# Format Specification of the TF4CHE Dataset

The Traffic Flow Dataset for China’s Congested Highways & Expressways (TF4CHE) is derived from AD4CHE (Aerial Dataset for China's Congested Highways & Expressways). AD4CHE collects data using unmanned aerial vehicles (UAVs) operating at an altitude of 100 meters and employs advanced calibration techniques to achieve a positioning accuracy of approximately 5 cm. It provides comprehensive vehicle metrics, including position, speed, classification, as well as unique parameters such as self-offset and yaw rate. AD4CHE covers highway and expressway data across 11 distinct scenarios in five cities in China, comprising a total of 68 data segments. Each data segment includes three files: xx_recordingMeta.csv, xx_tracks.csv, and xx_tracksMeta.csv, which respectively provide video metadata, vehicle trajectories, and trajectory metadata. However, AD4CHE is a pre-development dataset, and its raw parameters are not directly applicable to traffic flow time-series forecasting tasks. TF4CHE is constructed by processing the original frame rate data from AD4CHE into sequential data and applying theoretical formulations from the rail transit domain. It serves as an application-oriented dataset suitable for direct use in research on traffic flow prediction and congestion identification.

## 1 TF4CHE Files and Parameter Description

The TF4CHE dataset is divided into 11 groups according to different recorded road segments, named "Road_segment_1.csv" through "Road_segment_11.csv". The specific road segments corresponding to each group are illustrated in Fig. 1.

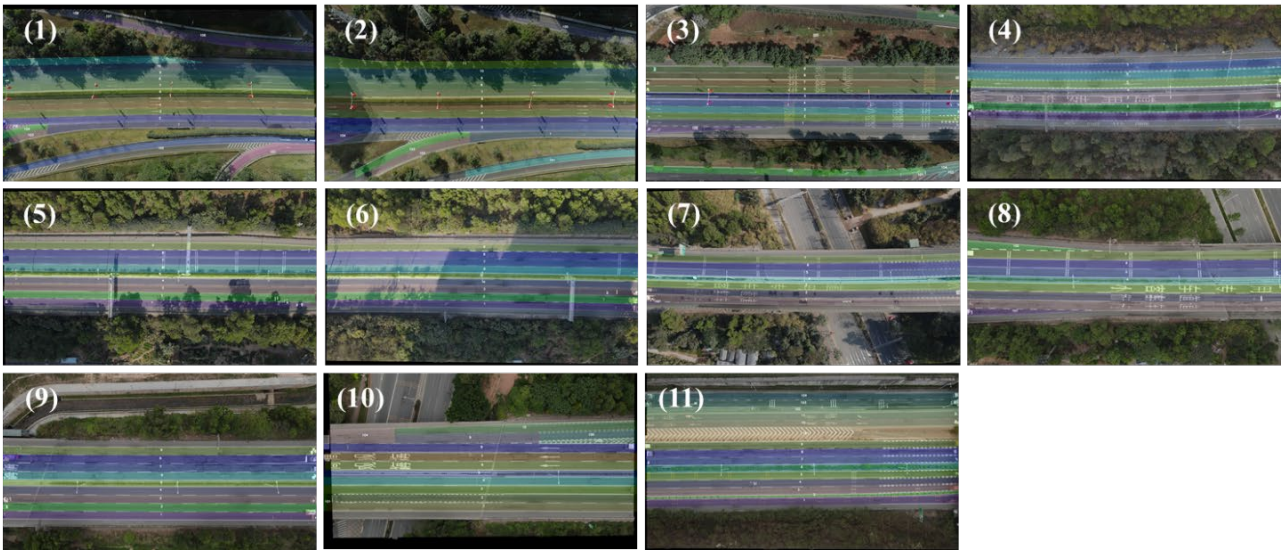


Fig. 1. Eleven recorded road segments in the TF4CHE dataset.

The lane layout of the recorded road segments is illustrated in Fig. 2. In the world co ordinate system, the origin is defined at the top-left corner of the image. The horizontal X-axis corresponds to the vehicle traveling direction and increases toward the right, whereas t he vertical Y-axis increases downward. There are four lanes traveling in the positive X-axis direction, numbered "1–3" along with an emergency lane, and four lanes traveling in the ne gative X-axis direction, numbered "5–7" along with an emergency lane.

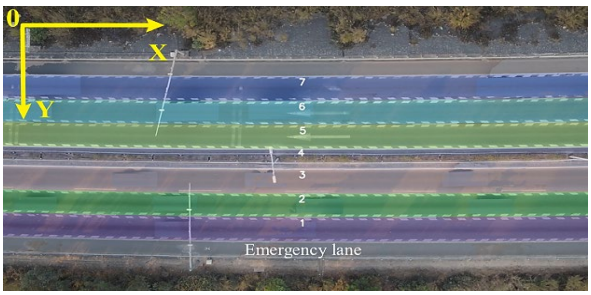

Fig. 2. Lane layout of the TF4CHE dataset.

