Supplement of
DRIVE v1.0: A data-driven framework to estimate road transport emissions and temporal profiles

Daniel Kühbacher et. al.

Correspondence: Daniel Kühbacher (daniel.kuehbacher@tum.de) and Jia Chen (jia.chen@tum.de)

1 Implementation and Application

This supplement contains extended model and implementation descriptions tailored to the application of the DRIVE (Data-driven Road-Transport Inventory for Vehicle Emissions) framework. It complements the README documents in the code repository to serve as a user manual.

The framework is developed in Python and targets users with programming skills. For all scripts that require user interaction, Jupyter Notebooks[1] were used, which combine code, visualization, and documentation in one file. All notebooks are set up in the same way, starting with a description of the content, the import of libraries, the user-defined processing settings, data imports, dedicated notebook functions, and all downstream processing steps. These notebooks access a custom utils module that implements functions for calculating cold and hot emissions, processing the traffic counts, accessing calendar functions, and rasterizing the line source emissions.

Additionally, QGIS [2], a free and open-source geographic information system, is used for manual data curation (e.g., assigning the locations of traffic counting stations to their respective road links) and visualization.

1.1 Workflow Overview

The framework is intended to be used according to the following workflow:

1. Data curation: Preparation of input files before automatic processing is conducted.
2. Data pre-processing: Combine data with information from different sources and clean the dataset.
3. Model setup and run: Define model parameters and run emission calculation.
4. Processing of the model output: Rasterize the line source emissions and visualize the data.

[1]https://jupyter.org/

[2]https://www.qgis.org/

2 Data Curation

Before automatic data processing can occur, the user needs to prepare the datasets and convert them to a machine-readable format. The required steps depend entirely on the availability of the data and its structure. This section describes the procedures that were carried out in Munich.

- Allocate traffic counting stations to the respective road links in the traffic model. Where and how these detectors are allocated depends on the representation of the road links in the traffic model (e.g., different driving directions could be represented as a single road link or as separate road links). Figure S1 shows how a single counting station could include several detectors (e.g., multiple lanes on a cross-section). Subsequently, the count data of multiple detectors on a single road link is aggregated. We performed this step manually for all detectors that can be assigned to road links in the traffic model. Several detectors were excluded near the trade fair (Messe München) and Munich's football stadium (Allianz Arena). They are solely used for traffic control during special events.

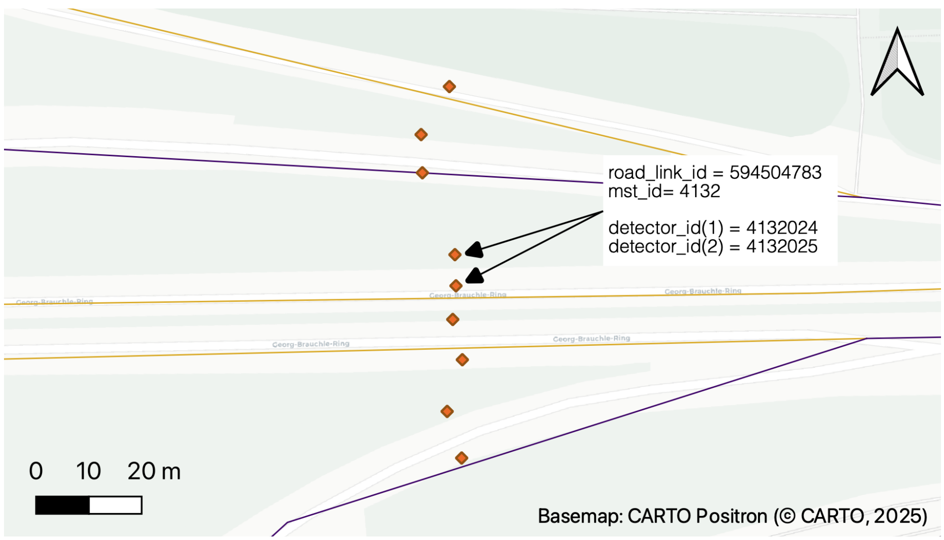

Figure S1. Example of four two-lane roads and one single-lane road, each equipped with individual detectors identified with a detector_id that form a counting station (mst_id) together. These multiple detectors must be correctly assigned to their corresponding road links (road_link_id) in the traffic model.

- Convert the traffic counting data to a standard data format for subsequent automatic processing.
- Convert the information from the traffic model to an HBEFA-compatible format (e.g., speed, gradient, road type).
- Create an Excel calendar file that lists local public holidays and vacation periods.
- Export the emission factors as tabular *.csv data from the HBEFA MS-Access application. Cold-start excess emissions and hot exhaust emission factors are available in separate tables.

