A pattern recognition algorithm for assessing trajectory completeness

Anshuman Sharma[a], Zuduo Zheng[a,*], Ashish Bhaskar[b]

[a]*School of Civil Engineering, The University of Queensland, St. Lucia 4072, Brisbane, Australia*
[b]*School of Civil Engineering and Built Environment, Science and Engineering Faculty, Queensland University of Technology (QUT), 2 George St, Brisbane, Qld. 4001, Australia*

Abstract
Vehicular trajectories are widely used for car-following (CF) model calibration and validation, as they embody characteristics of individual driving behaviour (each trajectory reflects an individual driver). Previous studies have highlighted that the trajectories should contain all the major vehicular interactions (driving regimes) between the leader and the follower for reliable CF model calibration and validation. Based on Dynamic Time Warping and Bottom-Up algorithms, this paper develops a pattern recognition algorithm for vehicle trajectories (PRAVT) to objectively, accurately, and automatically differentiate different driving regimes in a trajectory and then select the most complete trajectories (i.e. trajectories containing a maximum number of regimes). PRAVT is rigorously tested using synthetic data and then applied to the NGSIM data. We have observed that the NGSIM data are dominated by the trajectories which contain only three regimes, namely acceleration, deceleration, and following, 77% of the trajectories lack the standstill regime, and no trajectory in the NGSIM data is complete. These findings' impact on how to properly utilize NGSIM data can be profound. Given the extensive use of the NGSIM data in the traffic flow community, this paper also provides insights about the types of regimes contained in each trajectory of the NGSIM data.

Keywords
Trajectory completeness; Dynamic Time Warping; NGSIM; Car-following; Calibration; Validation

1. Introduction

Researchers and Practitioners in traffic engineering heavily depend on trajectory data to understand, analyse, and predict traffic flow phenomena such as car-following (CF) (Gipps, 1981; Hoogendoorn et al., 2011; Newell, 2002; Ossen and Hoogendoorn, 2005; Saifuzzaman et al., 2015), lane changing (Gipps, 1986; Laval and Daganzo, 2006; Wang et al., 2015), traffic oscillations (Laval and Leclercq, 2010; Zheng et al., 2011a, 2011b), queue estimation (Cheng et al., 2011; Liu et al., 2009), capacity analysis (Chen et al., 2014; Srivastava and Geroliminis, 2013), vehicle fuel consumption and emissions (Treiber et al., 2008), and etc.

Vehicle trajectory data (see Figure 1 (a)) are time series of vehicles' positions from which many other useful measures can be easily derived, e.g., individual speed and acceleration profiles, spacing, and time headway. As the trajectory embodies characteristics of individual driving behaviour (each trajectory reflects an individual driver), vehicular trajectories are widely used

* Corresponding author. Tel.: +61 7 3138 9989.
E-mail address: zuduo.zheng@uq.edu.au (Z. Zheng)

for CF model calibration and validation (Brockfeld et al., 2004; Ciuffo et al., 2012; Kesting and Treiber, 2008; Montanino et al., 2012; Monteil et al., 2014; Ossen and Hoogendoorn, 2008; Punzo et al., 2012, 2015; Punzo and Ciuffo, 2009; Treiber and Kesting, 2013a; Vieira da Rocha et al., 2015)

Needless to say, trajectory data play a critical role in CF model calibration and validation, and the quality of trajectory data can significantly influence the results. More specifically, two aspects of trajectory data quality are of interest to CF model developers: data noise and data completeness (hereon trajectory completeness). Data noise (e.g., measurement errors) is inevitable in trajectory data because vehicle trajectories are usually extracted from video cameras and other sensors, which are error-prone. Recently, several studies have focussed on noise in trajectory data (Duret et al., 2008; Hamdar and Mahmassani, 2008; Herrera and Bayen, 2008; Thiemann et al., 2008). Particularly, Ossen and Hoogendoorn (2008) investigated how the measurement errors in the trajectory data impact CF model calibration. In addition, thanks to the popularity of the data collected by Next Generation Simulation program (NGSIM) (NGSIM, 2010) in traffic flow community, notable efforts have been dedicated to data noise in the NGSIM data. For example, Thiemann et al. (2008) first reported the presence of significant errors in NGSIM data, and later Punzo et al. (2011) comprehensively investigated this issue and developed methods to effectively remedy or suppress the impact of data noise in NGSIM on CF model calibration. Moreover, Montanino and Punzo (2015) in a laudable effort reconstructed the first 15-min of I-80 data by filtering out most of the noise and inconsistencies, and the reconstructed data are more suitable for CF model calibration and validation.

Another important aspect of the trajectory data quality is the completeness of information contained in the trajectories. That is, whether or not the trajectory data contain sufficient information for the purpose of CF model calibration. As pointed out by Treiber and Kesting (2013a, 2013b), a trajectory is complete if it constitutes all the 6 driving regimes, namely free acceleration (Fa), cruising at the desired speed (C), following the leader at a constant speed (F), accelerating behind a leader (A), decelerating behind a leader (D), and standing behind a leader (S). In this paper, we adopt the same definition of the completeness of a trajectory.

Previously, researchers have reported that trajectory data used for CF model calibration should contain sufficient observations pertaining to all the longitudinal CF scenarios for reliable calibration, and to capture a complete picture of CF behaviour (Hoogendoorn and Hoogendoorn, 2010; Punzo and Simonelli, 2005; Treiber and Kesting, 2013a). In particular, Treiber and Kesting (2013a) have empirically demonstrated the importance of trajectory completeness on model calibration. As evident from their analysis, model calibration using less complete trajectories results in unrealistic values of parameters relevant to the missing regimes. Punzo et al. (2015) demonstrated that the variance of the simulation error is lower for longer trajectories than for shorter ones, thereby indicating that longer trajectories with different driving regimes should be preferred for model calibration. Furthermore, Sharma et al. (2018) concluded that the average calibrated parameters obtained from complete trajectories perform better in validation and leads to smaller validation errors. They also reported the importance of standing behind a leader regime in CF model calibration and validation.

Aforementioned studies demonstrate the importance of trajectory completeness. However, to the best of authors' knowledge no algorithm has been developed in the literature to objectively and automatically differentiate among different driving regimes contained in a trajectory and then select most complete trajectories for CF model calibration. Motivated by this research need, we develop a robust pattern recognition algorithm for vehicle trajectories (PRAVTVT), which is capable of automatically selecting the complete (or most complete) trajectories from the data, and is the prerequisite for CF model calibration and validation.

Towards this end, driving regimes are clustered as CF (or interaction) or FF (or non-interaction) sections. In the CF section, the following vehicle (hereon follower) responds to the stimulus of leading vehicle (hereon leader). The CF section includes four driving regimes: following the leader at a constant speed, accelerating behind a leader, decelerating behind a leader, and standing behind a leader. Meanwhile, the FF section is that in which the follower's behaviour is independent of the leader's. The FF section includes two driving regimes: free acceleration, and cruising at the desired speed. All the 6 regimes are displayed in Figure 1.

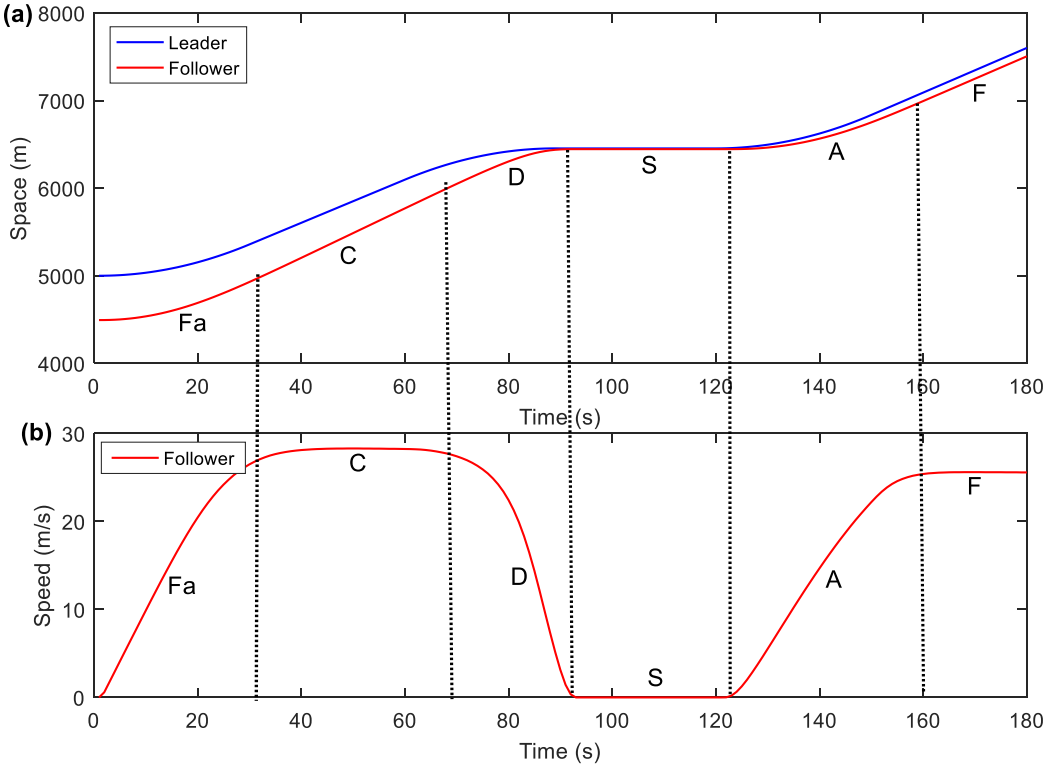


Figure 1 Example of a follower's complete trajectory. a) Driving regimes marked in the follower's trajectory; b) Driving regimes marked in the follower's speed profile.

To unambiguously differentiate different driving regimes, the proposed PRAVT consists of three stages:

- Stage I: Segmentation of the leader-follower trajectories and speed profiles
- Stage II: Distinguishing between CF and FF sections in the follower's trajectory

- Stage III: Identifying regimes present in CF and FF sections. At the end of this stage, driving regimes for each trajectory are identified and each trajectory's completeness is assessed.

The remaining of the paper is organized as follows: Section 2 presents the study sites and data; Section 3 develops the PRAVT in detail. More specifically, Stage I proposes an extended Bottom-Up algorithm, Stage II introduces an algorithm based on Dynamic Time Warping (DTW) to distinguish between the CF and the FF sections, and Stage III presents a method of slopes to identify different driving regimes. Section 4 evaluates the performance of PRAVT and presents findings from PRAVT's implementation on NGSIM data. Finally, Section 5 summarises the major contributions and sheds light on future work.

2. Data

Two datasets are used in this study, i.e., the NGSIM data and a synthetic data, as described below in detail.

2.1 NGSIM data

NGSIM (NGSIM, 2010) vehicle trajectory data from the I-80 and US-101 were collected at Emeryville and Los Angeles California. The geometric configurations of both the sites are shown in Figure 2. As discussed previously, notable noises and errors exist in the original NGSIM data. Recently, Montanino and Punzo (2015) conducted an excellent investigation on the accuracy of the original NGSIM data, and reconstructed a portion of the vehicle trajectories on I-80 (the first 15 minutes, 4:00 p.m. to 4:15 p.m.). Zheng et al (2011a) has also denoised both the I-80 and US-101 data, and the denoised data have been used in many recent publications (Chen et al., 2014, 2012b; Saifuzzaman et al., 2017; Zheng et al., 2013, 2011a). In this study, to take advantage of the best resources available, both the first 15-min I-80 data reconstructed by Montanino and Punzo (2015) and the denoised US-101 data by Zheng et al. (2011a) are used. As another strategy to diminish the impact of data noise and errors, this study avoids the use of accelerations because of their notorious inaccuracy in the NGSIM data. In addition, the HOV lane (Lane 1) on I-80 and the auxiliary lane on US-101 are excluded from our analysis because different driving behaviour is likely to prevail in these lanes as compared to driving behaviour in other lanes. Note that NGSIM data also include data collected at Lankershim Boulevard in Los Angeles, and at Peachtree Street in Atlanta. In this research, the term "NGSIM data" refers to the reconstructed I-80 and US-101 data only.

2.2 Synthetic data

Although real trajectories are invaluable for CF model calibration and validation, it can be difficult to objectively and comprehensively assess a CF model's performance based on real trajectories for three primary reasons: i) true parameters for generating real trajectories are unknown; ii) variety (corresponding to driving behaviour diversity) in real trajectories is often limited and certainly cannot be manipulated; and iii) noises and errors often plague real trajectories. To address these issues, synthetic data are sometimes used in the literature (Ciuffo et al., 2012; Montanino et al., 2012). To generate synthetic trajectories, a CF model of interest

is first selected and the corresponding model parameters are predetermined. Then this model is used to simulate a follower's trajectory in response to a leader's.

Evidently, using synthetic data enables researchers to compare the calibrated model parameters with the ground truth (i.e., the predetermined parameter values), which would be impossible with real trajectories. Moreover, using synthetic data can allow researchers to flexibly manipulate the presence of different driving regimes either in a noise-free environment or in a noise-controlled (i.e., a certain type of noise is intentionally added) environment, which is particularly important for this study. Thus, synthetic trajectories are generated and used in this study.