The TF4CHE dataset covers various typical time-series traffic flow parameters for different traveling directions on expressways, including the number of three types of vehicles (car, bus, and truck), the total number of equivalent vehicles $G(t)$, average traffic density $k(t)$, average flow $q(t)$, average traffic speeds and accelerations along the X and Y directions, and lane space occupancy OccupancyRatio. The specific traffic flow parameters contained in each scenario are listed in the following table.

Table 1. Specific traffic flow parameters contained in each recorded road segment.

| Name | Description | Unit |
|---|---|---|
| Month/Year | Month and year of video recording (virtual date information) | [-] |
| Weekday | Completion date of video recording (virtual date information) | [-] |
| TimeCode | Specific start time of video recording (virtual time information) | [-] |
| second | Video time sequence in seconds | [s] |
| drivingDirection | Traveling direction of the recorded segment (1: along negative X-axis; 2: along positive X-axis) | [-] |
| car | Number of cars present in the frame at the given time | [veh] |
| bus | Number of buses present in the frame at the given time | [veh] |
| truck | Number of trucks present in the frame at the given time | [veh] |
| $G(t)$ | Total number of equivalent vehicles at the given time, with cars as the reference (conversion coefficients: $\alpha_{bus}$=2，$\alpha_{truck}$=2.5) | [veh] |
| $k(t)$ | Average vehicle density at the given time | [veh/m] |
| $q(t)$ | Average traffic flow at the given time | [veh/s] |
| xVelocity($t$) | Average traffic speed along the X-axis at the given time | [m/s] |
| yVelocity($t$) | Average traffic speed along the Y-axis at the given time | [m/s] |
| xAcceleration($t$) | Average traffic acceleration along the X-axis at the given time | [m/s$^2$] |
| yAcceleration($t$) | Average traffic acceleration along the Y-axis at the given time | [m/s$^2$] |
| OccupancyRatio | Lane space occupancy at the given time | [-] |
| File_ID | Corresponding segment index from the original AD4CHE dataset (out of 68 segments) | [-] |

A partial data example from Road Segment 1 is shown below:

