.4% and 7.3%. To estimate the accuracy of velocities derived from FCD, we compare FCD-based velocities with those measured at traffic stations for various time intervals and numbers of hits. To directly relate the analysis to the measured velocity data, we apply the following methodology outlined in (Lelke et al., 2024). This approach utilizes the SQV (Standard Quality Value) metric to evaluate the quality of velocities derived from different numbers of vehicle trajectories.

The SQV is calculated as:

$$g_{SQV} = \frac{1}{1+\sqrt{\frac{(m-c)^2}{f*c}}}$$

where m represents the measured velocity, c is the reference velocity, and f is a scaling factor. (Friedrich et al., 2019b) The scaling factor f is determined based on the typical magnitudes of the evaluated traffic conditions, taking into account velocities of 80 km/h for trucks.

The SQV metric measures the quality of the velocity estimates derived from FCD compared to ground truth data (traffic station velocities). A higher SQV indicates better agreement, with an SQV of 0.9 or above considered a "very good match". (Friedrich et al., 2019b; Friedrich et al., 2019a) We aggregated the velocity measured on traffic stations by computing the harmonic mean for different time levels. For all aggregation levels (5, 10, and 15 minutes), the SQV improves with an increasing number of hits. The average SQV consistently exceeds the 0.9 threshold across all levels, indicating a strong match between FCD-derived and ground truth velocities. Both traffic stations (MQ_46.14a_HFB_SW and MQ_46.16a_HFB_NO) show similar trends, with some variability. Notably, for lower hit counts, SQV at MQ_46.16a_HFB_NO tends to be slightly higher than at MQ_46.14a_HFB_SW, suggesting marginally better performance at the former station. Longer aggregation intervals slightly enhance SQV but may sacrifice temporal resolution. For instance, a 15-minute aggregation yields higher SQV but may be less responsive to short-term traffic fluctuations. Overall, the figure 3 demonstrates that an SQV value of 0.9 or higher is achievable with adequate FCD hits, even with shorter intervals like 5 minutes. These findings validate the use of FCD for velocity estimation, particularly when penetration rates are sufficient.
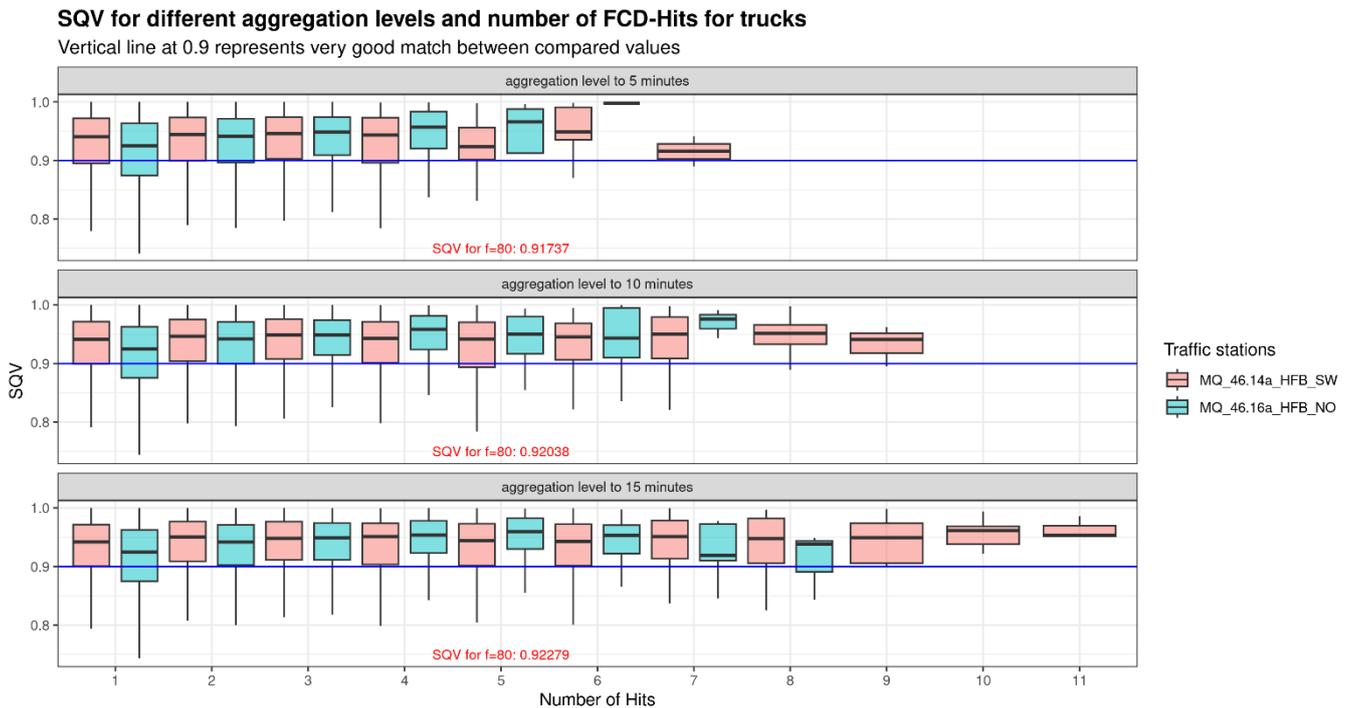