3 Optimization of the Volume-Capacity Ratio Thresholds

The volume-capacity ratio (VCR) was utilized to estimate the traffic condition on each road link. Schmaus et. al. (2023) investigated the distribution of the vehicle kilometers traveled (VKT) among different traffic conditions in Germany based on floating car data. Table S1 shows the distribution for different road types in urban areas. These shares were used as reference for the optimization process and are suitable for urban roads in Germany.

The optimization process requires manual parameterization of the VCR thresholds. A script is provided that calculates the VKT distribution across different traffic conditions on a given road type. By iteratively changing the VCR thresholds, the target distribution is achieved. Table S2 shows the resulting optimized VCR thresholds for different road types.

Table S1. Share of vehicle kilometers traveled (VKT) in different HBEFA traffic conditions for different road types in German urban areas. The shares were retrieved form Schmaus et. al. (2023)

	Freeflow	*Heavy*	*Saturated*	*Stop&Go*	*Stop&Go 2*
Motorway	38 %	31 %	28 %	2,7 %	0,3 %
Primary-National	55 %	27 %	12 %	2,3 %	4,2 %
Primary-City	57 %	25 %	11 %	3,2 %	4,7 %
Secondary	54 %	23 %	14 %	5 %	3,6 %
Residential	55 %	23 %	13 %	4,7 %	3,4 %

Table S2. Optimized VCR-thresholds as applied in Munich to achieve the targeted distribution of VKT in different traffic conditions.

Road Type	*Freeflow*	*Heavy*	*Saturated*	*Stop&Go*	*Stop&Go 2*
Reference (HBS , 2015)	0.55	0.9	1	>1	-
Motorway National	0.5	0.71	0.98	1.1	> 1.1
Primary-National	0.33	0.5	0.7	1	> 1
Primary-City	0.67	0.82	0.92	1.02	> 1.02
Distributor/Secondary	0.37	0.5	0.63	0.8	> 0.8
Access/ Residential	0.122	0.25	0.38	0.5	> 0.5

4 Cross Validation of the Traffic Model and Counting Data

The VISUM model used in Munich represents the average annual norm-weekday (Tuesday-Thursday outside the holiday season) traffic volume $q_{i,vc}^{model}$ for passenger cars (PC), light commercial vehicles (LCV), heavy goods vehicles (HGV). To compare it with the counting data and evaluate the fit, we calculated the average weekday count from 2019. This allows for estimating the modeled and observed values for multiple road links, enabling the computation of the coefficient of determination (R^2) and the scalable quality value (SQV).

The coefficient of determination indicates how effectively the model represents the traffic volumes across the city area and different road types. It also reveals systematic model errors like an over- or underestimation of traffic volumes on particular road types. Figure S2 shows scatter plots for the total traffic volume (SUM) and all vehicle-specific traffic volumes available in the traffic model (PC, LCV, and HGV). The LCV category shows higher values in the model than the counting stations. Upon request, the city department responsible argued that no LCV-specific calibration had been performed. The manual counts used for calibrating the traffic model do not distinguish between PC and LCV. This likely causes the model-data mismatch.

Secondly, the Scalable Quality Value (Equation S1), a quality measure used in traffic engineering, provides an additional, magnitude-independent criterion for traffic model assessment. Proposed by Friedrich et. al. (2019), the SQV introduces a scaling factor f to assess traffic volumes at different time aggregations (hourly traffic volume: $f = 10^3$, daily traffic volume: $f = 10^4$). M is the modeled traffic volume and C the counted volume. Furthermore, he proposes SQV thresholds for perfect matches ($SQV \geq 0.9$) to acceptable matches ($SQV \geq 0.8$). Given the extensive size of the road network and the fact that data from these counting stations was not utilized in optimizing the traffic model, a lower threshold was justified. We use a $SQV \geq 0.6$ to select trustful counting stations. Figure S3 shows a histogram of the SQV statistics of all counting stations in our area of interest. 57.1% of all counting stations on different road types show a $SQV \geq 0.9$ while 90.5% of all stations have a $SQV \geq 0.6$.

