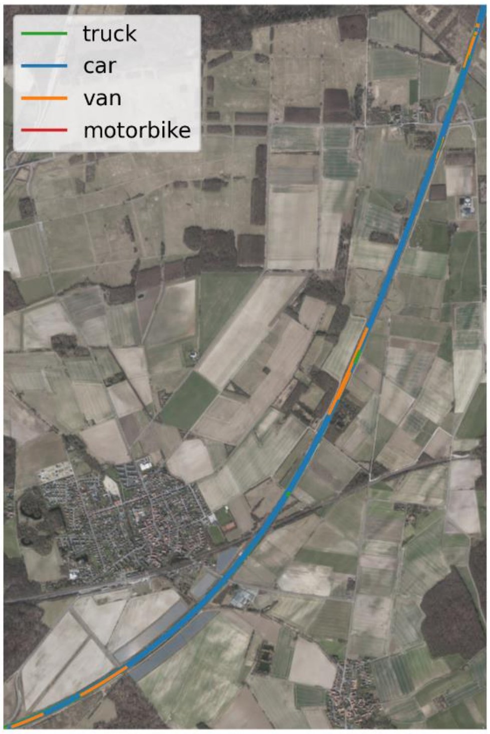
DLR Institute of Transportation Systems

DLR Highway Traffic dataset (DLR HT)

Version 1.0.0
Dataset Documentation



Contents

List of figures

List of tables

1. Introduction

One of DLR's strategic positions is to foster innovation and transfer by embedding parts of DLR tooling frameworks into external value chains (process-driven) and by linking innovative data and models to R&D and V&V procedures (data-driven). Among other pillars, DLR operates research infrastructures that provide such tooling and data frameworks for potentially embedding them into external tooling frameworks.

One strategic main development corridor is the thematic area of virtualized verification and validation processes e.g. as performed in research projects like VVM or SET Level. In both projects methodological aspects of a generic safety case - leading to a process view on the safety case - and a data- and model-driven tooling approach - leading to an open source V&V toolchain called Open Source Toolchain for Automotive Reserach (OSTAR) has been established by DLR to support the realization of safety cases by providing data, models and simulation services.

Consequently, following the above strategy, it is pursued to link examples according to data, models, and services to external technical ecosystems to generate win-win situations like professional exchange and initiation of cooperation and common research projects.

The real-world trajectory dataset offered here serves as an example for potential partners from Industry and Science to integrate DLR-generated data into their analysis and R&D toolchain procedures. The dataset was generated at the Test Bed Lower Saxony in Braunschweig, Germany, which is one part of DLR's acquisition technology. The dataset is intended to be used in research and development only. Neither the completeness nor the correctness of the dataset can be guaranteed in any way. The dataset author does not take legal responsibility for the use of this dataset for any purposes or endangerment that occurs during usage.

2. Quick Facts

Topic	**Content**
Dataset name	DLR Highway Traffic dataset
Dataset acronym	DLR HT
Dataset version	1.0.0
Date of emission	December 2024, 13th
Emitting entity	DLR Institute of Transportation Systems
Download from	https://doi.org/10.5281/zenodo.14012006
Filename	DLR-Highway-Traffic-dataset_v1-0-0.zip
Size (uncompressed)	15.9 GB
Duration	10 hours of data
Collection Period	07.10.2024 06:00:00 - 07.10.2024 16:00:00 UTC+0
License	CC BY-NC-SA 4.0
Contact	opendata-ts@dlr.de

3. General Description

3.1. Brief Description

The dataset comprises 38,215 trajectories of traffic participants, along with current local weather data, road condition data, and traffic volume data from the Test Bed Lower Saxony in Braunschweig, Germany. The trajectory data is indexed by object ID and timestamps, including detailed information about the position, speed, acceleration, dimensions, and classification of each object. The weather data provide information on wind, sunlight, precipitation, visibility, and more. The road surface data provide information on surface temperature, water layer thickness, and more. The traffic volume data provides the number of objects per lane at a specific location on the test bed near the weather station.