Figure 4 SQV for different aggregation levels and number of FCD-Hits for trucks

The analysis indicates that a 10-minute aggregation level offers a high temporal resolution for assessing truck travel time reliability using the available FCD sample. Given the spatial and temporal variability in FCD penetration, interpolation is necessary to derive a continuous velocity function for specific time intervals and spatial resolutions. The sample includes approximately 19 truck FCD hits per kilometre of freeway per hour, which is insufficient to detect traffic disruptions and their temporal and spatial

fluctuations. Therefore, the Adaptive Smoothing Method (ASM) by (Treiber and Helbing, 2002; Treiber and Kesting, 2010) is applied. Spatiotemporal interpolation algorithms provide continuous average velocities as functions of space and time, derived from traffic stations. Since the FCD sample is assigned to the network model, they can be spatially and temporally aggregated, making the methodology applicable to the available data.(Treiber and Helbing, 2002; Treiber and Kesting, 2010)

Figure 4 illustrates the ASM-method results, showing a central region from 10:00 to 17:00 with a significant red zone indicating reduced truck velocity due to congestion or other factors, suggesting a peak traffic impact during this period.
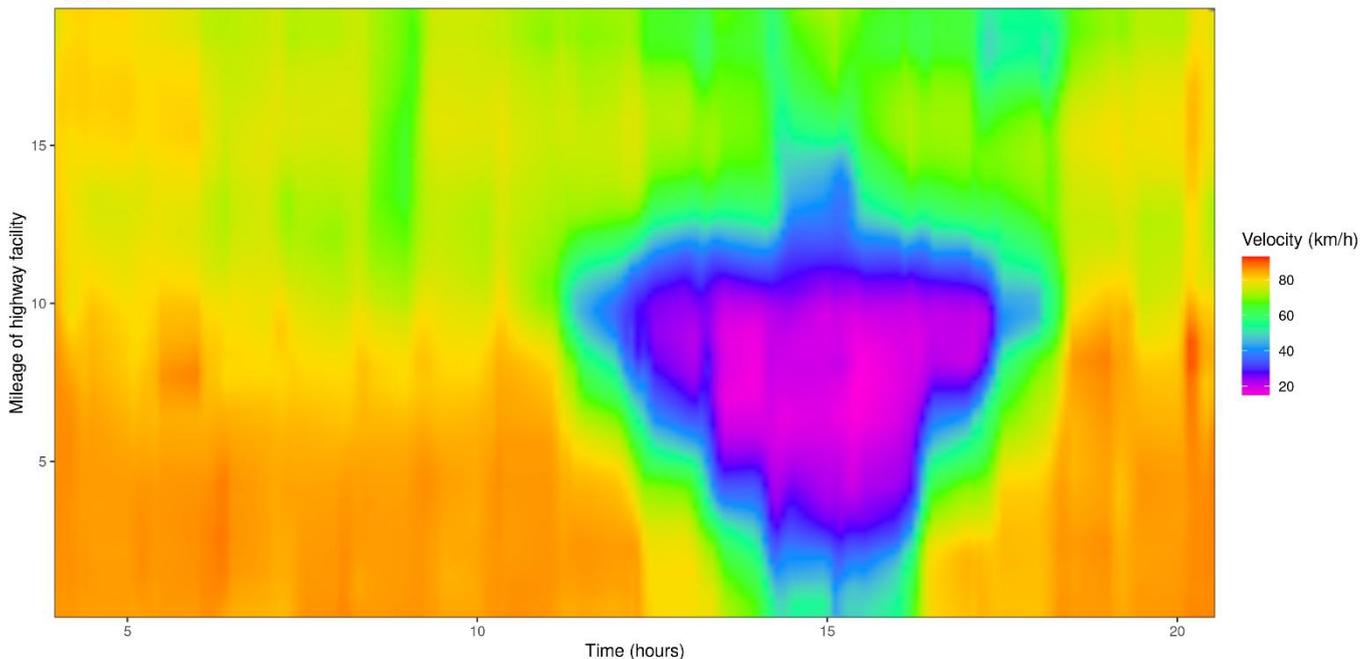