| Month/Year | weekday | TimeCode | second | drivingDirection | car | bus | truck | G(t) | K(t) | q(t) | xVelocity(t) | yVelocity(t) | xAcceleration(t) | yAcceleration(t) | Occupancy Ratio | File_ID |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4.2021 | Mon | 0:00:00 | 0 | 1 | 18 | 0 | 2 | 23 | 0.038721 | 20 | -12.7589 | -0.617 | -0.11188 | 0.153529 | 0.148418 | 1 |
| 4.2021 | Mon | 0:00:01 | 1 | 1 | 18 | 0 | 2 | 23 | 0.038721 | 20 | -12.785 | -0.552 | -0.07722 | 0.0785 | 0.146852 | 1 |
| 4.2021 | Mon | 0:00:02 | 2 | 1 | 18 | 0 | 2 | 23 | 0.038721 | 20 | -13.2875 | -0.4855 | -0.10895 | 0.06 | 0.147761 | 1 |
| 4.2021 | Mon | 0:00:03 | 3 | 1 | 17 | 0 | 2 | 22 | 0.037037 | 19 | -13.2879 | -0.076 | -0.14105 | 0.087895 | 0.141448 | 1 |
| 4.2021 | Mon | 0:00:04 | 4 | 1 | 17 | 0 | 2 | 22 | 0.037037 | 19 | -13.5705 | -0.0275 | -0.08947 | 0.085556 | 0.141229 | 1 |
| 4.2021 | Mon | 0:00:05 | 5 | 1 | 19 | 0 | 2 | 24 | 0.040404 | 21 | -13.2839 | -0.17895 | -0.05429 | 0.094737 | 0.154226 | 1 |
| 4.2021 | Mon | 0:00:06 | 6 | 1 | 17 | 0 | 3 | 24.5 | 0.041246 | 20 | -13.2539 | 0.027222 | -0.09389 | 0.07 | 0.150034 | 1 |
| 4.2021 | Mon | 0:00:07 | 7 | 1 | 18 | 0 | 2 | 23 | 0.038721 | 20 | -13.4172 | -0.01833 | -0.07059 | 0.07 | 0.131414 | 1 |
| 4.2021 | Mon | 0:00:08 | 8 | 1 | 15 | 0 | 3 | 22.5 | 0.037879 | 18 | -13.6156 | -0.09765 | -0.082 | 0.060667 | 0.119394 | 1 |
| 4.2021 | Mon | 0:00:09 | 9 | 1 | 16 | 0 | 3 | 23.5 | 0.039562 | 19 | -13.83 | -0.27944 | -0.09778 | 0.07 | 0.126195 | 1 |
| 4.2021 | Mon | 0:00:10 | 10 | 1 | 17 | 0 | 3 | 24.5 | 0.041246 | 20 | -13.5226 | -0.16444 | -0.10158 | 0.02 | 0.140067 | 1 |
| 4.2021 | Mon | 0:00:11 | 11 | 1 | 11 | 0 | 3 | 18.5 | 0.031145 | 14 | -13.5857 | -0.19308 | -0.11923 | 0.112308 | 0.100589 | 1 |
| 4.2021 | Mon | 0:00:12 | 12 | 1 | 13 | 0 | 2 | 18 | 0.030303 | 15 | -13.8713 | -0.168 | -0.09133 | 0.111538 | 0.106077 | 1 |
| 4.2021 | Mon | 0:00:13 | 13 | 1 | 16 | 0 | 2 | 21 | 0.035354 | 18 | -13.62 | -0.14063 | -0.06 | 0.094667 | 0.126869 | 1 |
| 4.2021 | Mon | 0:00:14 | 14 | 1 | 17 | 0 | 2 | 22 | 0.037037 | 19 | -13.6884 | -0.06941 | -0.05933 | 0.064 | 0.133316 | 1 |
| 4.2021 | Mon | 0:00:15 | 15 | 1 | 17 | 0 | 3 | 24.5 | 0.041246 | 20 | -13.7116 | -0.28056 | -0.05778 | 0.102941 | 0.149882 | 1 |
| 4.2021 | Mon | 0:00:16 | 16 | 1 | 18 | 0 | 3 | 25.5 | 0.042929 | 21 | -13.7465 | -0.09833 | -0.0235 | 0.052632 | 0.155623 | 1 |
| 4.2021 | Mon | 0:00:17 | 17 | 1 | 16 | 0 | 3 | 23.5 | 0.039562 | 19 | -13.55 | -0.14438 | -0.05789 | 0.0675 | 0.135539 | 1 |
| 4.2021 | Mon | 0:00:18 | 18 | 1 | 16 | 0 | 3 | 23.5 | 0.039562 | 19 | -13.2956 | -0.20882 | -0.02211 | 0.095556 | 0.13335 | 1 |
| 4.2021 | Mon | 0:00:19 | 19 | 1 | 15 | 0 | 2 | 20 | 0.03367 | 17 | -13.5635 | -0.08938 | 0.008824 | 0.085714 | 0.119394 | 1 |
| 4.2021 | Mon | 0:00:20 | 20 | 1 | 17 | 0 | 2 | 22 | 0.037037 | 19 | -13.3021 | -0.08611 | 0.040588 | 0.057778 | 0.132155 | 1 |

## 2 Visualization of typical traffic flow parameters in TF4CHE

Most of the metadata in the TF4CHE dataset were collected under congested conditions on expressways and arterial roads. To illustrate the overall distribution patterns of the dataset parameters, a visualization analysis of representative traffic flow parameters is performed.