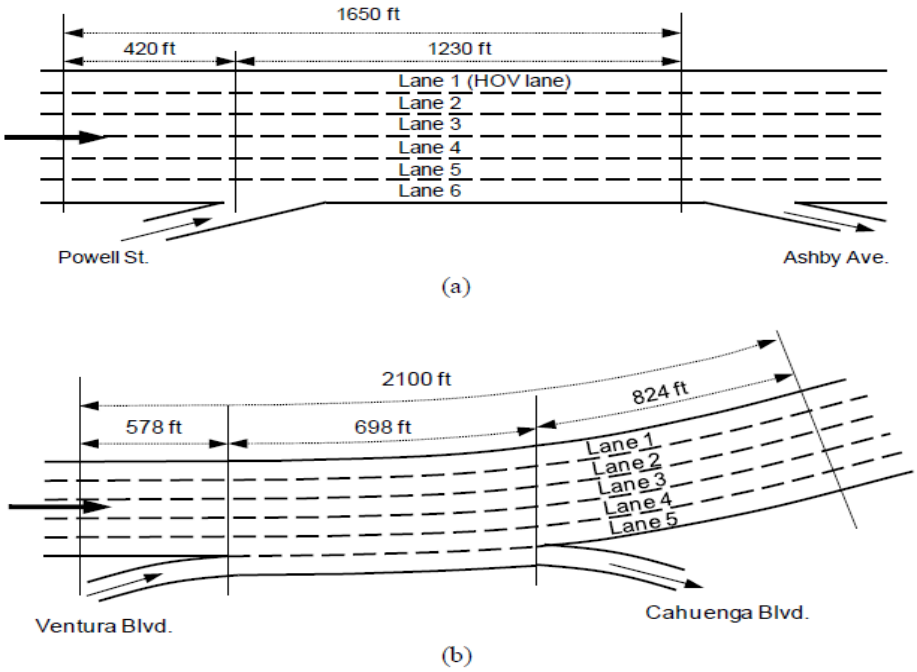


Figure 2 Schematic diagram of study sites. (a) I-80; (b) US-101.

3 PRAVT development

Both the leader's and the follower's trajectories and their corresponding speed profiles are required to assess the completeness of a follower's trajectory. The proposed PRAVT algorithm consists of three stages which are described below.

3.1 Stage I: Segmentation of the leader-follower trajectories and speed profiles

At first, the speed profile associated with a given trajectory is segmented using segmentation algorithm. Segmenting the speed profile assists in a clearer and more accurate representation of different regimes, thereby facilitating easier identification of these regimes.

Piecewise linear representation of a time series is called segmentation, and the algorithms which perform this transformation are referred to as segmentation algorithms (Keogh et al., 1993). The segmentation algorithms can be grouped into three categories: Sliding Windows, Top-Down, and Bottom-Up. The Bottom-Up algorithm is used in this study because it often outperforms Top-Down and Sliding Windows algorithms (Keogh et al., 1993). Bottom-Up algorithm starts with segmenting the time series in the finest possible segments (e.g., a single

data point) and then merging neighbouring segments until the error of the merged segment reaches the tolerance limit. Of note, the tolerance limit (also called as a stopping criterion) is a threshold error defined by the user. Another stopping criterion is the maximum number of segments. If the maximum number of segments is "N" then the segmentation algorithm produces the best piecewise linear representation with N number of segments.

The following steps summarise the procedure of the Bottom-Up algorithm:

a) Divide the time series into the finest possible approximations and calculate the cost of merging each pair of adjacent segments;
b) Identify the pair of segments with the minimum cost, and merge them;
c) Update the cost of merging each pair of adjacent segments;
d) Repeat steps (b) and (c) until the stopping criterion is met.

In this study, the maximum number of segments is adopted as the stopping criterion. Importantly, one should judiciously decide the maximum number of segments for a particular time series as this factor can have a significant impact on the result (Keogh et al., 1993). To avoid the ambiguity in deciding the optimal number of segments, and to obtain the segmented speed profile which best suits our investigation, we extend the Bottom-Up algorithm as illustrated in the steps below:

i) Follow the steps (a) to (d) of the Bottom-Up algorithm by choosing a maximum number of segments as the stopping criterion. The maximum number of segments is selected by using the average segment length, as further discussed later.
ii) Merge the successive segments if they are of the same slope (i.e., the same driving regime), e.g., if the successive segments are acceleration segments, then these segments are merged until a segment that belongs to a different driving regime is encountered.

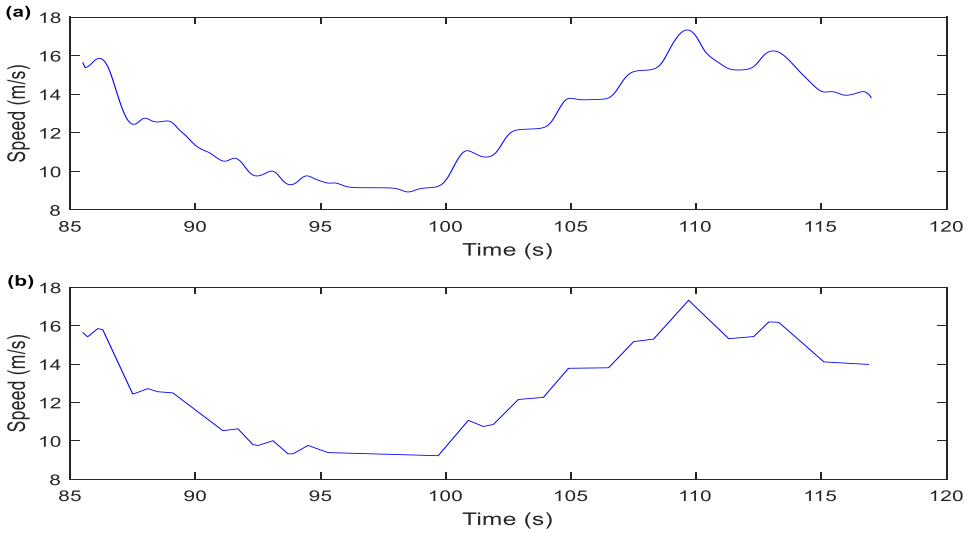


Figure 3 Speed profile a vehicle. a) Original speed profile; b) Segmented speed profile by implementing the extended Bottom-Up algorithm.

Obviously, the maximum number of segments depends on the average length of one segment (the maximum number of segments = total length of the time series/average length of one segment). In our analysis, we ensured that the average segment length should not be too small. Otherwise, the driving regime reflected in a segment can be too short to be used to calibrate

CF model. Moreover, the smaller the segment length, the greater the probability for DTW used at Stage II to encounter the singularity problem (see Section 3.2 & Appendix B for more discussion).

By considering above mentioned two factors and after a rigorous analysis of trial and error, 0.5 s is adopted as the average segment length for the NGSIM data. For demonstration purpose, the extended Bottom-Up algorithm is applied to the speed profile of a vehicle from the reconstructed I-80 data. The original and the segmented speed profiles are shown in Figure 3. After segmenting the speed profiles, the inflection points are transferred to the associated trajectories to obtain the segmented trajectories.

3.2 Stage II: Distinguishing CF and FF sections in the follower's trajectory

Once the segmented trajectory is obtained, the next step is to distinguish CF and FF sections present in the trajectory.

Previously, researchers have considered using a headway (space or time) threshold to distinguish between CF and FF sections by assuming either a deterministic value (Aycin and Benekohal, 1998; Herman and Potts, 1961; Panwai and Dia, 2007; Punzo et al., 2011) or a distribution (Ahmed, 1999; Subramanian, 1996; Toledo, 2003). The significant limitation of assuming a deterministic value is its incapability to capture the inter-driver heterogeneity. Assuming a distribution can overcome this limitation. Often headway threshold is assumed to follow a truncated normal distribution, and the distribution parameters are estimated jointly with other model parameters. Obviously, the parameters obtained by this method are dependent on the model and estimation setting, hence, they are likely to change with the change in the model or estimation setting. Furthermore, it is unclear whether this method is consistent with the behavioural characteristics of CF and FF driving or not. To overcome these limitations, this research proposes a methodology based on the fundamental relationship of CF i.e. stimulus-response (S-R) relationship. Particularly, we also introduce time gap (τ_n for n^{th} vehicle) as a new metric to define the threshold because of its agreement with the S-R relationship and the Newell's theory (Newell, 2002). As in Newell's theory, τ_n is the time taken by the follower to adapt to the change in the leader's speed.

To obtain a threshold, the major challenge is to detect S-R relationship in the trajectory data. To overcome this, we adopted DTW algorithm for reasons discussed below.

In time series analysis, DTW is an algorithm for measuring the similarity of two time series by achieving a better alignment between two temporal sequences which may vary in time. Moreover, DTW allows "elastic" transformations which minimise the effects of shifts and distortion in time to cater for the time series of different time lengths (Senin, 2008). Some well-known applications of DTW are in the fields of speech recognition (Rabiner and Juang, 1993), manufacturing (Gollmer and Posten, 1995), data mining (Keogh and Pazzani, 1998), robotics (Schmill et al., 1999) and medicine (Caiani et al., 1998).

In the field of transportation engineering, the full potential of DTW is still to be explored. DTW has been employed to detect characteristic points in the leader's and the follower's trajectories (Przybyla et al., 2012) and for lane change calibration (Nishiwaki et al., 2012). Taylor et al. (2015) calibrated the Newell's CF model by applying DTW. However, previous studies which used DTW for extracting the S-R relationship (Przybyla et al., 2012; Taylor et al., 2015) simply

assumed that DTW is capable of doing so. To the best of our knowledge, none of these studies has verified whether DTW indeed has such capability or not. Therefore, to fill this important gap, before utilizing DTW, this study has rigorously assessed its capability of accurately extracting the S-R relationship.

3.2.1 DTW algorithm

Before introducing the algorithm, the terminologies used in the algorithm are defined below:

Time series: A sequence of data measured sequentially in time. For instance, time series of speed, position, acceleration, etc.

Optimal distance: The shortest distance between two points or two time series.

Cost: A local distance measure between elements of two time series.

Cost matrix: A matrix that contains the cost values computed for each element of the time series.

Cell: an element of the cost matrix.

The primary objective of DTW is to find the optimal distance between two time series, e.g., $A = (a_1, a_2, a_3, \dots\dots\dots, a_l)$ of length $l \in \mathbb{N}$ and $B = (b_1, b_2, b_3, \dots\dots\dots, b_m)$ of length $m \in \mathbb{N}$.

The first step of determining the optimal distance is to compute the cost matrix. The cost matrix contains cost values (C), which signifies how different a_i (an element of A) is from b_j (an element of B). If C is small (low cost), then a_i and b_j have some similar features and vice versa. A commonly chosen cost measure is the Euclidean distance as shown in Equation (1).

$$C(i,j)_{1\le i\le l, 1\le i\le m} = \|a_i - b_j\| = \sqrt{(a_i - b_j)^2} \tag{1}$$

The cost C is calculated for each pair of data points between each data point of A and all the points of B.

The next step is to calculate the cumulative cost matrix containing the cumulative least cost required to arrive at any cell when traversing from the first cell to the last cell. The cumulative least cost of a cell is given in Equation (2).

$$If\ i = 1, j = 1;\ C^{cum}(1,1) = C(1,1)$$

$$If\ i = 1;\ C^{cum}(1,j) = C(1,j) + C^{cum}(1,j-1)$$

$$If\ j = 1;\ C^{cum}(i,1) = C(i,1) + C^{cum}(i-1,j)$$

$$Otherwise$$

$$C^{cum}(i,j) = C(i,j) + min(C^{cum}(i-1,j-1), C^{cum}(i,j-1), C^{cum}(i-1,j)) \tag{2}$$

The last step is to find the optimal warping path, which is the shortest path joining matrix elements of the cumulative cost matrix C^{cum} and thereby identifying the optimal mapping between the elements of the time series A and B. A warping path is denoted as $W_p = (w_1, w_2, w_3, \dots, w_k, \dots\dots, w_L)$ where w_L is cell (l, m) and w_1 is cell $(1, 1)$, and w_k is a generic cell in the warping path representing cell (i, j) of the matrix C^{cum}, and L is the length of the warping path. W_p needs to satisfy the following criteria:

a) **Boundary condition:** The starting and ending points of the cumulative cost matrix should be the start and end points of the warping path. This guarantees the entire time series alignment, as the two time series might be of different lengths. Furthermore, this avoids that the algorithm only partially considers any series.
b) **Monotonicity condition:** The warping path must step forward and cannot step backwards. This retains the time-ordering of all the points in the two time series.
c) **Step size condition:** $w_k - w_{k-1} \in \{(1,0),(0,1),(1,1)\}$ for all $k \in [1:L]$. The warp path must step forward by one time step only. This ensures that the algorithm does not miss any important feature by jumping more than one time step.

The optimal warping path is traced from cell (l,m) to cell$(1,1)$. However, there can be many warping paths satisfying the aforementioned conditions (a), (b), and (c). The optimal warping path is the one that minimizes the warping cost i.e., $min \sum_{k=1}^{L} w_k$. Furthermore, DTW distance (the optimal distance) between A and B is given by the minimum warping cost, i.e. $DTW\ (A,B) = min \sum_{k=1}^{L} w_k$. For illustrative examples of DTW algorithm application, see Appendix A and Taylor et al. (2015). For more detail on DTW, refer to Albrecht and Muller (2009), Berndt and Clifford (1994), and Keogh and Ratanamahatana (2005).

Importantly, in the case of trajectory data, DTW assists in identifying a pair of similar points (optimal mapping) on two time series of positions (or speed) via elements of the optimal warping path.

3.2.2 Assessing DTW

Synthetic data are generated to assess DTW's capability of extracting S-R relationship. The following paragraph describes the experiment design to generate the synthetic data.

The experiment involves two vehicles in heavy traffic: the leader $n-1$ and the follower n. At the start of the experiment ($t = 0\ s$) the leader is standing in a queue at a traffic signal and it's $20\ m$ ahead of the reference point (leader's position $x_{n-1}(0) = 20\ m$, speed $v_{n-1}(0) = 0\ m/s$, and acceleration $a_{n-1}(0) = 0\ m/s^2$). With the onset of the green light, at $t = 0.1\ s$, the leader starts moving with a constant $a_{n-1}(0.1) = 1\ m/s^2$ and maintains this acceleration until $t = 5\ s$. At this time, $x_{n-1}(5) = 32.5\ m$ and $v_{n-1}(5) = 5\ m/s$. From $t = 5.1\ s$ to $t = 10\ s$, the leader travels with $a_{n-1}(t = 5.1\ to\ t = 10) = 0\ m/s^2$. In the remaining experiment the leader maintains three constant accelerations $1\ m/s^2$, $0\ m/s^2$, and $-1\ m/s^2$ (deceleration) at different times, respectively. The experiment ends at $t = 60\ s$.