Figure 5 Based on ASM-method derived continuous velocity fucntion for trucks on highway facilities

To calculate travel times dynamically, we developed the following method:

As the path of a segment or freeway facility is predefined , we start at the first 100m-edge of a segment or freeway facility and end at the last 100m-edge, along with a particular start time—for example, 08:00 hours (this can be adjusted based on requirements). Starting from the initial time, as the segments and freeway facilities are divided into 100m-edges we calculate the travel time for each 100m-edge at the starting time, keeping track of the cumulative travel time spent as we move from one 100m-edge to the next. This process continues until the total time reaches 10 minutes (assuming the last 100m-edge has not yet been reached). At this point, we update the velocity to reflect the conditions at 08:10 hours and repeat the process for the next 10-minute interval. This cycle of updating the velocity every 10 minutes and calculating travel times for each segment continues until the last element is reached. The 10-minute interval can also be adjusted based on the desired level of detail. Finally, we have a plot illustrating how travel time is affected or increased based on different start times considered in the method.

Derived synthesized trajectories from ASM-method for trucks on highway facilites
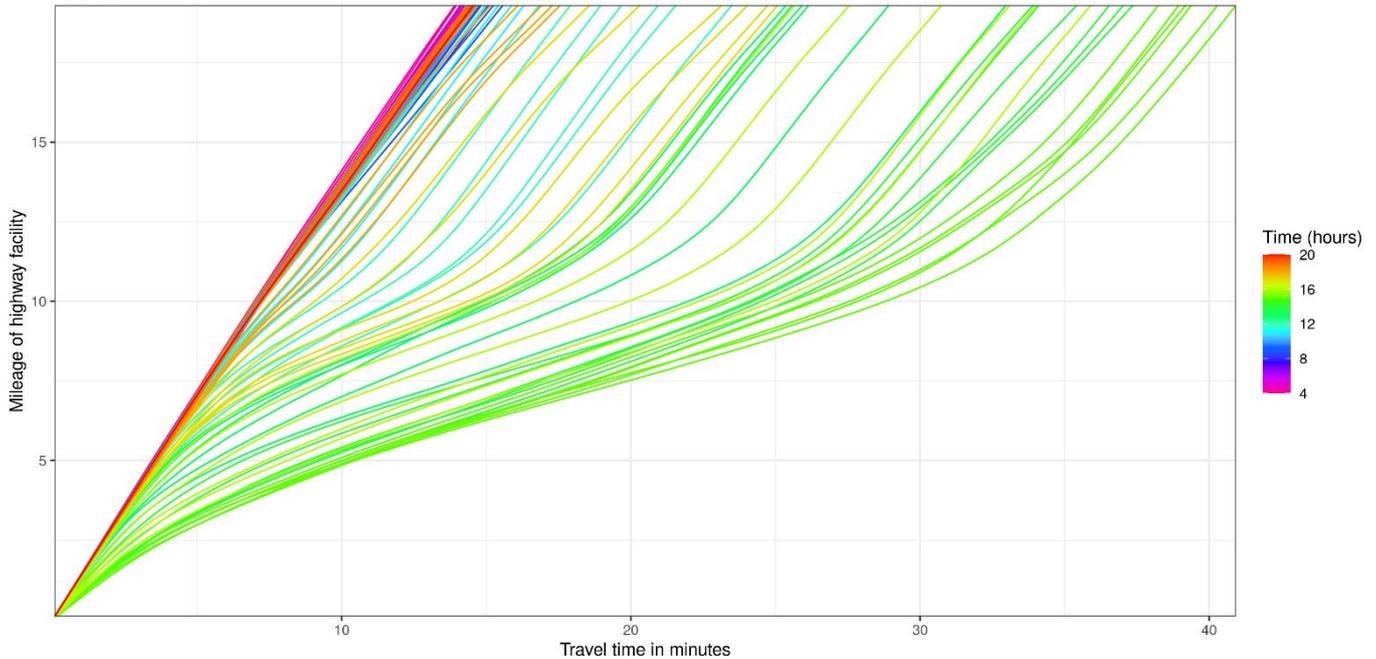Aggregated to 10 Minutes interval

Figure 6 Derived synthesized trajectories from ASM-method for trucks on highway facilities