3.2. Data Source

The trajectory data is extracted from video recordings from 118 multi-sensor systems that are mounted at 59 different poles along the A39 highway between the Wolfsburg/Königslutter highway interchange and the Cremlingen exit. See the detailed explanation in section 5.1.1. Weather and road surface data is collected from sensors mounted on a pole. Traffic volume data is derived from the trajectory data.

3.3. Versioning

We use Semantic Versioning in the MAJOR.MINOR.PATCH format (e.g. v1.0.0) to differentiate potential future modifications to the dataset.

- The MAJOR version is incremented when data from a different time range is published.
- The MINOR version is incremented when new raw data is added to the dataset (e.g., V2X data).
- The PATCH version is incremented when errors in the data are corrected (e.g., merging two trajectories that belong to the same object) or when changes are made to the documentation.

3.4. Change Log

1.0.0

- Initial publication

3.5. Geographical Coverage

Description of the area where the data was captured:

- German Highway A39 at Google Maps
- DLR: Test Bed Lower Saxony

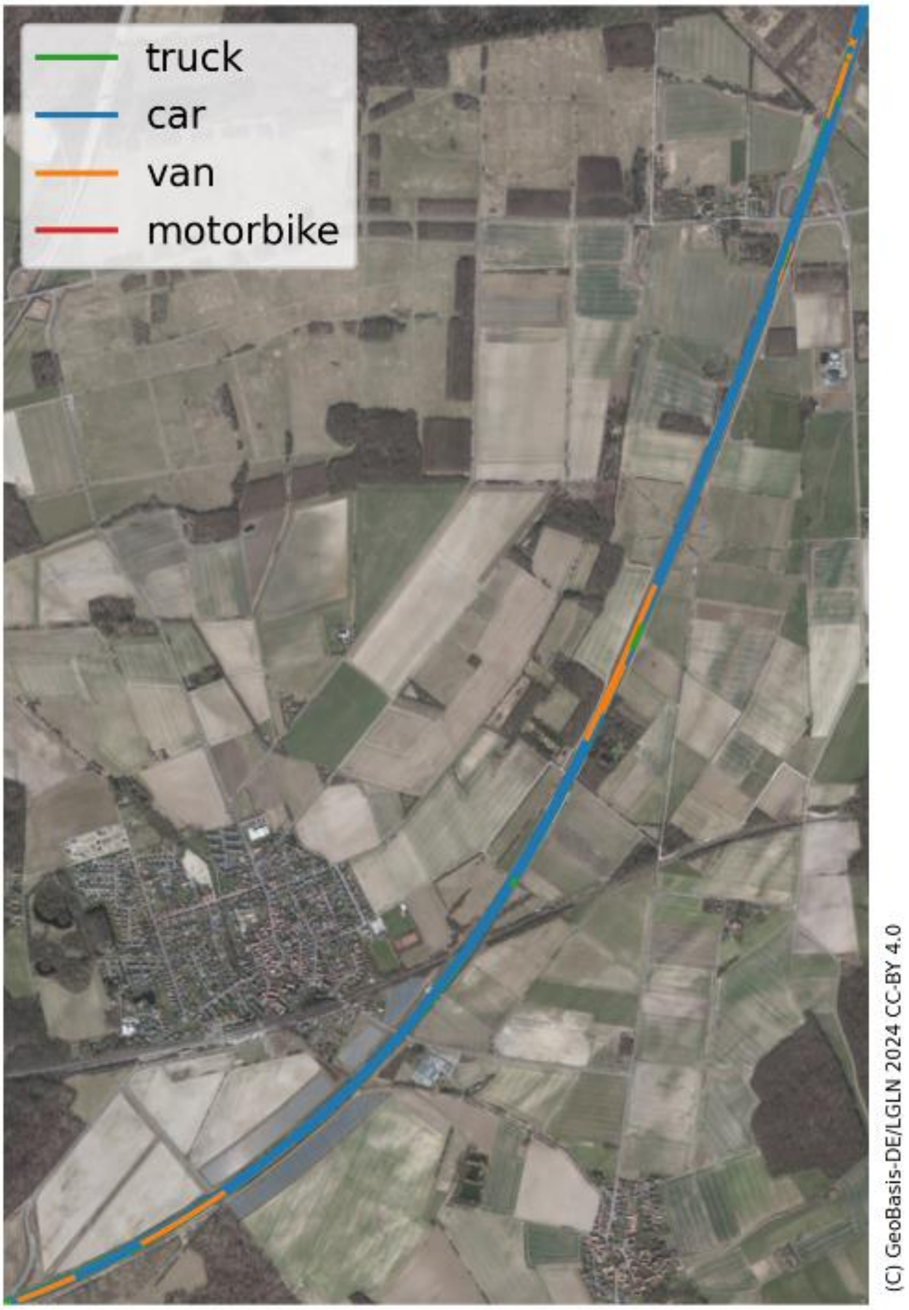


Figure 1: Geographical coverage and trajectory data from 06:15 until 06:20.

4. Data Structure and Content

4.1. General

4.1.1. Format and File Type

- File format: CSV
- Structure: The data is organized in a tabular format, with a unique index combining the timestamp and object ID to ensure each entry is distinct.

4.1.2. Data Segmentation

- The data is divided into 5-minute batches to manage file size.

4.1.3. Sampling Rate

- Trajectory data: 20 Hz
- Weather data: every 10, or 30 seconds
- Road condition data: every 30 seconds
- Traffic volume data: 1 Hz

4.2. Trajectory Data

Table 1: Trajectory data format

Name	Unit	Data type	Description
timestamp	-	ISO timestamp	The timestamp indicating when the data was recorded.