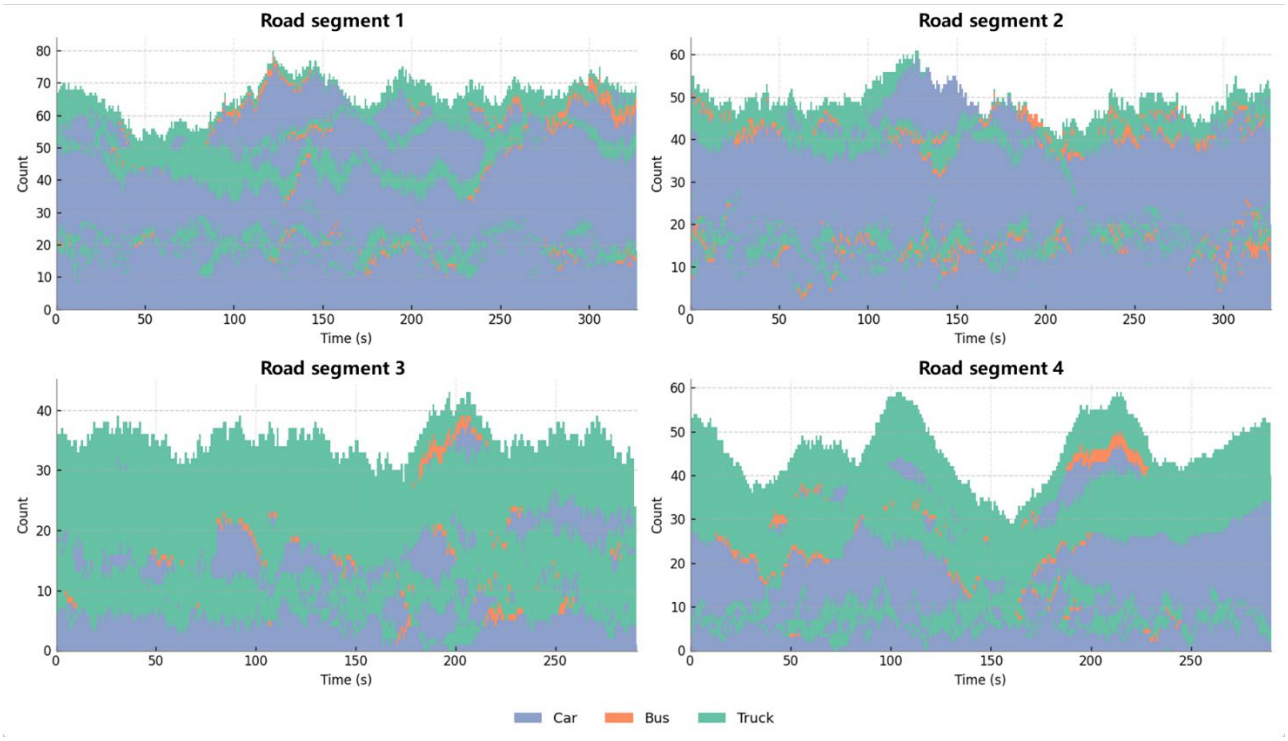


Fig. 3. Time-series stacked bar chart of car, bus, and truck counts in a typical road segment.

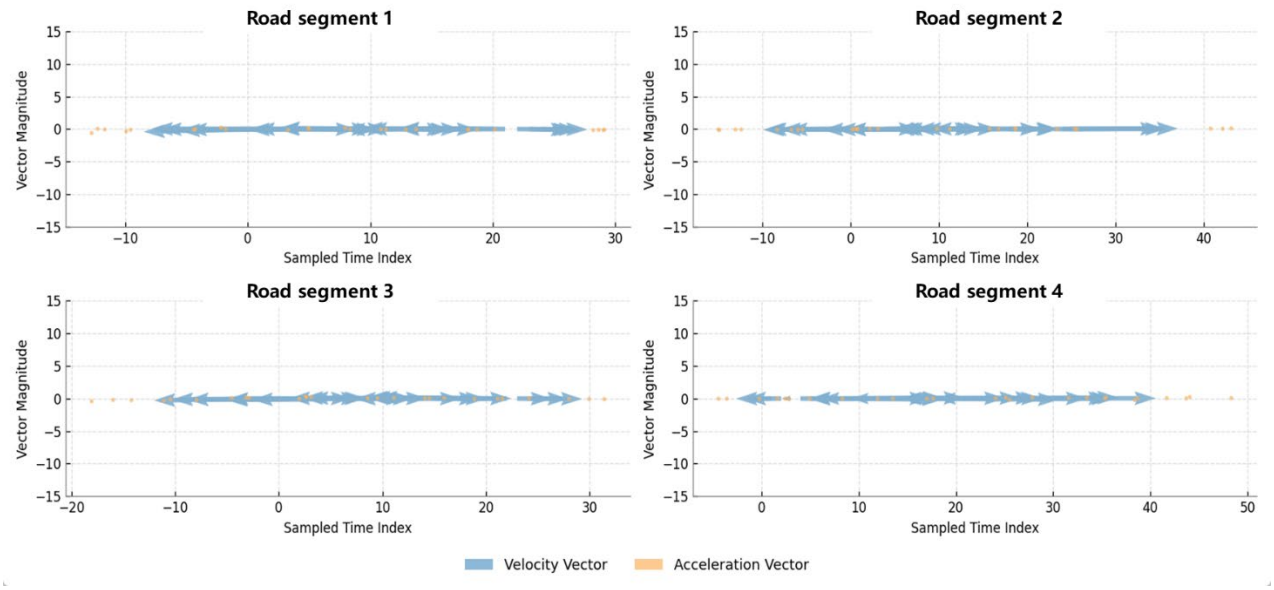


Fig. 4. Time-series vector field analysis of average speed and corresponding average acceleration in a typical road segment.