At $t = 0\ s$ the follower is standing at the reference point (the follower's position $x_n(0) = 0\ m$, speed $v_n(0) = 0\ m/s$, and acceleration $a_n(0) = 0\ m/s^2$). The follower starts moving after 2 s and its response to the leader's movement is according to Newell's CF model (Newell, 2002). Note that any CF model can be used to generate the follower's response.

Newell postulated that under homogeneous and congested traffic state the follower's trajectory is a translation of the leader's trajectory by a space gap d_n and τ_n, while in FF vehicles always travel at the desired speed V_0, as mathematically formulated in Equation (3).

$$x_n(t+\tau_n) = \min\{\underbrace{x_n(t) + V_0 \times \tau_n}_{free-flow}, \underbrace{x_{n-1}(t) - d_n}_{congestion}\} \quad (3)$$

Where $x_n(t)$ represents the position of the follower n at time t and $x_{n-1}(t)$ represents the position of the leader $n-1$ at time t.

In this experiment, the values of parameters τ_n and d_n are 2 s and 5 m, respectively. Figure 4 presents the speed profiles of the leader and the follower. Additionally, Figure 4 depicts the actual S-R points (Ground truth). More specifically, the inflection points A, B, C, D, E, F and G represent stimulus points in the leader's speed profile, and $A', B', C', D', E',\ F'$ and G' represent the response points where the follower has adapted to the speed change of the leader.

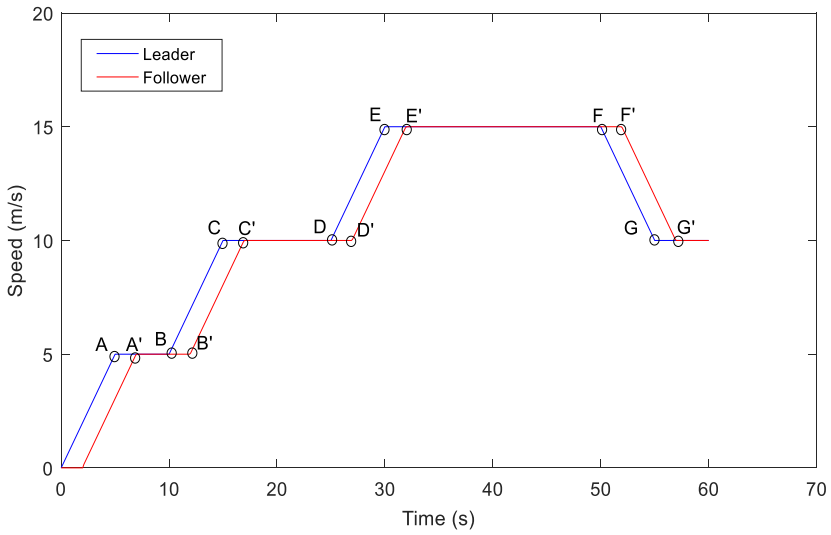


Figure 4 Synthetic leader-follower speed profiles with S-R inflection points.

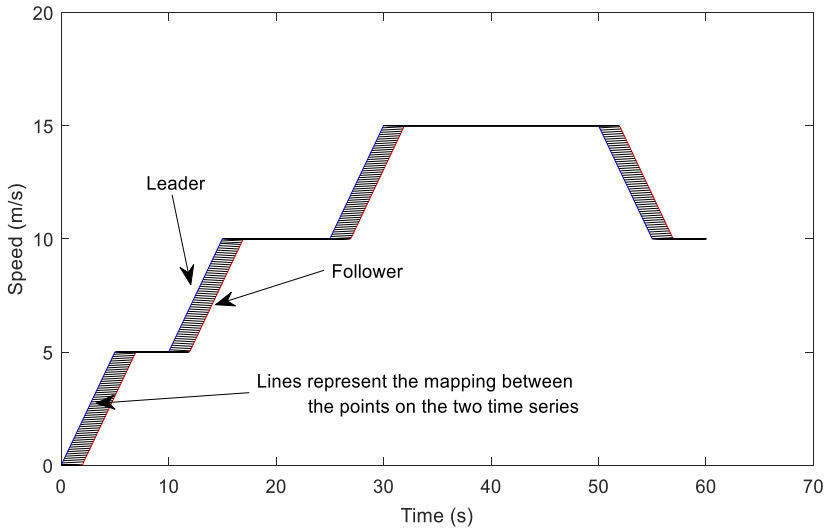


Figure 5 DTW match solution for synthetic leader-follower speed profiles.

To test DTW's capability of extracting these S-R points, we applied DTW to the leader's and the follower's speed time series as well as time series of positions (i.e. trajectories). The similar points on the two time series are depicted by the line joining these points on the time series of speed (see Figure 5). The point on the leader's time series is the estimated stimulus point and a time-delayed similar point at the other end of the line is the follower's estimated response point.

A comparison between the ground truth and the estimated S-R points is illustrated in Table 1. As shown in Table 1, the small errors in stimulus, response, and time gap clearly depict that the simulated S-R pairs closely match the observed ones, which confirms DTW's excellent performance of extracting S-R relationships. Note that DTW has also performed well when the trajectories instead of the speed profiles are given as input.

Table 1 Comparison of ground truth and estimated S-R points, and time gaps (noise free synthetic data).

Ground truth			**Estimated**			**Error**		
Stimulus (time, s)	**Response (time, s)**	**Time gap (s)**	**Stimulus (time, s)**	**Response (time, s)**	**Time gap(s)**	**Stimulus (s)**	**Response (s)**	**Time gap (s)**
5	7	2	4.90	6.90	2	0.1	0.1	0
10	12	2	9.90	11.90	2	0.1	0.1	0
15	17	2	14.90	16.90	2	0.1	0.1	0
25	27	2	24.90	26.90	2	0.1	0.1	0
30	32	2	29.90	31.90	2	0.1	0.1	0
50	52	2	49.90	51.90	2	0.1	0.1	0
55	57	2	54.90	56.80	1.9	0.1	0.2	0.2

All the values are in seconds. The time gap is the difference between stimulus and response points on the time-axis

In the real world, the trajectory data are often contaminated by noise. Therefore, noise at different signal-to-noise ratio (SNR) levels is added to the synthetic data to more convincingly test DTW's capability to extract the S-R relationship. A noise level SNR15 depicts that the signal-to-noise ratio per sample is 15 dB. Note that as SNR increases, noise in the data decreases. To be comprehensive, numerous scenarios are designed, as elaborated below.

First, noise is added to the position values of both the leader and the follower, and then their speed profiles are derived. The noise level for the leader's trajectory is fixed at zero (no noise), while the noise level for the follower varies at each run. Seven noise levels are considered, i.e., no noise, SNR15, SNR20, SNR25, SNR30, SNR35 and SNR40. Next, we applied DTW to the speed profiles for each combination of the leader-follower couple. After this, we changed the SNR level in the leader's trajectory and repeated the procedure above. The SNR levels considered for the leader include SNR15, SNR20, SNR25, SNR30, SNR35 and SNR40 in addition to no noise. In total, 49 scenarios (i.e., 7×7) are generated and assessed.

For the demonstration purpose, we present the results for the scenario where the SNR for both the leader and the follower is 35 (see Figure 6(a)). Figure 6 (b) shows the S-R relationship mapping on their speed profiles, and Table 2 summarises the comparison between the observed and the estimated S-R points. Clearly, as evidenced by the small errors in Table 2, the S-R relationship extracted by DTW is acceptable even when the data are noisy. Note that we obtained an acceptable DTW mapping for the scenarios where the SNR level for both the leader and the follower is 30 or more. Below 30, the data become too noisy.

Overall, the aforementioned analysis shows that DTW is an effective tool to extract the S-R relationship from trajectories. However, DTW mapping for a pair of trajectories occasionally results in unrealistic values of τ_n and d_n. In addition, at some points the mapping can encounter singularities (i.e., one-to-many mapping). One of the primary reasons for singularities in DTW

is high frequency data. The singularity issue in applying DTW has been reported in Taylor et al. (2015), and how to deal with this issue is explained in Appendix B.

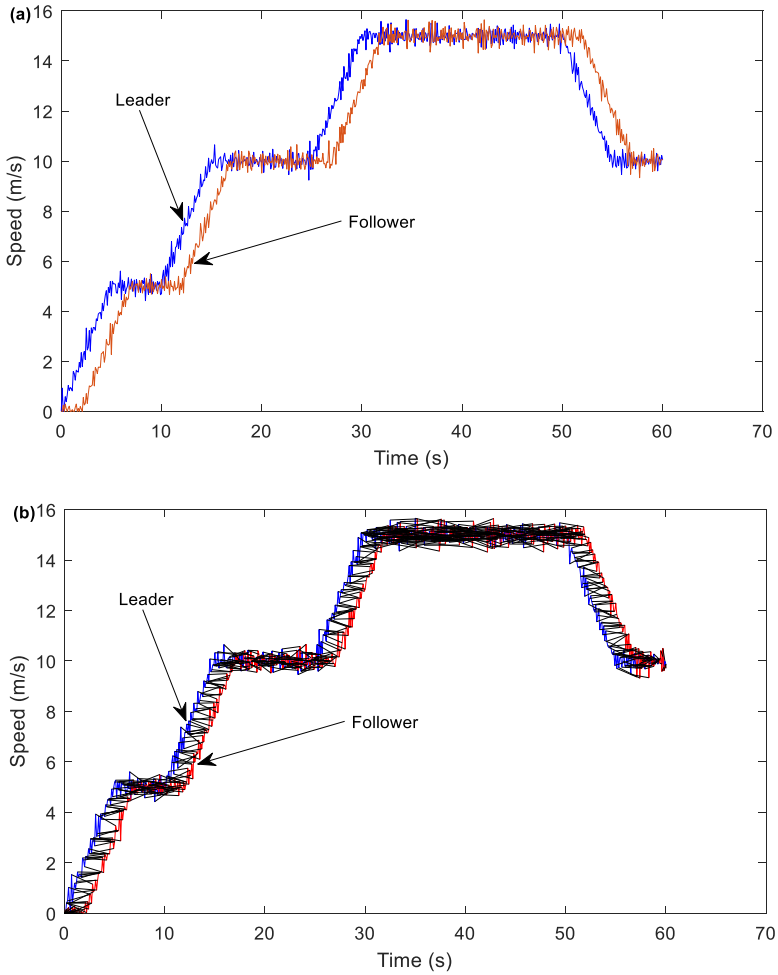


Figure 6 DTW match solution. a) Noisy (SNR 35) synthetic leader-follower speed profiles; b) DTW mapping on noisy synthetic leader-follower speed profiles.

Table 2 Comparison of the ground truth and the estimated S-R points, and time gaps of noisy synthetic data (SNR 35).

Ground truth			Estimated			Error		
Stimulus (s)	**Response (s)**	**Time gap (s)**	**Stimulus (s)**	**Response (s)**	**Time gap(s)**	**Stimulus (s)**	**Response (s)**	**Time gap (s)**
5	7	2	4.7	6.8	2.1	0.3	0.2	0.1
10	12	2	10.3	12.4	2.1	0.3	0.4	0.1
15	17	2	15.3	17.9	2.6	0.3	0.9	0.6
25	27	2	25.3	27	1.7	0.3	0.0	0.3
30	32	2	30.1	32.1	2	0.1	0.1	0
50	52	2	49.9	52	2.1	0.1	0.0	0.1
55	57	2	54.5	56.2	1.7	0.5	0.8	0.3

All values are in seconds

3.2.3 Distinguishing between CF and FF

This section introduces a DTW-based method to determine a threshold for reliably distinguishing between the CF and the FF. This method is inspired by Newell's CF theory (Newell, 2002). One of the two parameters in Newell's CF theory, τ_n , is utilised to distinguish CF from FF because τ_n represents the follower's interaction with the leader. Roughly, a small τ_n indicate the follower responds to the leader (i.e., CF), and a large τ_n may indicate the negligible influence of the leader on the follower (i.e., FF). More specifically, the method consists of 5 main steps, as listed below:

a) Apply DTW on leader-follower trajectories to obtain the optimal mapping;
b) Calculate τ_n of a vehicle n for each pair of the mapping points[a];
c) Obtain the mean μ_τ and standard deviation σ_τ of τ_n values;
d) If $\mu_\tau > \mu_{lim}$ or $\sigma_\tau > \sigma_{lim}$, the threshold is μ_τ; otherwise, the threshold is $\mu_\tau + 2\,\sigma_\tau$;
e) At any point, if τ_n is below the threshold, then this point on the follower's trajectory belongs to CF, otherwise FF.

Here, μ_{lim} and σ_{lim} are the upper limits of mean time gap (μ_τ) and standard deviation of time gap (σ_τ). While applying the above methodology one might confront a case where one FF (CF) mapping point exists in between two CF (FF) mapping points (for simplicity, we call such points as isolated points). These isolated points are updated to the same as the neighbouring points unless the time difference between an isolated point and a neighbouring point is sufficiently large (e.g., larger than 5 seconds). Note that if the data are collected at low frequency, Stage II can be applied before Stage I, however, data shall necessarily pass through both Stage I and Stage II before entering stage III.

3.3 Stage III: Identifying regimes present in CF and FF sections

After CF and FF sections are distinguished using Stage II, the next step is to detect the presence of specific driving regimes, i.e., the accelerating behind the leader, decelerating behind the leader, following the leader at constant speed, and standing behind the leader regimes in CF sections, and free acceleration and cruising at desired speed regimes in FF sections.