id	-	int	The traffic participant's ID is derived from the timestamp in microseconds of their first detection.
center_easting	m	float	The x-coordinate of the object's center in UTM zone 32N.
center_northing	m	float	The y-coordinate of the object's center in UTM zone 32N
velocity_easting	m/s	float	The velocity vector's component along the east direction.
velocity_northing	m/s	float	The velocity vector's component along the north direction.

velocity_magnitude	m/s	float	The speed of the object, calculated as the absolute value of the velocity vector.
acceleration_easting	m/s^2	float	The acceleration vector's component along the east direction.
acceleration_northing	m/s^2	float	The acceleration vector's component along the north direction.
acceleration_magnitude	m/s^2	float	The magnitude of the acceleration, calculated as the absolute value of the acceleration vector.
acceleration_signed	m/s^2	float	The signed acceleration is calculated as the change in velocity magnitude between two consecutive timestamps.
yaw	°	float	The heading of the bounding box in degrees, where 0° is east, increasing counterclockwise up to 360°.
dimension_length	m	float	The length of the object's bounding box.
dimension_width	m	float	The width of the object's bounding box.
dimension_height	m	float	The height of the object's bounding box.
classifications_pedestrian	-	float	The probability that the object is a pedestrian.
classifications_bicycle	-	float	The probability that the object is a bicycle.
classifications_motorbike	-	float	The probability that the object is a motorbike.
classifications_car	-	float	The probability that the object is a car.
classifications_van	-	float	The probability that the object is a van.
classifications_truck	-	float	The probability that the object is a truck.
interpolated	-	boolean	True if data was added as a result of connecting two trajectories; False otherwise.