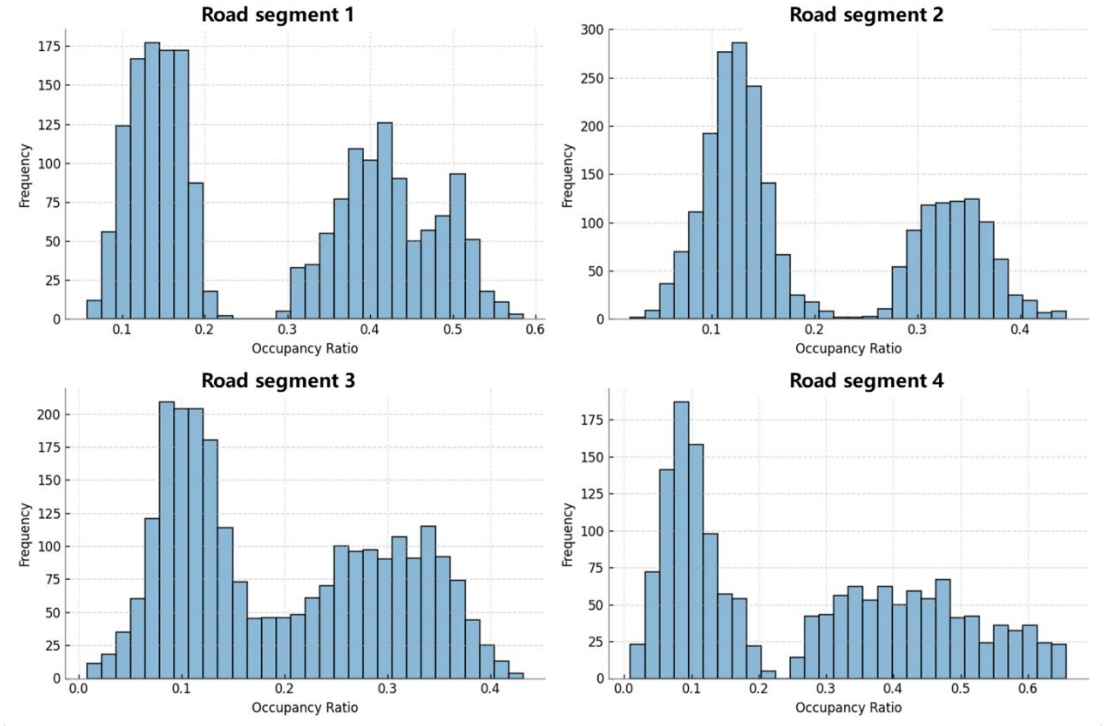


Fig. 5. Time-series bar chart of lane space occupancy in a typical road segment.