$$SQV = \frac{1}{1 + \sqrt{\frac{(M-C)^2}{f \cdot C}}} \tag{S1}$$

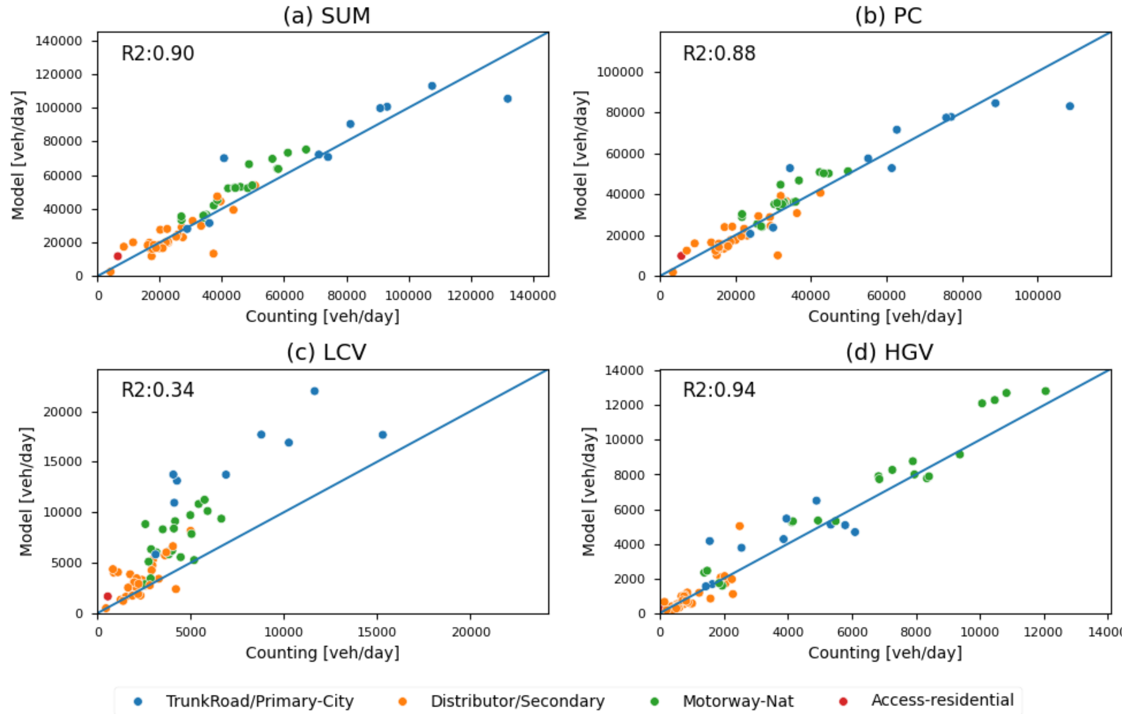

Figure S2. Scatter plot of the average annual norm-weekday traffic volume in the traffic model and calculated from the traffic counting data. The total traffic (SUM), passenger cars (PC), and heavy goods vehicles (HGV) show a good fit for all road types. Light cargo vehicle volume (LCV) is not well calibrated in the model.

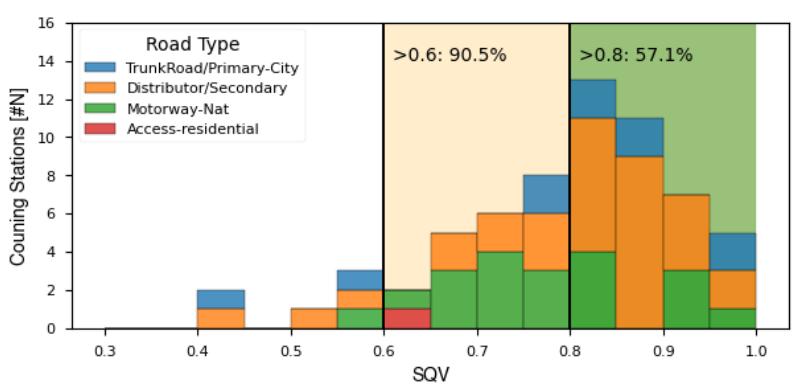

Figure S3. Histogram of the scalable quality value (SQV) of the average annual norm-weekday traffic volume in the traffic model and calculated from the traffic counting data. 57% of all counting stations have an acceptable fit according to the threshold provided by Friedrich et. al. (2019) ($SQV \geq 0.8$). We relaxed the criterion and used it to select trustful counting stations with an $SQV \geq 0.6$, which includes $> 90\%$ of all stations in our region of interest.

References

Friedrich, M., Pestel, E., Schiller, C., & Simon, R. (2019). Scalable GEH: A Quality Measure for Comparing Observed and Modeled Single Values in a Travel Demand Model Validation. Transportation Research Record, 2673(4), 722-732. https://doi.org/10.1177/0361198119838849

Schmaus, M., Bawidamann, J., Friedrich, M., Haberl, M., Trenkwalder, L., Fellendorf, M., Uhlig, J., Lohse, R., and Pestel, E.: Flüssiger Verkehr für Klimaschutz und Luftreinhaltung, Final report 14/2023, Umweltbundesamt (UBA), 2023.

Forschungsgesellschaft für Straßen- und Verkehrswesen FSGV (Ed.). (2015). Handbuch für die Bemessung von Straßenverkehrsanlagen: HBS 2015. Teil S - Stadtstraßen (Ausg. 2015, Stand: 18.9.2015). FGSV-Verl.