2 Validation

We conducted a validation at each step of the data processing workflow (cp. to the sections). To validate the dynamic travel times we compare them against empirically measured travel times on section level. To derive empirically measured travel times, the FCD sample must first be segmented into individual trips per vehicle. This segmentation is carried out using a threshold-based approach, which defines the boundaries of a trip. Since this study focuses solely on travel times at the level of sections, extensive segmentation is not necessary. Only "section-specific trips" need to be segmented, meaning that when deriving travel times at the level of sections, it must be ensured that the derived travel times belong to a single trip. Consequently, FCD hits from a vehicle are assigned to the same trip as long as the time interval between consecutive FCD hits does not exceed 10 minutes. If this threshold is exceeded, the trip is considered to have ended. All vehicles that transmit an FCD hit within a trip on both the first and last section element are identified. The travel time of a vehicle on a specific section of a trip is calculated by subtracting the timestamp of the first hit from that of the last hit. However, the average length of the first and last elements of a section is approximately 85.5 m, which must be considered. In some cases, the first hit may be recorded near the end of the first element, or the last hit may occur near the beginning of the last element. After that we have a sample of empirically measured travel times on sections. To compare both samples we compute the travel time share (dynamically travel times to empirically measured travel times) and the travel time difference. The distribution of the comparison between dynamically computed travel times and empirically measured travel times for trucks on German highway sections are depicted in the following figure. The histogram(figure 6) exhibits a highly skewed distribution, with the majority of the data concentrated near 1, indicating minimal deviation for most comparisons. A long tail extends towards larger comparison values, suggesting a smaller number of cases with significant deviations. he median deviation (1.03) suggests that dynamically computed travel times are close to empirically measured travel times for most cases as well as relatively low median difference (3.94 seconds) indicates good agreement between

the two measures for travel times. The long tail in the distribution and the relatively high average difference (15.79 seconds compared to the median of 3.94 seconds) highlight the presence of outliers or cases with substantial deviations. The larger average values suggest that a subset of long sections with significant deviations is skewing the results, indicating that section length might play a key role. Nevertheless the low median values demonstrate that the method for estimating travel times dynamically is reliable for the majority of cases.
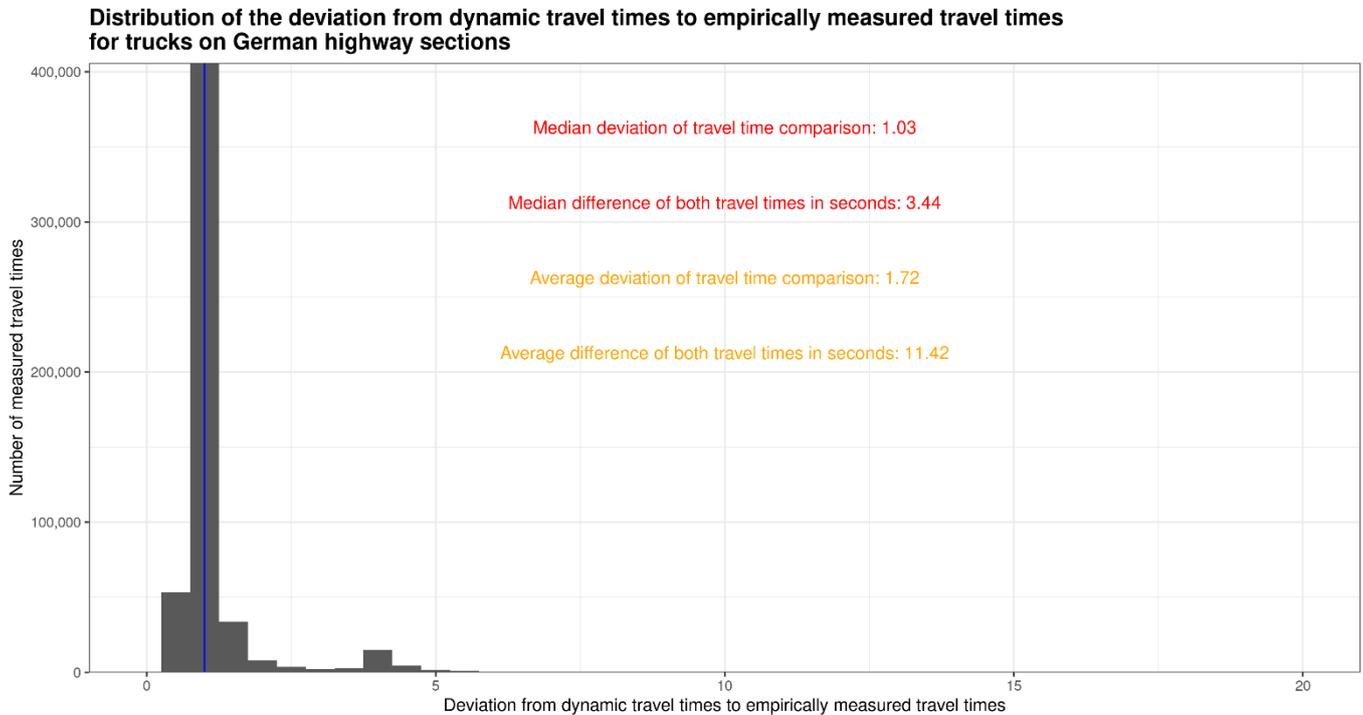