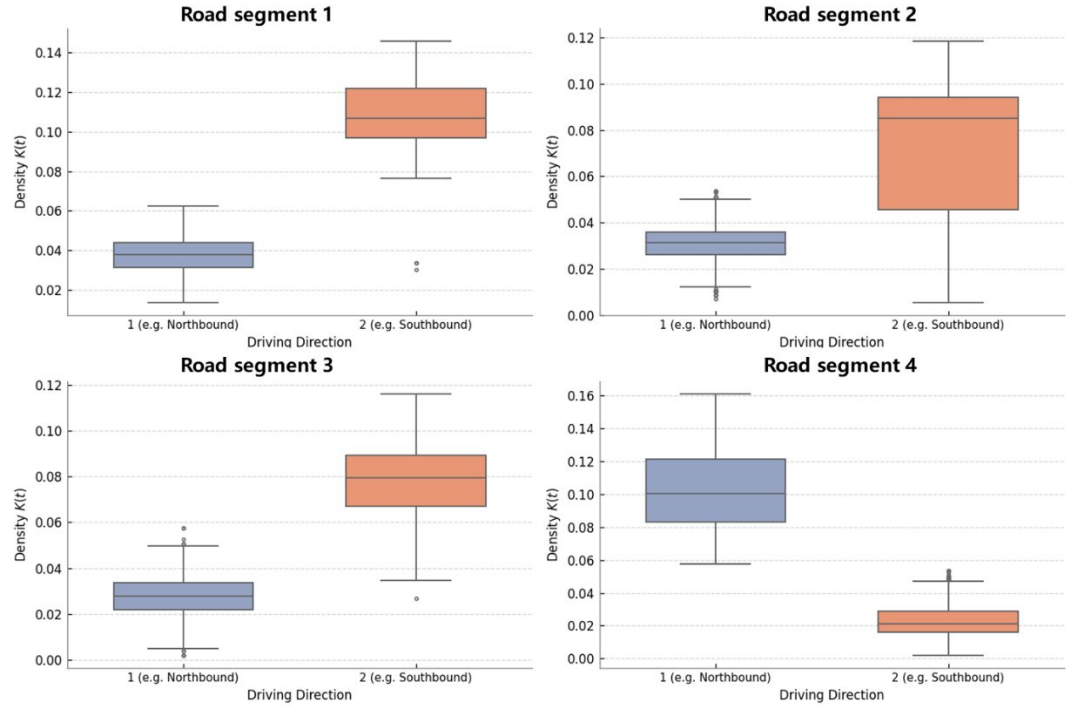


Fig. 6. Time-series box plot of traffic density for different traveling directions in a typical road segment.

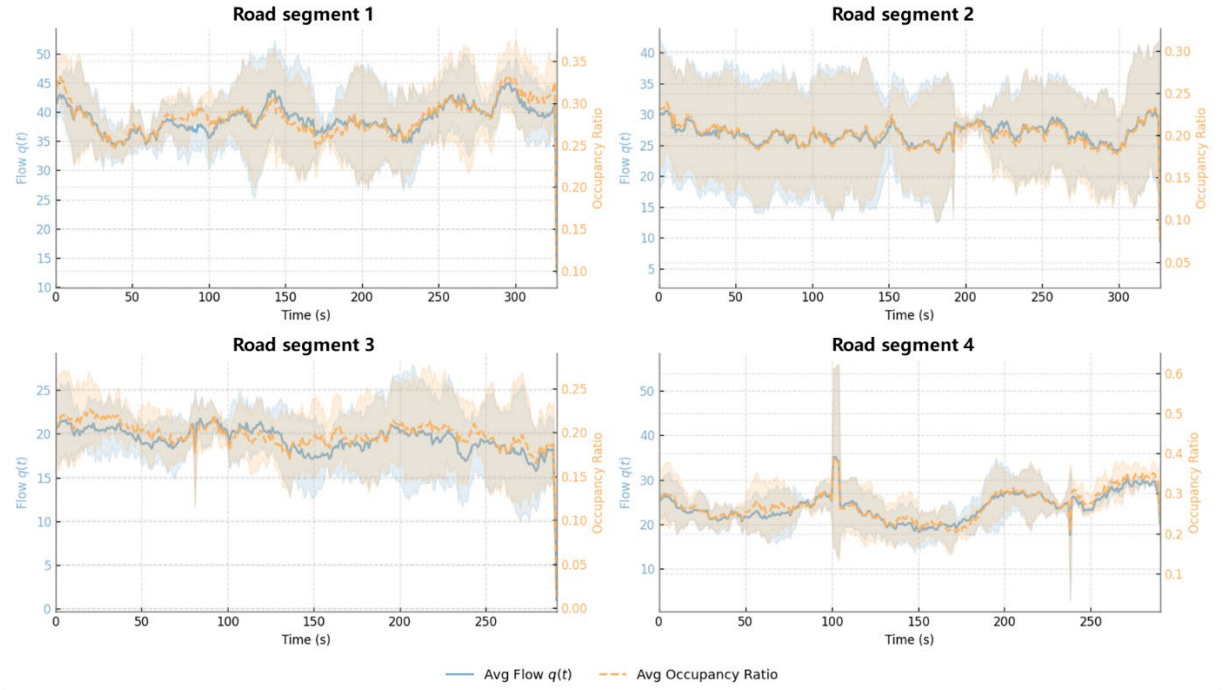


Fig. 7. Time-series curves with error bands for average flow and lane space occupancy in a typical road segment.

# 3 Comparison between TF4CHE and AD4CHE

The TF4CHE dataset is derived from AD4CHE; however, its time-series structure makes it more accessible for practical application. The differences between the two datasets in terms of file format and data structure are summarized as follows:

(1) File Format

AD4CHE is a specialized natural driving dataset focused on congested scenarios across four cities in China, containing 5.12 hours of high-precision aerial data. It covers expressway and arterial road data across 11 distinct scenarios in five cities, comprising a total of 68 data segments. Each

segment contains three files: xx_recordingMeta.csv, xx_tracks.csv, and xx_tracksMeta.csv, which respectively provide video metadata, vehicle trajectories, and trajectory metadata.

In contrast, TF4CHE is constructed through the analysis, extraction, and processing of the AD4CHE data. It organizes the data into 11 groups based on different recorded road segments, with each group saved as a single CSV file named "Road_segment_1.csv" through "Road_segment_11.csv." Compared to AD4CHE, TF4CHE adopts a simpler and more consolidated file structure.