4.3. Weather Data

Table 2: Weather data

Name	Unit	Data type	Data collection interval in seconds	Description
timestamp	-	ISO timestamp	10	The exact time when the measurement was taken.
air_temperature	°C	float	10	The temperature of the surrounding air, measured in degrees Celsius.
relative_humidity	%	float	10	The percentage of moisture in the air relative to the maximum amount the air can hold at the current temperature.
dew_point_temperature	°C	float	10	The temperature at which the air becomes saturated, and dew begins to form, measured in degrees Celsius.
wet_bulb_temperature	°C	float	10	The lowest temperature that can be achieved by evaporating water into the air, measured in degrees Celsius.
air_pressure_msl	hPa	float	10	The atmospheric pressure at mean sea level, given in hectopascals (hPa).
wind_direction	°	float	10	The direction from which the wind is blowing, measured in degrees (0° = North, 90° = East, etc.).
wind_speed	m/s	float	10	The speed of the wind, measured in meters per second (m/s).

wind_gust_direction	°	int	10	The direction from which the strongest wind gusts during a measurement period originated, measured in degrees.
hail_intensity	hits/cm²	int	10	The intensity of hail falling, based on the size and number of hailstones.
visibility	m	float	30	The horizontal visibility distance, measured in meters, indicating how far one can see clearly.
present_weather	-	float	30	The current weather conditions according to World Meteorological Organization (WMO) standards, based on standardized codes (Codes).
rain_intensity	mm/h	float	30	The intensity of rainfall, measured in millimeters per hour (mm/h).
rain_accumulation	mm	float	30	The total amount of rainfall accumulated over a specified period, measured in millimeters.
snow_accumulation	mm	float	30	The total amount of snow that has accumulated over a specified period, measured in millimeters.

4.4. Road Condition Data

Table 3: Road condition data

Name	Unit	Data type	Data collection interval in seconds	Description
timestamp	-	ISO timestamp	30	The exact time when the measurement was taken.
surface_temperature	°C	float	30	This measurement reflects the thermal state of the road surface.
surface_state	-	int	30	The condition of the road surface.
surface_grip	-	float	30	The traction of the road surface. (0: lowest grip, 0.82: highest grip)
water_layer_thickness	mm	float	30	The thickness of the water layer on the road surface, measured in millimeters.
ice_layer_thickness	mm	float	30	The thickness of the ice layer on the road surface, measured in millimeters.
snow_layer_thickness	mm	float	30	The thickness of the snow layer on the road surface, measured in millimeters.

4.4.1. Explanation of Surface States

Table 4: Surface states

State	Description
0	Error
1	Dry
2	Moist
3	Wet

5	Forsty
6	Snowy
7	Icy
9	Slushy

4.5. Traffic Volume Data

Table 5: Traffic volume data

Name	Unit	Data type	Data collection interval in seconds	Description
timestamp	-	ISO timestamp	1	Start timestamp of the second in which the objects were counted.
north_south_left	objects	int	1	Number of objects passing the left lane (in the direction of travel) from north to south at UTM Northing 5,794,749.
north_south_right	objects	int	1	Number of objects passing the right lane (in the direction of travel) from north to south at UTM Northing 5,794,749.
south_north_left	objects	int	1	Number of objects passing the left lane (in the direction of travel) from south to north at UTM Northing 5,794,749.
south_north_right	objects	int	1	Number of objects passing the right lane (in the direction of travel) from south to north at UTM Northing 5,794,749.

5. Data Collection and Processing

5.1. Collection Methodology

Description of the methods and devices used to collect the trajectories.

5.1.1. Trajectory Data

The trajectory data is extracted from video recordings from 118 multi-sensor systems that are mounted along the A39 highway at 59 different poles between the Wolfsburg/Königslutter highway interchange and the Cremlingen exit as depicted in Figure 2.