Figure 7 Distribution of the deviation from dynamic travel times to empirically measured travel times for trucks on German highway sections

3   Acknowledgements

4   Author contributions

The authors confirm their contributions to the method report as follows: M. Schlott and B. Leerkamp conceived and designed the study. M. Schlott assigned the FCD to the network model, classified the vehicles and implemented with L. Abdul the ASM-method as well as the computation of dynamic travel times, Rafay A. supported throughout the data processing.

5   Literature

Bundesministerium für Verkehr und Digitales, 2018. Verbindungsfunktionsstufen 0/1 im Zielnetz der Bundesfernstraßen (BPl 2016, VB/WB*). Bundesministerium für Verkehr und Digitales. https://bmdv.bund.de/SharedDocs/DE/Artikel/StB/Verbindungsfunktionsstufen-0-1.html. Accessed 9 December 2024.

Forschungsgesellschaft für Straßen- und Verkehrswesen, 2015. Handbuch für die Bemessung von Straßenverkehrsanlagen: HBS 2015. FGSV W1 - Wissensdokumente FGSV 299 B, Köln, CD-ROM.

Friedrich, M., Pestel, E., Heidl, U., Pillat, J., Schiller, C., Simon, R., 2019a. Anforderungen an städtische Verkehrsnachfragemodelle.

Friedrich, M., Pestel, E., Schiller, C., Simon, R., 2019b. Scalable GEH: A Quality Measure for Comparing Observed and Modeled Single Values in a Travel Demand Model Validation. Transportation Research Record: Journal of the Transportation Research Board 2673, 722–732.

Geistefeldt, J., 2015. Verkehrsqualität auf Autobahnen. Forschungsgesellschaft für Straßen- und Verkehrswesen. HBS-Einführungskolloqium, 2015.

Holthaus, T., o. J. Open Source gestützte Anwendung der Richtlinien für integrierte Netzgestaltung im ländlichen Raum: Unveröffentliches Manuskript. Dissertation, Wuppertal.

Jin, X., Han, J. K-Means Clustering. In: pp. 695–697.

Lelke, T., Fuchs, L., Friedrich, B., Vortisch, P., 2024. Assessing the Accuracy of Average Travel Speeds from Floating Car Data on Urban Street Facilities. Transportation Research Procedia 78, 570–577.

Lemke, K., 2016. The New German Highway Capacity Manual (HBS 2015). Transportation Research Procedia 15, 26–35.

Luxen, D., Vetter, C., 2011. Real-time routing with OpenStreetMap data, in: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. GIS '11: 19th SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago Illinois. 01 11 2011 04 11 2011. ACM, New York, NY, USA, pp. 513–516.

Newson, P., Krumm, J., 2009. Hidden Markov map matching through noise and sparseness, in: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. GIS '09: 17th SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle Washington. 04 11 2009 06 11 2009. ACM Digital Library, New York, NY, USA, pp. 336–343.

Peter, L., Janko, J., Schick, N., Waßmuth, V., Friedrich, M., Bawidamann, J. (Eds.), 2021. Entwicklung eines aktuellen, echtzeit-verfügbaren Key Performance Indicator (KPI) Systems für das deutsche Autobahnnetz: = Development of an up-to-date, real-time available Key Performance Indicator (KPI) system for the German motorway network. Fachverlag NW in der Carl Schünemann Verlag GmbH, Bremen, 151 pp.

Treiber, M., Helbing, D., 2002. Reconstructing the Spatio-Temporal Traffic Dynamics from Stationary Detector Data. Cooperative Transportation Dynamics 1.

Treiber, M., Kesting, A., 2010. Datengestützte Analyse der Stauentstehung und -ausbreitung auf Autobahnen. Straßenverkehrstechnik, 5–11.

Vortisch, P., Friedrich, B., 2022. Verkehrsdatenerfassung 2.0 - Nutzung von FCD zur Bewertung der Verkehrsqualität von innerstädtischen Hauptverkehrsstraßen. Deutscher Straßen- und Verkehrskongress, 2022, Dortmund.