(2) Data Structure and Key Parameters

In AD4CHE, the xx_tracks.csv file records frame-level information for each vehicle trajectory, resulting in a more complex data structure and larger data volume, as illustrated in Fig. 8 TF4CHE processes metadata from xx_tracks.csv by integrating frame rate information from xx_recordingMeta.csv and vehicle attributes from xx_tracksMeta.csv. It converts the data into time-series information based on "seconds" and computes typical traffic flow parameters, including vehicle density, traffic flow, average speed, average acceleration, and lane space occupancy, with data organized by recording scenario.

In terms of data structure, TF4CHE more closely adheres to the style of classical time-series datasets, thereby making it an application-oriented dataset that can be directly utilized in intelligent transportation research areas such as traffic flow prediction and congestion identification.

| frame | id | x | y | width | height | xVelocity | yVelocity | xAcceleration | yAcceleration | frontSightDistance | backSightDistance | dhw | thw | ttc | precedingXVelocity | precedingId | followingId | leftPrecedingId | leftAlongsideId | leftFollowingId | rightPrecedingId | rightAlongsideId | rightFollowingId | laneId | angle | orientation | yaw_rate | ego_offset |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 48.73 | 52.39 | 14.14 | 2.1 | 2.82 | -0.09 | 0.22 | 0.02 | 95.27 | 48.73 | 10.72 | 3.81 | -12.66 | 3.67 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | 0 | 0.02 |
| 1 | 1 | 48.81 | 52.38 | 14.14 | 2.1 | 2.82 | -0.09 | 0.22 | 0.02 | 95.19 | 48.81 | 10.76 | 3.82 | -12.71 | 3.67 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.04 | 0.01 |
| 2 | 1 | 48.96 | 52.39 | 14.14 | 2.1 | 2.83 | -0.09 | 0.39 | 0.06 | 95.04 | 48.96 | 10.72 | 3.79 | -12.77 | 3.67 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.04 | 0.02 |
| 3 | 1 | 49.03 | 52.38 | 14.14 | 2.1 | 2.84 | -0.09 | 0.39 | 0.06 | 94.97 | 49.03 | 10.8 | 3.8 | -12.98 | 3.68 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.04 | 0.01 |
| 4 | 1 | 49.11 | 52.39 | 14.14 | 2.1 | 2.86 | -0.08 | 0.39 | 0.05 | 94.89 | 49.11 | 10.84 | 3.79 | -13.15 | 3.68 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.04 | 0.02 |
| 5 | 1 | 49.26 | 52.39 | 14.14 | 2.1 | 2.87 | -0.08 | 0.38 | 0.05 | 94.74 | 49.26 | 10.8 | 3.76 | -13.22 | 3.69 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.04 | 0.02 |
| 6 | 1 | 49.29 | 52.41 | 14.14 | 2.1 | 2.88 | -0.08 | 0.39 | 0.04 | 94.71 | 49.29 | 10.91 | 3.78 | -13.5 | 3.69 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.03 | 0.04 |
| 7 | 1 | 49.37 | 52.41 | 14.14 | 2.1 | 2.9 | -0.08 | 0.38 | 0.03 | 94.63 | 49.37 | 10.95 | 3.78 | -13.69 | 3.7 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.03 | 0.04 |
| 8 | 1 | 49.56 | 52.41 | 14.14 | 2.1 | 2.91 | -0.08 | 0.35 | 0.03 | 94.44 | 49.56 | 10.91 | 3.75 | -13.77 | 3.7 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.03 | 0.06 |
| 9 | 1 | 49.74 | 52.41 | 14.14 | 2.1 | 2.92 | -0.08 | 0.31 | 0.03 | 94.26 | 49.74 | 10.84 | 3.71 | -13.8 | 3.7 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.03 | 0.07 |
| 10 | 1 | 49.82 | 52.39 | 14.14 | 2.1 | 2.93 | -0.08 | 0.3 | 0.03 | 94.18 | 49.82 | 10.95 | 3.74 | -14.07 | 3.71 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.03 | 0.06 |
| 11 | 1 | 49.86 | 52.39 | 14.14 | 2.1 | 2.94 | -0.08 | 0.31 | 0.03 | 94.14 | 49.86 | 11.03 | 3.75 | -14.3 | 3.71 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.03 | 0.06 |