The procedure to identify the types of regimes present in each section is schematically presented in Figure 7, and called as the method of slopes. As described in Figure 7, segmented speed profile of the follower specific to CF and FF sections are provided as input, and then the segments are divided into two groups: one with positive slopes and one with negative slopes. To check for the constant speed segments, the definition given by Ozaki (Ozaki, 1993) is followed. According to the definition, a regime is a steady-state (i.e., constant speed) regime if the acceleration or deceleration rate is within 0.05 g (g is the gravitational acceleration). Hence, the segments with a slope between 0.5 to -0.5 are classified as constant speed segments. Furthermore, from these constant speed segments, the segments with zero speeds are classified as standstill segments.

When the method of slopes is applied to FF sections, the acceleration segments represent Fa regimes and constant speed segments represent C regimes. Obviously, FF sections will be devoid of standstill and deceleration segments. When the method is applied to CF sections, the acceleration segments represent A regimes, constant speed segments represent F regimes,

[a] τ_n values lower than 0.01 s and d_n lower than 0.01 m are removed because they are very likely to be outliers.

deceleration segments represent D regimes, and standstill segments represent S regimes. A trajectory is complete if all the regimes are present.

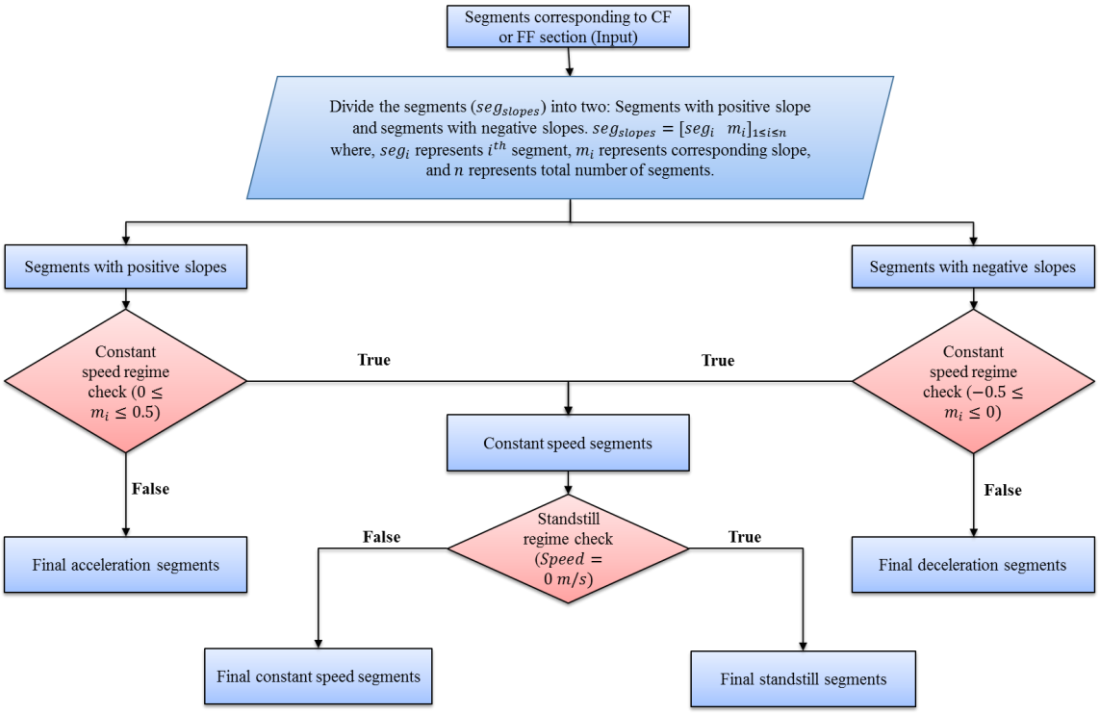


Figure 7 Flow chart describing method of slopes.

Importantly, while assessing the trajectory completeness all the three stages are applied to each leader-follower pair available in the data. Note that influence of road geometry, type of lead vehicle, grade, or weather conditions are trivial for trajectory completeness. For example, irrespective of the type of leader (heavy vehicle or a car) or any weather condition, a following vehicle will experience any or all of these driving regimes during the course of his/her drive. In addition, lane changing (see Zheng (2014) for a review) is regarded as a primary driving task rather than a driving regime of a CF process.

4 PRAVT evaluation and implementation

4.1 Evaluating the performance of PRAVT using synthetic data

4.1.1 Synthetic data generation

100 trajectories are randomly drawn from the reconstructed I-80 data that serve as leaders' trajectories. The followers' trajectories are generated using Gipps' CF model (Gipps, 1981). Other CF models such as Intelligent Driver Model (Treiber et al., 2000), can also be used. Equations (4), (5), and (6) provide the mathematical formulation of Gipps' CF model.

$$V_{FF} = V_n(t) + 2.5a\tau\left(1 - \frac{V_n(t)}{V_0}\right)\left(0.025 + \frac{V_n(t)}{V_0}\right)^{1/2} \tag{4}$$

$$V_{CF} = b\tau + \sqrt{(b\tau)^2 - b[2(\Delta X_n(t) - s_0 - l) - V_n(t)\tau - \frac{V_{n-1}(t)^2}{\hat{b}}} \quad (5)$$

$$V_n(t+\tau) = \min\{V_{FF}, V_{CF}\} \quad (6)$$

where a is the desired acceleration $(\mathrm{m/s^2})$, b is the desired deceleration $(\mathrm{m/s^2})$, $\Delta X_n(t)$ is the spacing between the leader and the follower at time t, s_0 is the standstill distance (m), l is the length of the leader (m), $\hat{b}$ is an estimate of the deceleration applied by the leader $(\mathrm{m/s^2})$, V_0 is the desired speed $(\mathrm{m/s})$, τ is the reaction time (s), and $V_{n-1}(t)$ is the speed of the $(\mathrm{n}-1)^{th}$ vehicle (leader) at time t $(\mathrm{m/s})$. $V_n(t)$ is the speed of the n^{th} vehicle (follower) at time $(\mathrm{m/s})$, V_{FF} is the speed in FF $(\mathrm{m/s})$ and V_{CF} is the speed in CF $(\mathrm{m/s})$. As evident from Equations (6), the Gipps' model switches from the FF speed to the CF speed and vice versa, therefore, it explicitly defines transitions between FF and CF sections.

For each leader, 50 followers are simulated using Gipps' model resulting in 50 leader-follower pairs. The Gipps' parameters for the 50 followers are sampled using a Latin Hypercube Design (LHD). In LHD, parameters are randomly sampled from the entire parameter range divided into segments of equal probabilities. Refer to McKay et al. (1979) for a discussion on Latin Hypercube, and Park and Qi (2005) for details on experiment design using LHD in the field of transportation engineering. The Gipps' parameters ranges in LHD are $V_0 \in [15, 30]$, $\tau \in [0.1, 3]$, $s_0 \in [1, 10]$, $a \in [0.1, 4]$, $b \in [0.1, 4.5]$, and $\hat{b} \in [0.1, 4.5]$. Note that l is fixed to 4 m.

Real trajectories are utilised as leaders because they contain randomly arranged different driving regimes, which is essential for the rigorous testing of PRAVT. Moreover, the space/separation between the starting positions of the leader and the follower directly influences the type of driving regimes that can be generated afterwards in the follower's trajectory. The separation values for the 50 simulated followers are sampled from the range $[10, 1000]$ using LHD. As a result, 5000 leader-follower couples (100×50) have been generated, containing different motion phases in the leaders' trajectories and different percentages of CF and FF regimes in the followers' trajectories.

4.1.2 Description of the performance measures and PRAVT implementation

This study employs the following performance measures to evaluate the efficacy of PRAVT:

a) CF present: Percentage of the CF present in the trajectory (Equation (7));
b) FF present: Percentage of the FF present in the trajectory (Equation (8));
c) True CF estimated: Percentage of the true CF estimated by PRAVT (Equation (9));
d) True FF estimated: Percentage of the true FF estimated by PRAVT (Equation (10)).

$$CF\ present = \frac{L_{GT}^{CF}}{L_{GT}^{Total}} \times 100 \quad (7)$$

$$True\ CF\ estimated = \frac{L_{EST:GT}^{CF}}{L_{EST}^{CF}} \times 100 \quad (8)$$

$$FF\ present = \frac{L_{GT}^{FF}}{L_{GT}^{Total}} \times 100 \quad (9)$$

$$True\ FF\ estimated = \frac{L_{EST:GT}^{FF}}{L_{EST}^{FF}} \times 100 \quad (10)$$

L_{GT}^{Total} = Length of the trajectory (s)

L_{GT}^{CF} = Length of CF section present in the trajectory i.e. ground truth of the CF section (s)
L_{GT}^{FF} = Length of FF section present in the trajectory i.e. ground truth of the FF section (s)
L_{EST}^{CF} = Length of the CF section estimated by PRAVT (s)
L_{EST}^{FF} = Length of the FF section estimated by PRAVT (s)
$L_{EST:GT}^{CF}$ = Length of CF section estimated by PRAVT that matches with ground truth of the CF section (s)
$L_{EST:GT}^{FF}$ = Length of FF section estimated by PRAVT that matches with ground truth of the FF section (s)

When implementing PRAVT on a leader-follower pair, μ_{lim} and σ_{lim} (see step (d) of the procedure to distinguish between CF and FF sections in Section 3.2) are also sampled from the ranges $[1, 10]$ and [0.5, 5], respectively, using LHD. As 50 followers are generated for each leader, 50 samples are drawn for μ_{lim} and σ_{lim}.

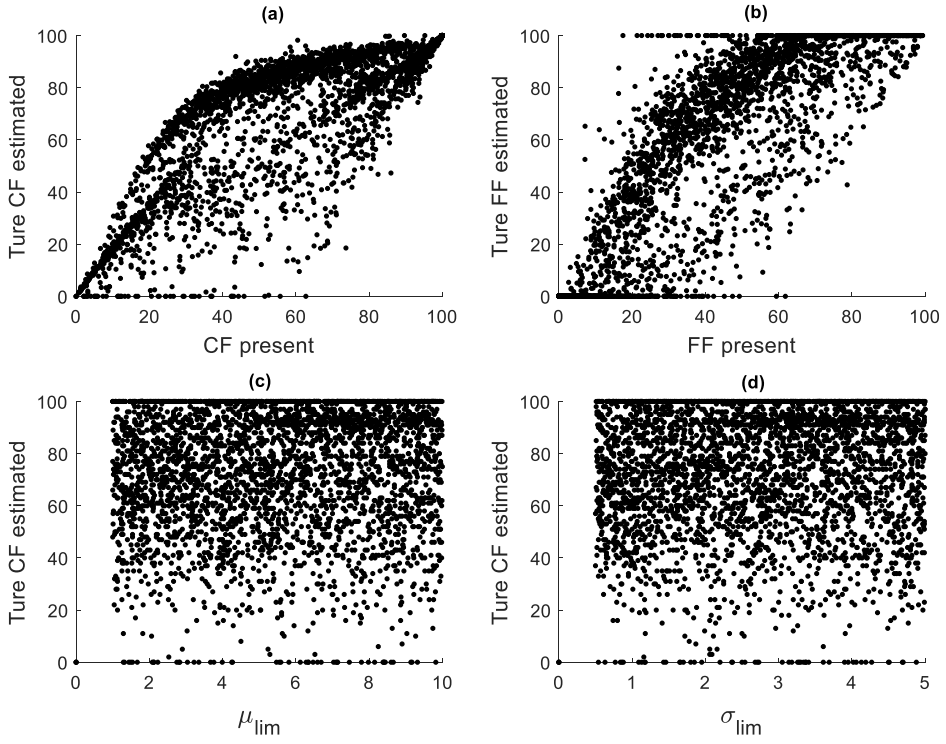


Figure 8 Plots of PRAVT performance measures. a) True CF estimated versues CF present; b) True FF estimated versues FF present; c) True CF estimated versues μ_{lim}; and d) True CF estimated versues σ_{lim}.

The performance measures are calculated for all the 5000 leader-follower pairs and are depicted in Figure 8. In general, the true CF and the true FF estimated increase with the increase in the percentages of CF and FF sections, respectively (Figures 8(a) and 8(b)). PRAVT estimates at least 70% of the true CF when the CF section in the trajectory is 30% or above. Similarly, PRAVT estimates at least 70% of the true FF when the FF section in the trajectory is 40% or above. The results reveal high recovering capability, accuracy, and reliability of PRAVT, and thereby validate PRAVT's effectiveness. After estimating CF and FF sections, PRAVT identifies the regimes accurately as demonstrated by a typical example in Figure 9.

Figure 8(c) and 8(d) depict that for a given μ_{lim} or σ_{lim} the true CF estimated ranges from 40% to 100% in most of the cases. Note that similar findings are observed for the true FF estimated also (not shown in the plots). These findings reveal the importance of choosing an appropriate μ_{lim} and σ_{lim} for a given data for extracting best results from PRAVT.