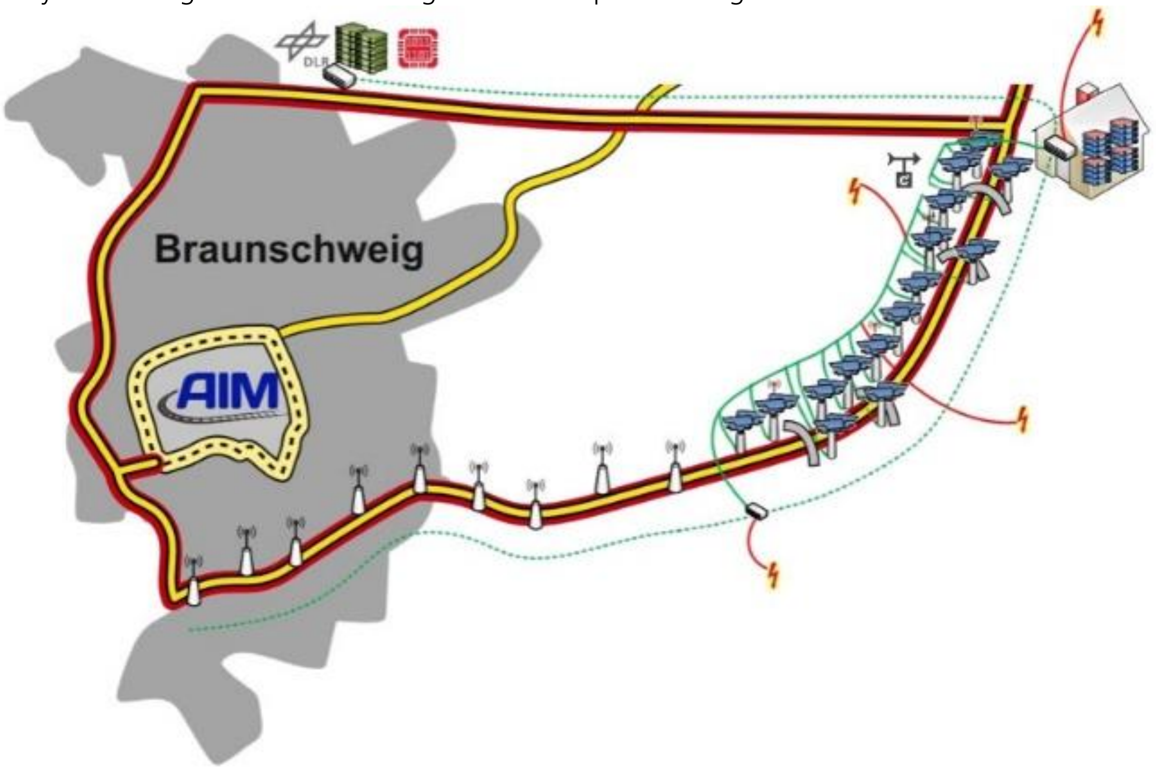


Figure 2: Symbolic representation of Testbed Lower Saxony, with cameras and weather station at upper right part of Highway A39.

The field of view of each sensor reach the field of view of the next sensor as illustrated in Figure 3. This setup creates a redundant system that effectively captures all relevant objects on the highway.

Figure 3: Example of camera field of views (FOVs).

The view of one camera is shown in Figure 4.

Figure 4: Example of camera image.

Each multi-sensor system includes a pair of cameras (GiGEVision mvBlueCOUGAR-109b) and active infrared lighting (between the cameras) to improve scene visibility. The cameras (orange) and infrared lighting (green) are shown in Figure 5.

Figure 5: Multi-sensor system mounted on pole next to the A39 highway, cameras and infrared lighting highlighted

Object detection and tracking are based on the principle of optical flow. Objects are identified by their movement patterns at the pixel level in consecutive images. From the resulting flow field, 3D voxels are generated (white and colored pillars in the left image), which are then aggregated into objects in the form of bounding boxes (green box in the right image) as depicted in Figure 6.

Figure 6: Object detection

5.1.2. Weather and Road Condition Data

The weather data is collected from sensors mounted on one of the poles at coordinates 52.29042 10.71662 (see Figure 2). General weather information is collected using the WXT536 sensor. The PWD22 sensor is responsible for determining visibility. The DSC211 and DST111 sensors are employed to assess the condition of the road surface.

5.1.3. Traffic Volume Data

The traffic volume was derived from the trajectory data by calculating the time and lane at which a trajectory crossed the UTM Northing value 5,794,749.