| 12 | 1 | 49.97 | 52.41 | 14.14 | 2.1 | 2.95 | -0.08 | 0.3 | 0.02 | 94.03 | 49.97 | 10.99 | 3.73 | -14.39 | 3.71 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.03 | 0.07 |
| 13 | 1 | 50.12 | 52.4 | 14.14 | 2.1 | 2.96 | -0.07 | 0.28 | 0.02 | 93.88 | 50.12 | 11.03 | 3.73 | -14.57 | 3.71 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.03 | 0.07 |
| 14 | 1 | 50.23 | 52.4 | 14.14 | 2.1 | 2.97 | -0.07 | 0.26 | 0.02 | 93.77 | 50.23 | 11.06 | 3.73 | -14.76 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.03 | 0.07 |
| 15 | 1 | 50.34 | 52.4 | 14.14 | 2.1 | 2.97 | -0.07 | 0.25 | 0.02 | 93.66 | 50.34 | 11.1 | 3.73 | -14.94 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.03 | 0.07 |
| 16 | 1 | 50.46 | 52.37 | 14.14 | 2.1 | 2.98 | -0.07 | 0.24 | 0.03 | 93.54 | 50.46 | 11.1 | 3.72 | -15.08 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.02 | 0.04 |
| 17 | 1 | 50.53 | 52.38 | 14.14 | 2.1 | 2.99 | -0.07 | 0.24 | 0.03 | 93.47 | 50.53 | 11.17 | 3.74 | -15.33 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.02 | 0.05 |
| 18 | 1 | 50.64 | 52.38 | 14.14 | 2.1 | 3 | -0.07 | 0.23 | 0.03 | 93.36 | 50.64 | 11.17 | 3.73 | -15.47 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.02 | 0.05 |
| 19 | 1 | 50.79 | 52.37 | 14.14 | 2.1 | 3.01 | -0.07 | 0.22 | 0.03 | 93.21 | 50.79 | 11.17 | 3.72 | -15.62 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.02 | 0.04 |
| 20 | 1 | 50.87 | 52.39 | 14.14 | 2.1 | 3.01 | -0.07 | 0.22 | 0.02 | 93.13 | 50.87 | 11.21 | 3.72 | -15.83 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.03 | -0.03 | -0.02 | 0.07 |
| 21 | 1 | 50.94 | 52.35 | 14.14 | 2.1 | 3.02 | -0.07 | 0.23 | 0.02 | 93.06 | 50.94 | 11.25 | 3.72 | -16.05 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.02 | 0.06 |
| 22 | 1 | 51.06 | 52.39 | 14.14 | 2.1 | 3.03 | -0.07 | 0.23 | 0.01 | 92.94 | 51.06 | 11.29 | 3.73 | -16.29 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.02 | 0.09 |
| 23 | 1 | 51.17 | 52.41 | 14.14 | 2.1 | 3.04 | -0.07 | 0.24 | 0 | 92.83 | 51.17 | 11.32 | 3.73 | -16.54 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.02 | 0.11 |
| 24 | 1 | 51.21 | 52.41 | 14.14 | 2.1 | 3.05 | -0.07 | 0.25 | 0 | 92.79 | 51.21 | 11.4 | 3.74 | -16.88 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.02 | 0.11 |
| 25 | 1 | 51.36 | 52.4 | 14.14 | 2.1 | 3.05 | -0.07 | 0.24 | 0 | 92.64 | 51.36 | 11.4 | 3.73 | -17.12 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.02 | 0.1 |
| 26 | 1 | 51.43 | 52.39 | 14.14 | 2.1 | 3.06 | -0.07 | 0.25 | 0.01 | 92.57 | 51.43 | 11.44 | 3.74 | -17.42 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.02 | 0.09 |
| 27 | 1 | 51.54 | 52.35 | 14.14 | 2.1 | 3.07 | -0.07 | 0.24 | 0.02 | 92.46 | 51.54 | 11.47 | 3.74 | -17.74 | 3.72 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.02 | 0.06 |
| 28 | 1 | 51.69 | 52.36 | 14.14 | 2.1 | 3.08 | -0.07 | 0.23 | 0.02 | 92.31 | 51.69 | 11.44 | 3.72 | -17.95 | 3.71 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.02 | 0.07 |
| 29 | 1 | 51.77 | 52.37 | 14.14 | 2.1 | 3.09 | -0.07 | 0.23 | 0.01 | 92.23 | 51.77 | 11.51 | 3.73 | -18.34 | 3.71 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.02 | 0.07 |
| 30 | 1 | 51.84 | 52.37 | 14.14 | 2.1 | 3.09 | -0.07 | 0.23 | 0.01 | 92.16 | 51.84 | 11.55 | 3.73 | -18.7 | 3.71 | 10 | 54 | 69 | 51 | 7 | 29 | 0 | 61 | 2 | -0.02 | -0.03 | -0.01 | 0.07 |

Fig. 8. Data display of the xx_tracks.csv file in the AD4CHE dataset.