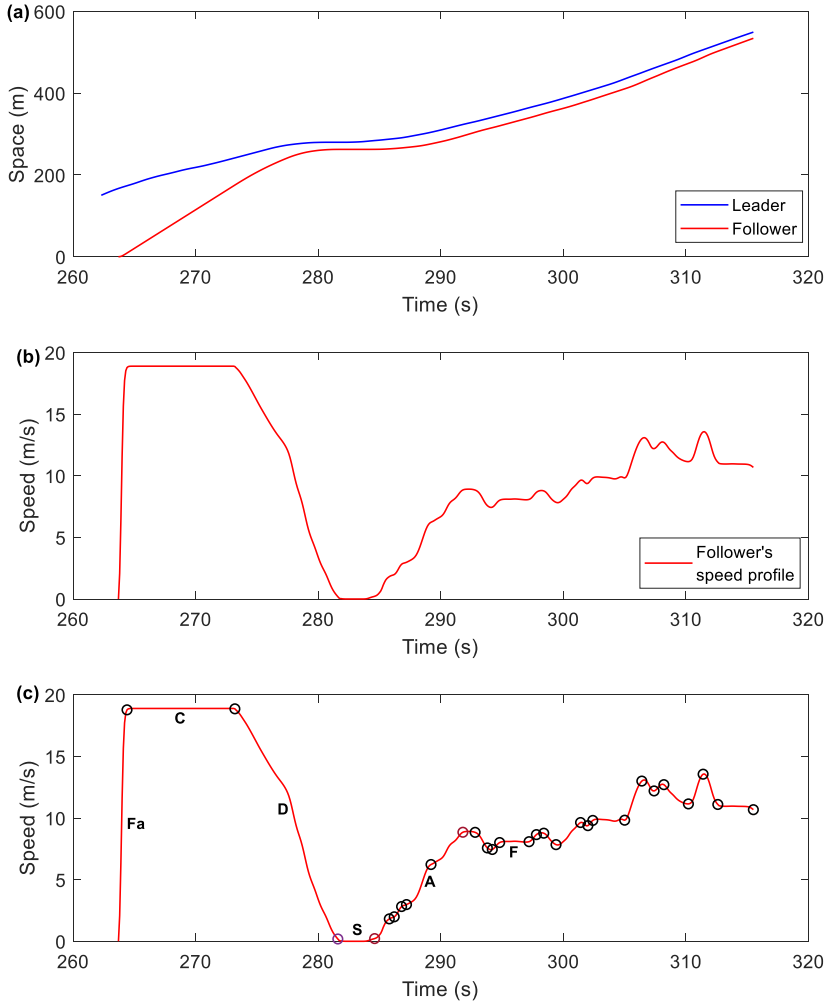


Figure 9 PRAVT applied to the synthetic trajectory. a) the trajectory input to PRAVT; b) the speed profile of the follower; and c) the output of PRAVT with regimes identified. Note that in (c) circles show the end points of the regimes, Fa and C are observed once, and for clarity, the names of the other identified regimes are displayed only once.

Nevertheless, PRAVT can recover at least 40% of the true CF and FF in most of the cases even with an inappropriate μ_{lim} and σ_{lim}. The next section presents a methodology to calculate μ_{lim} and σ_{lim}.

4.1.3 Calculating $\boldsymbol{\mu_{lim}}$ and $\boldsymbol{\sigma_{lim}}$

Calibrate Newell's CF model (Newell, 2002) using the trajectory pairs in the data. Next, calculate the mean time gap (μ_τ) and the standard deviation of time gap (σ_τ) using the calibrated time gap (τ) values. The limit μ_{lim} is equal to $\mu_\tau + 2\sigma_\tau$ and σ_{lim} is equal to σ_τ.

To demonstrate the above methodology, we randomly select 10 trajectory pairs from the previous numerical experiment. The limits μ_{lim} and σ_{lim} are calculated, and PRAVT is applied on the selected trajectory pairs using the calculated limits. The results presented in Table 3 are

consistent with the findings mentioned in the previous section, i.e., the larger the CF present and the FF present are, the larger the true CF and the true FF estimated become. Importantly, the true CF and true FF recoveries are above 80% when the CF present and the FF present are larger than 30%, respectively. Hence, the proposed methodology to calculate μ_{lim} and σ_{lim} is robust and reliable.

When implementing PRAVT on any trajectory data, first calculate μ_{lim} and σ_{lim}. If avaliable, the calibrated values of time gap (τ) from the litearture can be used to calculate μ_{lim} and σ_{lim}, otherwise the aforementioned procedure shall be followed to calculate the two limits. In case of NGSIM data, $\mu_{lim} = 5\ s$ and $\sigma_{lim} = 1.5\ s$ are calculated using the time gap (τ) values reported in Ahn et al. (2004), Zheng et al. (2013) and (2011b).

Table 3 PRAVT performance measures estimated using calculated μ_{lim} and σ_{lim}.

Leader-follower pair ID	CF present (%)	FF present (%)	True CF estimated (%)	True FF estimated (%)
1	43.38	56.62	83.33	95.71
2	2.15	97.85	3.98	100
3	71.71	28.29	89.43	46.40
4	18.14	81.86	23.85	88.26
5	94.49	5.51	89.58	0.00
6	54.87	45.13	92.46	92.50
7	95.70	4.30	94.46	0
8	80.93	19.07	89.87	37.18
9	99.78	0.22	99.77	0
10	99.79	0.21	99.56	0

4.1.4 Testing PRAVT in an urban street setting

PRAVT is also tested using another numerical experiment in which Newell's CF model (Newell, 2002) is used to generate the followers' trajectories in an urban street setting (refer Equations (3)). Newell's CF model is preferred because of its simplicity, behavioural consistency with the S-R approach, and a clear transition from FF to CF. To avoid redundancy and for brevity, only the experiment design is explained below.

This experiment mimics a typical single lane setting in the urban environment. More specifically, the leader and the follower are travelling on a one-lane road. At $t = 0\ s$, the leader stops $5000\ m$ ahead of the reference point (the leader's position $x_{n-1}(0) = 5000\ m$, speed $v_{n-1}(0) = 0\ m/s$, and acceleration $a_{n-1}(0) = 0\ m/s^2$). The starting point of the follower is $1000\ m$ behind the leader (the follower's position $x_n(0) = 4000\ m$, speed $v_n(0) = 0\ m/s$, and acceleration $a_n(0) = 0\ m/s^2$). At $t = 0.1\ s$ the leader and the follower start to accelerate freely with the maximum acceleration a_{max} of $1\ m/s^2$ and $1.5\ m/s^2$, respectively, and the desired speed V_0 of $26\ m/s$ and $30\ m/s$, respectively. The leader attains V_0 at $t = 55.2\ s$ and maintain it until $t = 70\ s$. The follower attains V_0 at $t = 45.6\ s$. At $250\ m$ ahead of this position a traffic signal is placed. From $t = 70.1\ s$, the leader starts to decelerate, comes to a standstill at $t = 91.4\ s$, and stops at the signal for a duration of $19\ s$. A long standstill period is deliberately kept here to ensure that the follower catches up, and CF occurs. The follower reaches the signal at $t = 104.5\ s$ and stops behind the leader.

With the onset of the green light at $t = 110.1\ s$, the leader and the follower accelerate, and CF commences. The leader attains a speed of $22\ m/s$ at $t = 153.9\ s$ and maintain it until $t = 165\ s$. Likewise, the follower attains the same speed at $t = 156\ s$. After $t = 165\ s$, the leader starts to decelerate due to the second traffic signal placed at $250\ m$ ahead. In response, the follower starts to decelerate and comes to a halt at $t = 187.5\ s$. The experiment ends at $t = 188\ s$ with the leader and the follower stop at the traffic signal.

PRAVT is applied to the leader-follower pairs generated from the above experiment and the performance measures are calculated. The true CF estimated and the true FF estimated are above 80% for 25 leader-follower pairs out of 30. Hence, the results are consistent with the previous numerical experiment results (using Gipps' CF model) and underscore the good performance of PRAVT.

4.2 Applying PRAVT to NGSIM data

Before implementing PRAVT, NGSIM data preparation is carried out for PRAVT's smooth implementation. This phase mainly reorders the NGSIM trajectories based on their spatial positions, removes motorcycles' trajectories, lane changers' trajectories, and those single trajectories that are sandwiched between two lane changers.

We implemented PRAVT to trajectories in the reconstructed I-80 data and the denoised US-101 data. For the demonstration purpose, Figure 10 presents the PRAVT application to I-80 lane 2 data. Particularly, Figure 10(b) displays different driving regimes identified by PRAVT for a leader-follower couple in lane 2. As observed from Figure 10(b), drivers can undergo several regime transitions when he/she is travelling at low speeds in stop/slow-and-go traffic (7 regime transitions from 618 s to 630 s).

The Followers' trajectories with different levels of trajectory completeness are divided into eight groups, namely FaCADFS, FaCADF, CADFS, CADF, FaADFS, FaADF, ADFS, and ADF. A trajectory belongs to a particular group if it constitutes all those regimes which defines that group (e.g. if the group is CADFS then any trajectory in this group constitutes all free acceleration, cruising at the desired speed, accelerating behind the leader, deceleration behind the leader, following the leader at constant speed, and standing behind the leader).

Note that these eight groups represent eight different levels of trajectory completeness. FaCADFS represents the highest level (six regimes) and ADF represents the lowest level (three regimes). The PRAVT implementation results are summarised in Table 4.

Surprisingly, none of the trajectories in the NGSIM data are complete. Particularly, no trajectory contains C regime, which is expected because the NGSIM data was collected for heavy or congested traffic. In addition, only 24 % of the trajectories contain the free acceleration regime (824 out of 3414), and the free acceleration regime mostly appears right after the standstill regime, as illustrated in Figure 10(b). Such pattern is reasonable because for various reasons (e.g., deliberately waiting for a comfortable gap or distracted by secondary tasks) a driver in the standstill regime can end up with a large spacing between his/her vehicle and the leader, which enables the driver to freely accelerate.

Table 4 PRAVT application results – NGSIM data (I-80 and US-101).

Freeway	Lanes	Total trajectories*	Number of trajectories in each group							
			FaCADFS	CADFS	FaADFS	ADFS	FaCADF	CADF	FaADF	ADF
I-80 (reconstructed data) (4:00 p.m. to 4:15 p.m.)	Lane 2	221	0	0	5	26	0	0	11	179
	Lane 3	112	0	0	4	5	0	0	12	91
	Lane 4	100	0	0	8	7	0	0	13	72
	Lane 5	66	0	0	6	4	0	0	6	50
	Lane 6	72	0	0	3	5	0	0	7	57
US-101-1 (7:50 a.m. to 8:05 a.m.)	Lane 1	253	0	0	32	50	0	0	32	139
	Lane 2	219	0	0	19	9	0	0	10	181
	Lane 3	203	0	0	15	10	0	0	4	174
	Lane 4	161	0	0	4	7	0	0	4	146
	Lane 5	145	0	0	11	7	0	0	5	122
US-101-2 (8:05 a.m. to 8:20 a.m.)	Lane 1	263	0	0	60	29	0	0	29	145
	Lane 2	196	0	0	46	13	0	0	11	126
	Lane 3	202	0	0	50	9	0	0	24	119
	Lane 4	190	0	0	32	22	0	0	14	122
	Lane 5	127	0	0	10	6	0	0	11	100
US-101-3 (8:20 a.m. to 8:35 a.m.)	Lane 1	241	0	0	36	25	0	0	48	132
	Lane 2	194	0	0	55	33	0	0	25	81
	Lane 3	168	0	0	39	16	0	0	28	85
	Lane 4	162	0	0	37	10	0	0	27	88
	Lane 5	119	0	0	13	5	0	0	18	83
Total		**3414**	**0**	**0**	**485 (14 %)**	**298 (9 %)**	**0**	**0**	**339 (10 %)**	**2292 (67 %)**

*The term trajectories here refer to the followers' trajectories only (i.e., the leaders' are excluded)

Furthermore, 77% of the NGSIM data lacks standstill regime. Another fact worth mentioning is that the I-80 and US-101 data are dominated (i.e., 67%) by the trajectories that belong to the group ADF (acceleration, deceleration, and following). The highest and lowest levels of trajectory completeness available in the NGSIM data are FaADFS (14%) and ADF (67%), respectively.

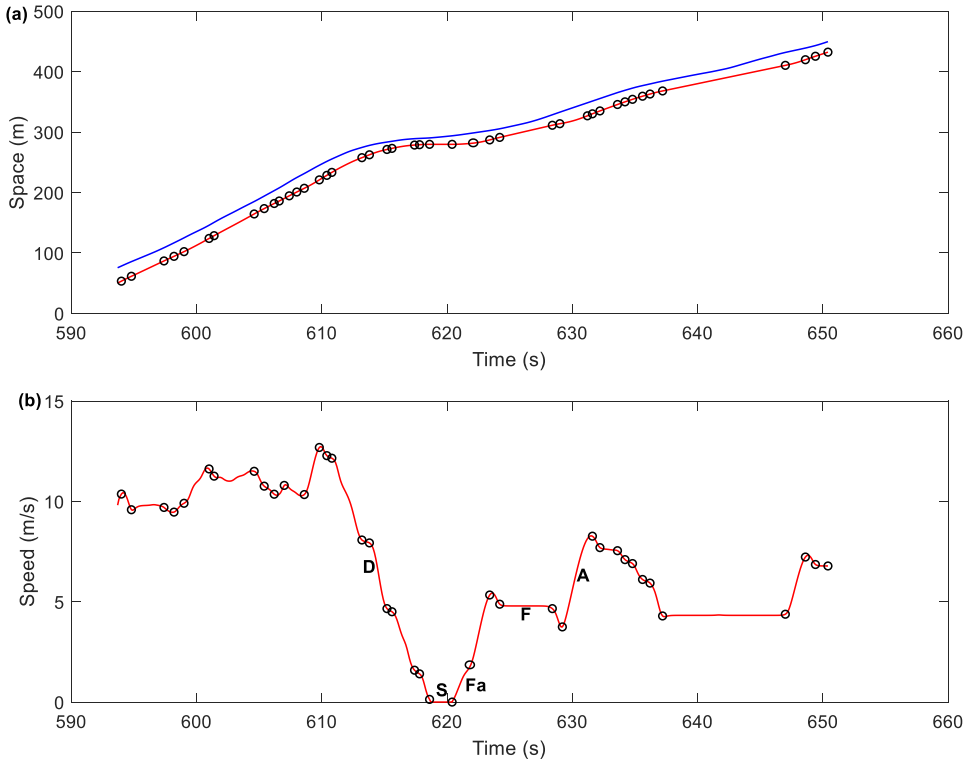


Figure 10 PRAVT application results. a) a trajectory pair form reconstructed I-80 data; b) the speed profile of the follower with regimes identified. Note that in (b) circles represent the end points of the regimes, Fa is observed once, and for clarity, the names of the other identified regimes are mentioned only once.