5.2. Preprocessing

Steps taken to process the raw data.

5.2.1. Trajectory Data

- Trajectories of objects lasting less than 4 seconds are removed.
- Trajectories of objects covering a distance of less than 40 meters are removed.
- Dimensions and object classifications are estimated for each timestamp in the raw data. During preprocessing, the median object size is calculated and assigned to every timestamp.
- Connecting broken trajectories: Each trajectory is predicted into the future. If a new trajectory starts within 40 meters of the predicted position, the two are connected. The ID of the first object is then used to overwrite the ID of the second object.
- A Kalman Filter is applied for smoothing and interpolating the position. Velocity and acceleration values are derived from the position data.

5.2.2. Weather, Road Condition and Traffic Volume Data

- Data is not preprocessed.

5.3. Data Quality

While the data may not be a perfect representation of reality, it provides a digital twin, allowing for traffic analyses. If you see any opportunities to improve the data quality, please feel free to contact us. Please be aware of the following limitations when interpreting the data.

5.3.1. Trajectory Data

- Accuracy:
  - The accuracy decreases as the distance from the nearest camera increases.
- Limitations:
  - The position, velocity, acceleration, heading, size, and object classification may not perfectly reflect reality.
- Known Issues:

- Preprocessing may not fully address all data limitations. Some objects may not have been detected due to insufficient contrast with the background.
- In shaded areas (e.g., under bridges), object tracking may be interrupted as detection can fail for several milliseconds.
- Objects might not be continuously tracked, leading to multiple trajectories being recorded for the same object.
- Conversely, objects traveling close to each other may be mistaken for a single object, leading to one trajectory being recorded for multiple objects.
- Occasionally, a trajectory may switch between different objects, meaning that a single trajectory ID could represent the movements of multiple objects.

5.3.2. Weather Data and Road Condition Data

Sensors are mounted on a pole in the northern area of the testbed. Measurements represent conditions at this location. The farther you move from the sensors, the less accurate the data becomes.

5.4. Anonymization

How personally identifiable information was protected: The raw high-resolution video data is processed directly at the testbed in a server house to receive anonymized trajectory data. High-resolution images are converted to anonymized low-resolution images for later DLR-internal analysis. Camera images have not been published yet. More information: https://www.dlr.de/en/ts/research-transfer/research-infrastructure/data-protection/data-protection-in-the-test-bed-lower-saxony

6. Citation and Usage

6.1. Access Rights

The dataset is published under CC BY-NC-SA 4.0 license.

6.2. Citation

How to cite the dataset in academic work: A publication about the dataset will follow. For now, refer to the Zenodo publication including a DOI.

6.3. Code

You can use our open source Python software TASI (TrAffic Situation analysis and Interpretation) to download, analyze, and visualize the data from the DLR HT dataset. For more details about TASI, please visit our GitHub repository https://github.com/DLR-TS/TASI.

6.4. Use Cases

These use cases demonstrate the versatility of the dataset in addressing a wide range of research and practical applications, from enhancing traffic safety to supporting the development of autonomous vehicles.

Autonomous Vehicle Development and Validation: The dataset can be instrumental in validating autonomous driving algorithms. By simulating various weather scenarios, developers can test how autonomous systems respond to real-world challenges, improving the safety and reliability of autonomous vehicles in diverse conditions.

Behavioral Analysis of Road Users: The data allows for in-depth analysis of how different types of road users (e.g., cars, trucks) adapt their behavior in response other traffic participants. This can lead to the development of more effective traffic management strategies and infrastructure improvements to enhance safety and comfort for all road users.

Safety Assessment and Accident Prediction: The dataset can be used to study the relationship between weather conditions and accident-prone areas. By analyzing trajectory deviations, abrupt stops, or unsafe maneuvers, researchers can predict potential accident hotspots and develop strategies to mitigate risks, especially during adverse weather conditions.