The aforementioned facts revealed by implementing PRAVT underscore the importance and necessity of carefully checking the trajectories' quality before they are used for calibrating CF models. Considering the importance of the information related to driving regimes and trajectory completeness in CF model calibration and validation, and the popularity of the NGSIM data in the traffic flow community, for each vehicle ID in NGSIM data we provide its trajectory completeness group in Appendix C.

5 Conclusion and future work

With the advancement in data collection technologies, access to high-resolution trajectory data provides an unprecedented opportunity for researchers to empirically assess performance of CF models and test their behavioural soundness. In particular, to reliably calibrate and validate CF models, trajectories should reflect all the main interactions between the follower and the leader in the real world. Although researchers have reported the importance of trajectory completeness, a robust algorithm to test the completeness of trajectories is missing. This paper presents a three-stage pattern recognition algorithm for vehicle trajectories (PRAVT) to test

the completeness of vehicular trajectories and select the most complete trajectories. The speed profiles are segmented in Stage I of PRAVT, and after identifying the Free-Flow and Car-following sections in Stage II, the slopes of speed segments i.e., acceleration/deceleration values are used to identify different driving regimes in Stage III.

PRAVT is applied to the reconstructed I-80 and US-101 data and several surprising results are revealed. Out of 3414 followers' trajectories, no trajectory is complete. More specifically, no trajectory constitutes cruising regime, 77% lack standstill regime, and the NGSIM data are dominated by the trajectories belonging to ADF group (i.e., trajectories with acceleration, deceleration, and following regimes). Given the celebrity status of the NGSIM data in the traffic flow community, these findings' impact on how to properly use NGSIM data can be profound.

Meanwhile, for the first time, the threshold between the CF and the FF section is determined based on the fundamental stimulus-response (S-R) relationship which is at the core of the CF phenomenon. In addition, a method based on DTW and an extended Bottom-Up algorithm are proposed to accurately determine the threshold for distinguishing between CF and FF for a pair of vehicles (PRAVT Stage I and II). This method can be easily extended to compute a global threshold for a lane (or a set of leader-follower couples).

Using synthetic and real data, this paper demonstrates DTW's capability to successfully detect the S-R relationship in CF (a limitation of previous studies). This finding is important as it implies that DTW mapping can be utilized to develop better S-R based CF models.

Approaching/Free deceleration (Fd) regime (i.e., decelerating while approaching a standing leader from a large distance) is another driving regime. Treiber and Kesting (2013b, 2013a) present a nice discussion on this regime. As PRAVT is developed on the S-R relationship between the leader and the follower, differentiating Fd regime from D regime using trajectory data is hard, because, in both Fd and D regimes the follower decelerates due to the leader's action (moving slowly or at the stopped position or any other action) i.e., the existence of similar S-R relationship. Thus, to avoid ambiguity Fd regime is excluded from the current analysis. However, PRAVT can be extended and enhanced to include Fd regime in the future. Moreover, other filtering/smoothing techniques (e.g., kernel-based smoothing) can be integrated into PRAVT as a data processing step, i.e., data can be smoothed first before they are entered as input into PRAVT algorithm. Intuitively, this would further enhance PRAVT's performance.

In this study, we mainly used data from freeways to test the performance of PRAVT. However, the numerical experiments presented in the study are designed in a general context. Moreover, in essence, PRAVT is a general pattern recognition method specifically designed for identifying typical driving regimes in vehicular trajectories, let it be freeways or urban streets. To apply this algorithm to a different environment, like applying any other traffic models it needs to be re-calibrated. However, the procedure will remain the same, and its performance is unlikely to significantly fluctuate as demonstrated consistently through the rigorous tests reported in this study, using both the freeway data and the (synthetic) urban traffic. Nevertheless, exploring the applicability of PRAVT in a wide range of environments is an interesting topic for future research.

Furthermore, to detect the segments with a constant speed, we used the limit 0.05g (g is the gravitational acceleration) which is based on Ozaki's definition (Ozaki, 1993). To the best of

our knowledge, no other studies have presented a methodology/empirical limits to differentiate the following regime with a constant speed from gradual accelerations/decelerations. A comprehensive research on this topic is needed. Finally, for NGSIM data, PRAVT has identified all the regimes of lengths greater than equal to 0.5 s. The minimum length of a regime is not critical because PRAVT is flexible to incorporate different minimum lengths of regimes as pre-defined by its users as per their specific situations. In furture research, it is interesting to investigate what should be the minimum length of a regime in the trajectory and how this factor influences the CF model calibration and validation.

In summary, using the methods developed in this paper, CF and FF sections in a trajectory can be accurately identified. Moreover, different driving regimes in the CF or FF section can also be extracted. Such information enables researchers to answer some fundamental issues related to CF calibration and validation, such as: i) is there any relationship between driving regimes and parameters of a CF model?; and ii) if some driving regimes are missing, how will it impact parameters estimates at the model calibration stage, and how will it impact the model's performance at the model validation stage? These topics are out of the scope of this paper, but the focus of Sharma et al. (2018).

Acknowledgements: This research was partially funded by the Australian Research Council (ARC) through Dr. Zuduo Zheng's Discovery Early Career Researcher Award (DECRA; DE160100449).

References

Ahmed, K.I., 1999. Modeling drivers' acceleration and lane changing behavior.

Ahn, S., Cassidy, M.J., Laval, J., 2004. Verification of a simplified car-following theory. Transportation Research Part B: Methodological 38, pp. 431–440.

Albrecht, T., Muller, M., 2009. Dynamic Time Warping (DTW). Information Retreival for Music and Motion, pp. 70–83.

Aycin, M., Benekohal, R., 1998. Linear acceleration car-following model development and validation. Transportation Research Record: Journal of the Transportation Research Board 1644, pp. 10–19.

Berndt, D.J., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. Presented at the KDD workshop, Seattle, WA, pp. 359–370.

Park, B., Qi, H., 2005. Development and Evaluation of a Procedure for the Calibration of Simulation Models. Transportation Research Record: Journal of the Transportation Research Board 1934, 208–217.

Brockfeld, E., Kühne, R., Wagner, P., 2004. Calibration and Validation of Microscopic Traffic Flow Models. Transportation Research Record: Journal of the Transportation Research Board 1876, pp. 62–70.

Caiani, E., Porta, A., Baselli, G., Turiel, M., Muzzupappa, S., Pieruzzi, F., Crema, C., Malliani, A., Cerutti, S., 1998. Warped-average template technique to track on a cycle-by-cycle basis the cardiac filling phases on left ventricular volume. Presented at the Computers in Cardiology 1998, IEEE, pp. 73–76.

Chen, D., Ahn, S., Laval, J., Zheng, Z., 2014. On the periodicity of traffic oscillations and capacity drop: The role of driver characteristics. Transportation Research Part B: Methodological 59, pp. 117–136.

Chen, D., Laval, J.A., Ahn, S., Zheng, Z., 2012. Microscopic traffic hysteresis in traffic oscillations: A behavioral perspective. Transportation Research Part B: Methodological 46, pp. 1440–1453.

Cheng, Y., Qin, X., Jin, J., Ran, B., Anderson, J., 2011. Cycle-by-Cycle Queue Length Estimation for Signalized Intersections Using Sampled Trajectory Data. Transportation Research Record: Journal of the Transportation Research Board 2257, pp. 87–94.

Ciuffo, B., Punzo, V., Montanino, M., 2012. The Calibration of Traffic Simulation Models: Report on the Assessment of Different Goodness of Fit Measures and Optimization Algorithms MULTITUDE Project–COST Action TU0903. EUR.

Duret, A., Buisson, C., Chiabaut, N., 2008. Estimating Individual Speed-Spacing Relationship and Assessing Ability of Newell's Car-Following Model to Reproduce Trajectories. Transportation Research Record: Journal of the Transportation Research Board 2088, pp. 188–197.

Esling, P., Agon, C., 2012. Time-series Data Mining. ACM Comput. Surv. 45, 12:1–12:34.

Gipps, P.G., 1986. A model for the structure of lane-changing decisions. Transportation Research Part B: Methodological 20, pp. 403–414.

Gipps, P.G., 1981. A behavioural car-following model for computer simulation. Transportation Research Part B: Methodological 15, pp. 105–111.

Gollmer, K., Posten, C., 1995. Detection of distorted pattern using dynamic time warping algorithm and application for supervision of bioprocesses. IFAC Proceedings Volumes 28(12), pp. 101-106.

Hamdar, S.H., Mahmassani, H.S., 2008. Driver car-following behavior: from discrete event process to continuous set of episodes. Presented at the Transportation Research Board 87th Annual Meeting Transportation Research Board, Washington, DC.

Herman, R., Potts, R.B., 1961. Single lane traffic theory and experiment.

Herrera, J.C., Bayen, A.M., 2008. Traffic Flow Reconstruction Using Mobile Sensors and Loop Detector Data. Presented at the Transportation Research Board 87th Annual Meeting Transportation Research Board, Washington, DC.

Hoogendoorn, S., Hoogendoorn, R., 2010. Calibration of microscopic traffic-flow models using multiple data sources. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 368, pp. 4497–4517.

Hoogendoorn, S., Hoogendoorn, R., Daamen, W., 2011. Wiedemann revisited: new trajectory filtering technique and its implications for car-following modeling. Transportation Research Record: Journal of the Transportation Research Board 2260, pp. 152–162.

Keogh, E., Chu, S., Hart, D., Pazzani, M., 1993. Segmenting time series: a survey and novel approach, in: In an Edited Volume, Data Mining in Time Series Databases. Data mining in time series databases 57, pp. 1–22.

Keogh, E., Pazzani, M., 2001. Derivative dynamic time warping, in: Proceedings of the 2001 SIAM International Conference on Data Mining, Proceedings. Society for Industrial and Applied Mathematics, pp. 1–11.

Keogh, E., Ratanamahatana, C.A., 2005. Exact indexing of dynamic time warping. Knowledge and Information Systems 7, pp. 358–386.

Keogh, E.J., Pazzani, M.J., 1998. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. Presented at the KDD 98, pp. 239–243.

Kesting, A., Treiber, M., 2008. Calibrating car-following models by using trajectory data: methodological study. Transportation Research Record: Journal of the Transportation Research Board 2088, pp. 148–156.

Laval, J.A., Daganzo, C.F., 2006. Lane-changing in traffic streams. Transportation Research Part B: Methodological 40, pp. 251–264.

Laval, J.A., Leclercq, L., 2010. A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 368, pp. 4519–4541.

Liu, H.X., Wu, X., Ma, W., Hu, H., 2009. Real-time queue length estimation for congested signalized intersections. Transportation Research Part C: Emerging Technologies 17, pp. 412–427.

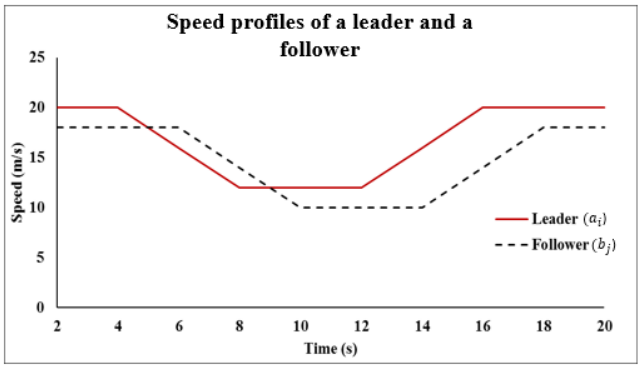
McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 21, pp. 239–245.

Montanino, M., Ciuffo, B., Punzo, V., 2012. Calibration of microscopic traffic flow models against time-series data. Presented at the 2012 15th International IEEE Conference on Intelligent Transportation Systems, pp. 108–114.

Montanino, M., Punzo, V., 2015. Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns. Transportation Research Part B: Methodological 80, pp. 82–106.
Monteil, J., Billot, R., Sau, J., Buisson, C., El Faouzi, N.-E., 2014. Calibration, estimation, and sampling issues of car-following parameters. Transportation Research Record: Journal of the Transportation Research Board 2422, 131–140.
Newell, G.F., 2002. A simplified car-following theory: a lower order model. Transportation Research Part B: Methodological 36, pp. 195–205.
NGSIM, 2010. Traffic Analysis Tools: Next Generation Simulation - FHWA Operations [WWW Document]. URL https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm (accessed 3.4.17).
Nishiwaki, Y., Miyajima, C., Kitaoka, N., Takeda, K., 2012. A Stochastic Approach for Modeling Lane-Change Trajectories, in: Hansen, J.H.L., Boyraz, P., Takeda, K., Abut, H. (Eds.), Digital Signal Processing for In-Vehicle Systems and Safety. Springer New York, pp. 271–282.
Ossen, S., Hoogendoorn, S., 2008. Validity of Trajectory-Based Calibration Approach of Car-Following Models in Presence of Measurement Errors. Transportation Research Record: Journal of the Transportation Research Board 2088, pp. 117–125.
Ossen, S., Hoogendoorn, S., 2005. Car-Following Behavior Analysis from Microscopic Trajectory Data. Transportation Research Record: Journal of the Transportation Research Board 1934, pp. 13–21.
Ozaki, H., 1993. Reaction and anticipation in the car-following behavior. Transportation and traffic theory.
Panwai, S., Dia, H., 2007. Neural Agent Car-Following Models. IEEE Transactions on Intelligent Transportation Systems 8, pp. 60–70.
Przybyla, J., Taylor, J., Jupe, J., Zhou, X., 2012. Simplified, data-driven, errorable car-following model to predict the safety effects of distracted driving. Presented at the 2012 15th International IEEE Conference on Intelligent Transportation Systems, pp. 1149–1154.
Punzo, V., Borzacchiello, M.T., Ciuffo, B., 2011. On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. Transportation Research Part C: Emerging Technologies 19, pp. 1243–1262.
Punzo, V., Ciuffo, B., 2009. How Parameters of Microscopic Traffic Flow Models Relate to Traffic Dynamics in Simulation. Transportation Research Record: Journal of the Transportation Research Board 2124, pp. 249–256.
Punzo, V., Ciuffo, B., Montanino, M., 2012. Can Results of Car-Following Model Calibration Based on Trajectory Data Be Trusted? Transportation Research Record: Journal of the Transportation Research Board 2315, pp. 11–24.
Punzo, V., Montanino, M., Ciuffo, B., 2015. Do we really need to calibrate all the parameters? Variance-based sensitivity analysis to simplify microscopic traffic flow models. IEEE Transactions on Intelligent Transportation Systems 16, pp. 184–193.
Punzo, V., Simonelli, F., 2005. Analysis and comparison of microscopic traffic flow models with real traffic microscopic data. Transportation Research Record: Journal of the Transportation Research Board 1934, pp. 53–63.
Rabiner, L., Juang, B.-H., 1993. Fundamentals of speech recognition. Prentice Hall.
Saifuzzaman, M., Zheng, Z., Haque, M.M., Washington, S., 2017. Understanding the mechanism of traffic hysteresis and traffic oscillations through the change in task difficulty level. Transportation Research Part B: Methodological 105, pp. 523–538.
Saifuzzaman, M., Zheng, Z., Mazharul Haque, M., Washington, S., 2015. Revisiting the Task–Capability Interface model for incorporating human factors into car-following models. Transportation Research Part B: Methodological 82, pp. 1–19.
Schmill, M.D., Oates, T., Cohen, P.R., 1999. Learned models for continuous planning. Presented at the AISTATS.
Senin, P., 2008. Dynamic time warping algorithm review. Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA, 855, pp. 1–23.
Sharma, A., Zheng, Z., Bhaskar, A., 2018. Is more always better? The impact of vehicular trajectory completeness on car-following model calibration and validation (Under review).

Srivastava, A., Geroliminis, N., 2013. Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model. Transportation Research Part C: Emerging Technologies 30,pp. 161–177.
Subramanian, H., 1996. Estimation of car-following models (Maters thesis). Massachusetts Institute of Technology.
Taylor, J., Zhou, X., Rouphail, N.M., Porter, R.J., 2015. Method for investigating intradriver heterogeneity using vehicle trajectory data: A Dynamic Time Warping approach. Transportation Research Part B: Methodological 73, pp. 59–80.
Thiemann, C., Treiber, M., Kesting, A., 2008. Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data. Transportation Research Record: Journal of the Transportation Research Board 2088, pp. 90–101.
Toledo, T., 2003. Integrated driving behavior modeling (Doctoral dissertation). Massachusetts Institute of Technology.
Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. Phys. Rev. E 62, pp. 1805–1824.
Treiber, M., Kesting, A., 2013a. Microscopic calibration and validation of car-following models – a systematic approach. Procedia - Social and Behavioral Sciences, 20th International Symposium on Transportation and Traffic Theory (ISTTT 2013) 80, pp. 922–939.
Treiber, M., Kesting, A., 2013b. Traffic flow dynamics. Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg.
Treiber, M., Kesting, A., Thiemann, C., 2008. How much does traffic congestion increase fuel consumption and emissions? Applying a fuel consumption model to the NGSIM trajectory data. Presented at the 87th Annual Meeting of the Transportation Research Board, Washington, DC.
Vieira da Rocha, T., Leclercq, L., Montanino, M., Parzani, C., Punzo, V., Ciuffo, B., Villegas, D., 2015. Does traffic-related calibration of car-following models provide accurate estimations of vehicle emissions? Transportation Research Part D: Transport and Environment 34, pp. 267–280.
Wagner, P., 2012. Analyzing fluctuations in car-following. Transportation Research Part B: Methodological 46, pp. 1384–1392.
Zheng, Z., 2014. Recent developments and research needs in modeling lane changing. Transportation Research Part B: Methodological 60, pp. 16–32.
Zheng, Z., Ahn, S., Chen, D., Laval, J., 2013. The effects of lane-changing on the immediate follower: Anticipation, relaxation, and change in driver characteristics. Transportation Research Part C: Emerging Technologies 26, pp. 367–379.
Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011a. Applications of wavelet transform for analysis of freeway traffic: Bottlenecks, transient traffic, and traffic oscillations. Transportation Research Part B: Methodological 45, pp. 372–384.
Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011b. Freeway traffic oscillations: Microscopic analysis of formations and propagations using Wavelet Transform. Transportation Research Part B: Methodological, Select Papers from the 19th ISTTT 45, pp. 1378–1388.

Appendix A. An example to demonstrate DTW algorithm application



STEP1: $C(i,j)$

a_i \ b_j	18	18	18	14	10	10	10	14	18	18
20	2	2	2	6	10	10	10	6	2	2
20	2	2	2	6	10	10	10	6	2	2
16	2	2	2	2	6	6	6	2	2	2
12	6	6	6	2	2	2	2	2	6	6
12	6	6	6	2	2	2	2	2	6	6
12	6	6	6	2	2	2	2	2	6	6
16	2	2	2	2	6	6	6	2	2	2
20	2	2	2	6	10	10	10	6	2	2
20	2	2	2	6	10	10	10	6	2	2
20	2	2	2	6	10	10	10	6	2	2

$C(4,3) = |a_4 - b_3| = \sqrt{(12-18)^2} = 6$

STEP2: $C^{cum}(i,\ j)$

	18	18	18	14	10	10	10	14	18	18
20	2	4	6	12	22	32	42	48	50	52
20	4	4	6	12	22	32	42	48	50	52
16	6	6	6	8	14	20	26	28	30	32
12	12	12	12	8	10	12	14	16	22	28
12	18	18	18	10	10	12	14	16	22	28
12	24	24	24	12	12	12	14	16	22	28
16	26	26	26	14	18	18	18	16	18	20
20	28	28	28	20	24	28	28	22	18	20
20	30	30	30	26	30	34	38	28	20	20
20	32	32	32	32	36	40	44	34	22	22

$C^{cum}(4,\ 3)$
$= C(4,3) + min(C^{cum}(4-1,\ 3-1), C^{cum}(4,\ 3-1), C^{cum}(4-1,\ 3))$
$= 6 + \min(6,6,12) = 12$

STEP3: w_p

	18	18	18	14	10	10	10	14	18	18
20	**2**	4	6	12	22	32	42	48	50	52
20	4	**4**	6	12	22	32	42	48	50	52
16	6	6	**6**	8	14	20	26	28	30	32
12	12	12	12	**8**	10	12	14	16	22	28
12	18	18	18	10	**10**	12	14	16	22	28
12	24	24	24	12	12	**12**	**14**	16	22	28
16	26	26	26	14	18	18	18	**16**	18	20
20	28	28	28	20	24	28	28	22	**18**	20
20	30	30	30	26	30	34	38	28	20	**20**
20	32	32	32	32	36	40	44	34	22	**22**

Shaded cells above show the warping path

Similar points on the two speed profiles via warping path

a_i	20	20	16	12	12	12	12	16	20	20	20
b_j	18	18	18	14	10	10	10	14	18	18	18

Figure A Self-explanatory example of DTW application.

Appendix B. Singularities in DTW and how to reconcile them

Singularities are an undesirable phenomenon in DTW where a single point in one time series maps onto more than one point of the other time series (one-to-many mapping). In our case where the objective is to detect the S-R relationship, this type of DTW behaviour is problematic. The primary reason for causing singularities is the incapability of DTW to distinguish between identical points present at a rising trend and a falling trend (note that DTW only considers Y-axis values) (Keogh and Pazzani, 2001). For solving this problem there are several attempts documented in the literature. For brevity two main approaches are introduced below.

The first approach is referred to as the derivative DTW algorithm, which was proposed by Keogh and Pazzani (2001). The only difference between DTW algorithm mentioned before and the derivative DTW algorithm is how the distance measure is calculated. The distance measure in the original DTW is Euclidean distance. But in the derivative DTW algorithm it is the square of the difference of the estimated derivatives of a_i and b_j. For simplicity and generality, the following derivative estimate (Equation A1) can be used:

$$D_x[a] = \frac{(a_i - a_{i-1}) + (\frac{a_{i+1} - a_{i-1}}{2})}{2} \quad \text{(A1)}$$

For a detailed description and application of the derivative DTW, see Keogh and Pazzani (2001). . Figure B displays the presence of singularities in the trajectory data (Figure B (a)) and singularity reduction when derivative DTW is applied to the trajectories (Figure B (b)).

The second effective way to reconcile the singularity is to reduce the high dimensionality (high dimension refers to a large number of data points in the time series) of the time series (used in this study). This can be achieved by segmentation (Esling and Agon, 2012). Segmenting the trajectory before applying DTW can prove useful because segmentation creates a low-dimensional representation of the actual trajectory

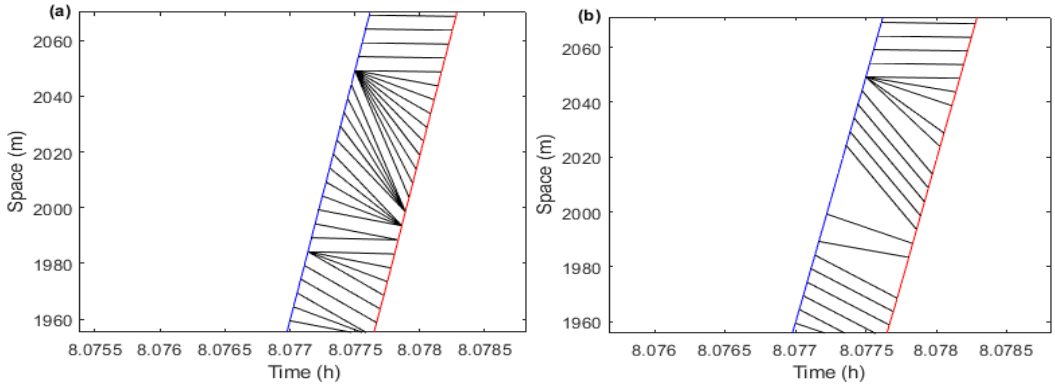


Figure B a) DTW mapping on original time series resulting in high singularities; b) Derivative DTW mapping on original time series resulting in low singularities.

Appendix C. Results: PRAVT applied to NGSIM data

The tables below present followers' vehicle IDs (identification numbers) categorised according to trajectory groups.

Table C.1 Vehicle IDs and corresponding trajectory groups in I-80 reconstructed data.

Lane ID	Trajectory groups			
	FaADFS	ADFS	FaADF	ADF
Lane 2	925, 1941, 1998, 2019, 2065	1780, 1804, 1829, 1845, 1887, 1961, 1992, 2006, 2022, 2035, 2042, 2058, 2071, 2095, 2092, 2113, 2122, 2128, 2132, 2151, 2158, 2169, 2176, 2186, 2190, 2199	190, 730, 804, 1565, 1717, 1765, 1795, 1868, 1951, 1984, 3152	11, 24, 87, 97, 117, 3340, 200, 224, 249, 255, 260, 282, 293, 348, 343, 354, 362, 368, 378, 381, 391, 417, 436, 459, 469, 472, 488, 494, 507, 510, 515, 533, 542, 565, 621, 636, 649, 660, 683, 695, 723, 793, 812, 839, 852, 868, 898, 905, 909, 937, 998, 1013, 1023, 1030, 1037, 1045, 1074, 1080, 1120, 1136, 1146, 1149, 1157, 1165, 1181, 1195, 1217, 1221, 1254, 1268, 1274, 1287, 1291, 1305, 1307, 1317, 1324, 1332, 1341, 1346, 1350, 1355, 1358, 1367, 1381, 1386, 1410, 1414, 1400, 1418, 1437, 1443, 1455, 1459, 1462, 1466, 1477, 1481, 1485, 1500, 1516, 1521, 1577, 1638, 1627, 1631, 1645, 1659, 1675, 1681, 1689, 1698, 1739, 1746, 1810, 1851, 1861, 1873, 1882, 1898, 1927, 2146, 2234, 2275, 2301, 2308, 2337, 2341, 2347, 2352, 2395, 2405, 2411, 2426, 2441, 2459, 2473, 2474, 2481, 2809, 2800, 2849, 2854, 2860, 2872, 2906, 2912, 2932, 2930, 2968, 2980, 2992, 2999, 3004, 3009, 3015, 3024, 3034, 3046, 3060, 3072, 3078, 3090, 3093, 3128, 3138, 3171, 3179, 3186, 3188, 3199, 3206, 3215, 3242, 3249, 3260, 3291, 3299, 3305
Lane 3	1497, 1702, 1803, 1828	1463, 1478, 1680, 1687, 1764	1406, 1441, 1779, 1794, 1836, 2165, 2404, 2499, 2897, 3084, 3172, 3238	232, 358, 371, 401, 412, 416, 424, 483, 492, 503, 546, 554, 675, 720, 748, 749, 773, 782, 787, 831, 876, 908, 913, 964, 986, 1003, 1026, 1077, 1135, 1147, 1167, 1175, 1180, 1231, 1238, 1304, 1309, 1314, 1333, 1338, 1343, 1353, 1365, 1382, 1390, 1682, 1842, 1847, 1852, 1888, 1897, 1905, 1913, 1925, 1932, 2004, 2061, 2072, 2076, 2121, 2139, 2133, 2175, 2266, 2342, 2384, 2394, 2419, 2422, 2770, 2777, 2842, 2866, 2880, 2913, 2931, 2938, 2942, 2986, 3018, 3094, 3143, 3156, 3193, 3219, 3253, 3252, 3280, 3286, 3290, 3301
Lane 4	1721, 1767, 2045, 2803, 2813, 2834, 2868, 3000	1578, 1597, 1816, 1995, 2855, 2893, 3166	107, 710, 1593, 1807, 1874, 1889, 2062, 2088, 2105, 2119, 2828, 2847, 3085	13, 122, 193, 251, 264, 269, 350, 396, 402, 407, 415, 418, 427, 440, 451, 479, 486, 559, 567, 587, 853, 902, 967, 973, 997, 1004, 1025, 1105, 1122, 1212, 1224, 1222, 1229, 1288, 1354, 1359, 1394, 1412, 1435, 1467, 1479, 1482, 1490, 1507, 1700, 1788, 1965, 2013, 2082, 2134, 2166, 2174, 2240, 2261, 2289, 2317, 2334, 2345, 2417, 3055, 3064, 3155, 3169, 3191, 3196, 3208, 3223, 3226, 3245, 3254, 3279, 3285
Lane 5	2117, 2120, 2141, 2143, 2759, 2812	1562, 2138, 2848 2856	1553, 1557, 1572, 1670, 2274, 2933	61, 94, 102, 123, 126, 203, 237, 247, 259, 287, 307, 360, 475, 482, 615, 700, 1007, 1016, 1043, 1118, 1125, 1131, 1192, 1198, 1213, 1223, 1259, 1372, 1399, 1407, 1711, 1722, 1727, 1735, 1753, 1784, 1877, 2162, 2370, 2940, 2958, 2982, 2991, 3027, 3044, 3054, 3059, 3148, 3175, 3181
Lane 6	2957, 2998, 3061	7, 72, 1879, 2188, 3100	687, 1411, 1694, 2171, 2421, 3121, 3154	68, 103, 3333, 168, 179, 197, 222, 234, 375, 458, 480, 484, 491, 818, 829, 982, 991, 1048, 1055, 1155, 1163, 1193, 1215, 1219, 1226, 1302, 1361, 1461, 1488, 1510, 1519, 1524, 1531, 1541, 1554, 1563, 1657, 1669, 1666, 1716, 1742, 1778, 1809, 1900, 2079, 2090, 2118, 2125, 2126, 2149, 2156, 2372, 2492, 2782, 3168, 3167, 3229