Conflict Analysis and Prediction: The dataset is valuable for analyzing traffic conflicts and predicting potential conflict scenarios on the road. Researchers can utilize the data to identify patterns and factors leading to conflicts between cars, or other road users. By advancing Surrogate Measures of Safety (SMoS), the dataset allows for the development of more accurate and reliable methods to assess safety performance. These insights can contribute to proactive traffic management strategies, enhancing road safety and reducing the likelihood of accidents before they occur.

7. Examples

7.1. Sample Data

Small excerpts from the dataset to show users what the data looks like.

7.1.1. Trajectory data

To improve readability in this documentation, the data in the following table is presented in a transposed format.

Table 6: Example: Trajectory data

timestamp	2024-10-07 06:00:00.004659+00:00	2024-10-07 06:00:00.004659+00:00	2024-10-07 06:00:00.004659+00:00
id	1728280701706084	1728280711579163	1728280715873385
center_easting	616517.884	616874.078	616362.445
center_northing	5793321.779	5794184.485	5793017.125
velocity_easting	14.053	-8.869	17.084
velocity_northing	29.593	-23.115	30.56
velocity_magnitude	32.76	24.758	35.011
acceleration_easting	-0.079	0.089	-0.253
acceleration_northing	0.064	0.593	0.207
acceleration_magnitude	0.101	0.6	0.327
acceleration_signed	0.14	-1.0	0.061
yaw	64.434	-110.835	60.679
dimension_length	6.91	4.603	4.44
dimension_width	2.386	1.862	1.892
dimension_height	2.229	1.896	1.603
classifications_pedestrian	0.0	0.0	0.0
classifications_bicycle	0.0	0.0	0.0
classifications_motorbike	0.0	0.0	0.004

classifications_car	0.0	0.363	0.768
classifications_van	0.395	0.502	0.08
classifications_truck	0.605	0.135	0.148
interpolated	False	False	False

7.1.2. Weather Data

To improve readability in this documentation, the data in the following table is presented in a transposed format.

Table 7: Example: Weather data

timestamp	2024-10-07 06:00:00+00:00	2024-10-07 06:00:10+00:00	2024-10-07 06:00:20+00:00
air_temperature	9.05	9.05	9.05
relative_humidity	81.567	81.583	81.6
dew_point_temperature	6.021	6.064	6.064
wet_bulb_temperature	7.482	7.526	7.526
air_pressure_msl	991.7	991.7	991.7
wind_direction	101.0	57.0	60.0
wind_speed	1.167	1.466	0.635
wind_gust_direction	107.0	107.0	107.0
hail_intensity	0.0	0.0	0.0
visibility	NaN	NaN	20000.0
present_weather	NaN	NaN	0.0
rain_intensity	NaN	NaN	0.0
rain_accumulation	NaN	NaN	0.0
snow_accumulation	NaN	NaN	0.0

7.1.3. Road Condition Data

To improve readability in this documentation, the data in the following table is presented in a transposed format.

Table 8: Example: Road condition data

timestamp	2024-10-07 00:00:20+00:00	2024-10-07 00:00:50+00:00	2024-10-07 00:01:20+00:00
surface_temperature	9.7	9.7	9.7
surface_state	1.0	1.0	1.0
surface_grip	0.82	0.82	0.82
water_layer_thickness	0.0	0.0	0.0
ice_layer_thickness	0.0	0.0	0.0
snow_layer_thickness	0.0	0.0	0.0

7.1.4. Traffic Volume Data

To improve readability in this documentation, the data in the following table is presented in a transposed format.

Table 9: Example: Traffic volume data

timestamp	2024-10-07 06:00:00+00:00	2024-10-07 06:00:01+00:00	2024-10-07 06:00:02+00:00
north_south_left	0	0	0
north_south_right	0	0	0
south_north_left	0	1	0
south_north_right	1	0	0

8. Additional Resources

- DLR Urban Traffic dataset v1.1.0: Traffic data from the AIM Research Intersection: https://zenodo.org/records/14025010