Table C.2 Vehicle IDs and corresponding trajectory groups in US-101-1 data.

Lane ID	Trajectory groups *			
	FaADFS	**ADFS**	**FaADF**	**ADF**
Lane 1	427, 554, 571, 605, 610, 624, 1058, 1104, 1431, 1657, 1768, 1782, 1787, 1835, 1840, 1850, 1867, 1870, 2035, 2041, 2079, 2144, 2470, 2603, 2829, 2841, 2848	280, 298, 322, 363, 417, 422, 442, 458, 468, 475, 485, 494, 507, 1025, 1031, 1033, 1053, 1124, 1129, 1443, 1668, 1764, 1795, 2033, 2059, 2073, 2134, 2170, 2175, 2226, 2372, 2399, 2594, 2598, 2633, 2819, 2823, 2836, 2858, 2865, 2877, 2910, 2918	558, 579, 588, 598, 987, 1012, 1091, 1135, 1143, 1245, 1337, 1348, 1454, 1570, 1604, 1772, 2105, 2130, 2162, 2181, 2184, 2202, 2209, 2219, 2478, 2487, 2494, 2515, 2524, 2533, 2658	12, 23, 27, 39, 43, 72, 95, 121, 137, 147, 152, 163, 176, 181, 194, 220, 223, 232, 239, 240, 241, 271, 282, 302, 566, 643, 653, 661, 672, 698, 730, 742, 762, 769, 776, 779, 788, 796, 801, 813, 822, 827, 834, 843, 852, 861, 865, 874, 876, 904, 913, 931, 948, 997, 1005, 1077, 1082, 1086, 1153, 1159, 1166, 1172, 1176, 1189, 1192, 1196, 1205, 1208, 1212, 1217, 1248, 1259, 1280, 1284, 1290, 1309, 1325, 1314, 1358, 1366, 1370, 1409, 1420, 1462, 1469, 1477, 1484, 1514, 1521, 1524, 1529, 1533, 1539, 1549, 1554, 1587, 1592, 1599, 1600, 1648, 1877, 1881, 1889, 1906, 1913, 1926, 1931, 1937, 1945, 1955, 1964, 1970, 1978, 1984, 1989, 2092, 2140, 2190, 2200, 2214, 2246, 2250, 2258, 2262, 2273, 2279, 2285, 2291, 2316, 2323, 2326, 2328, 2337, 2354, 2361, 2393, 2503, 2504, 2548
Lane 2	2444, 2462, 2474, 2480, 2500, 2531, 2538, 2549, 2554, 2562, 2581, 2714, 2780, 2850, 2855, 2861, 2948, 2991, 2993	1538, 2719, 2729, 2730, 2792, 2840, 2843, 2930, 2954	1423, 1681, 1848, 2174, 2176, 2290, 2516, 2571, 2578, 2625	22, 26, 32, 78, 94, 105, 127, 151, 159, 164, 175, 180, 184, 196, 212, 281, 292, 297, 304, 311, 344, 351, 364, 372, 379, 383, 399, 431, 440, 446, 484, 514, 564, 570, 578, 585, 618, 636, 669, 671, 728, 734, 738, 748, 767, 771, 793, 812, 825, 833, 839, 844, 851, 854, 873, 871, 882, 883, 891, 897, 905, 915, 976, 979, 1006, 1032, 1065, 1083, 1101, 1127, 1131, 1134, 1138, 1144, 1179, 1187, 1201, 1209, 1233, 1240, 1246, 1262, 1286, 1289, 1301, 1304, 1323, 1339, 1338, 1346, 1353, 1361, 1367, 1373, 1399, 1417, 1435, 1441, 1444, 1453, 1476, 1478, 1486, 1489, 1504, 1507, 1526, 1548, 1551, 1557, 1573, 1577, 1588, 1593, 1598, 1650, 1673, 1677, 1698, 1703, 1709, 1746, 1752, 1755, 1762, 1788, 1839, 1844, 1875, 1890, 1901, 1908, 1911, 1916, 1924, 1930, 1967, 1973, 1980, 1985, 1992, 1998, 2001, 2008, 2019, 2037, 2046, 2053, 2057, 2065, 2074, 2090, 2096, 2106, 2115, 2180, 2198, 2203, 2208, 2212, 2216, 2222, 2228, 2233, 2249, 2260, 2265, 2270, 2280, 2314, 2322, 2324, 2334, 2344, 2350, 2356, 2401, 2405, 2411, 2417, 2418
Lane 3	2463, 2523, 2528, 2568, 2595, 2619, 2649, 2660, 2668, 2672, 2700, 2827, 2832, 2838, 2860	2446, 2508, 2536, 2542, 2563, 2590, 2605, 2694, 2717, 2725	1468, 2205, 2628, 2891	20, 150, 161, 192, 199, 206, 213, 255, 257, 295, 301, 328, 342, 347, 368, 374, 380, 403, 419, 439, 443, 502, 533, 542, 602, 616, 623, 630, 649, 659, 667, 694, 702, 706, 717, 736, 739, 751, 765, 775, 782, 787, 791, 799, 845, 849, 860, 869, 886, 892, 903, 908, 912, 917, 929, 945, 962, 968, 972, 977, 1008, 1011, 1021, 1040, 1046, 1055, 1057, 1084, 1112, 1122, 1125, 1145, 1148, 1170, 1177, 1180, 1213, 1234, 1250, 1282, 1287, 1294, 1350, 1354, 1363, 1369, 1372, 1395, 1412, 1429, 1473, 1481, 1487, 1520, 1523, 1537, 1540, 1545, 1553, 1555, 1559, 1578, 1586, 1590, 1596, 1622, 1625, 1655, 1683, 1690, 1692, 1697, 1705, 1710, 1716, 1718, 1726, 1729, 1740, 1751, 1759, 1763, 1765, 1841, 1846, 1862, 1885, 1900, 1904, 1910, 1915, 1928, 1968, 1971, 1975, 1982, 1988, 1995, 1999, 2003, 2031, 2032, 2060, 2168, 2173, 2186, 2210, 2232, 2237, 2240, 2248, 2251, 2268, 2315, 2329, 2339, 2348, 2355, 2367, 2407, 2420, 2441, 2498, 2691, 2901, 2912, 2922, 2940, 2945, 2962, 2974, 2979, 3008, 3019
Lane 4	2822, 2834, 2907, 2921	2826, 2868, 2913, 2926, 2943, 2972, 2982	2525, 2646, 2853, 2874	8, 21, 25, 62, 119, 145, 158, 228, 277, 285, 300, 408, 481, 495, 501, 546, 562, 567, 611, 658, 665, 690, 697, 705, 715, 725, 733, 755, 778, 850, 858, 872, 879, 887, 896, 906, 911, 924, 978, 981, 1035, 1062, 1067, 1100, 1105, 1108, 1123, 1169, 1175, 1178, 1186, 1188, 1194, 1202, 1218, 1231, 1238, 1291, 1296, 1343, 1351, 1360, 1365, 1391, 1393, 1415, 1430, 1442, 1470, 1479, 1490, 1518, 1522, 1536, 1541, 1626, 1654, 1659, 1670, 1680, 1699, 1701, 1711, 1715, 1721, 1725, 1732, 1734, 1738, 1747, 1753, 1757, 1761, 1770, 1773, 1776, 1780, 1842, 1847, 1852, 1856, 1871, 1909, 1914, 1977, 1981, 2043, 2067, 2135, 2157, 2163, 2167, 2171, 2199, 2206, 2211, 2239, 2243, 2256, 2287, 2374, 2381, 2447, 2464, 2483, 2486, 2502, 2509, 2517, 2529, 2543, 2550, 2556, 2575, 2604, 2620, 2639, 2663, 2671, 2689, 2696, 2699, 2731, 2735, 3014, 3022

Lane 5	2687, 2693, 2727, 2754, 2763, 2783, 2811, 2863, 2900, 2906, 2946	1347, 2698, 2715, 2775, 2835, 2869, 2973	1503, 2557, 2624, 2682, 2915	109, 162, 169, 215, 226, 231, 469, 488, 500, 512, 520, 633, 641, 709, 880, 888, 889, 3103, 909, 927, 942, 975, 992, 1041, 1071, 1076, 1120, 1133, 1157, 1171, 1174, 1181, 1195, 1199, 1203, 1223, 1241, 1243, 1254, 1292, 1300, 1311, 1327, 1359, 1364, 1408, 1440, 1447, 1517, 1527, 1568, 1575, 1579, 1620, 1630, 1638, 1653, 1658, 1669, 1671, 1674, 1679, 1682, 1707, 1719, 1722, 1749, 1756, 1760, 1766, 1774, 1777, 1781, 1838, 1866, 1887, 1902, 1912, 1917, 1922, 1927, 1965, 1972, 1996, 2016, 2034, 2052, 2125, 2147, 2154, 2166, 2172, 2178, 2183, 2187, 2215, 2221, 2230, 2244, 2272, 2275, 2302, 3109, 2325, 2378, 2443, 2450, 2467, 2471, 2532, 2565, 2566, 2569, 2576, 2596, 2602, 2630, 2635, 2644, 3000, 3002, 3015

Table C.3 Vehicle IDs and corresponding trajectory groups in US-101-2 data.

Lane ID	Trajectory groups *			
	FaADFS	**ADFS**	**FaADF**	**ADF**
Lane 1	323, 328, 331, 370, 405, 413, 490, 498, 504, 507, 514, 521, 524, 539, 562, 573, 578, 594, 620, 640, 651, 655, 669, 687, 695, 699, 702, 706, 716, 721, 756, 759, 762, 869, 928, 1635, 1647, 1667, 1673, 1685, 1690, 1717, 1729, 1740, 1745, 1762, 1768, 1777, 1840, 1842, 1902, 1909, 1923, 1984, 2001, 2030, 2050, 2067, 2074, 2551	9, 17, 363, 375, 386, 484, 593, 661, 709, 778, 783, 837, 844, 856, 881, 926, 1008, 1619, 1631, 1653, 1699, 1706, 1731, 1973, 2037, 2105, 2572, 2588, 2594	12, 301, 312, 314, 339, 345, 358, 425, 746, 751, 825, 863, 890, 988, 1493, 1514, 1886, 1898, 1979, 1995, 2014, 2019, 2053, 2124, 2135, 2175, 2505, 2518, 2564	23, 28, 35, 41, 47, 59, 65, 68, 116, 123, 128, 132, 135, 143, 164, 168, 235, 240, 249, 251, 259, 262, 266, 273, 278, 282, 294, 307, 319, 343, 350, 380, 398, 430, 477, 979, 982, 995, 1001, 1015, 1029, 1033, 1040, 1047, 1051, 1061, 1068, 1073, 1097, 1112, 1125, 1145, 1164, 1171, 1180, 1188, 1192, 1194, 1198, 1204, 1215, 1234, 1240, 1245, 1249, 1255, 1262, 1269, 1275, 1281, 1289, 1296, 1304, 1305, 1309, 1320, 1330, 1336, 1344, 1361, 1370, 1375, 1385, 1391, 1410, 1418, 1423, 1425, 1432, 1441, 1475, 1526, 1532, 1538, 1543, 1551, 1578, 1593, 1600, 1614, 1941, 2119, 2148, 2153, 2161, 2166, 2189, 2196, 2197, 2206, 2225, 2240, 2249, 2257, 2284, 2299, 2311, 2312, 2320, 2328, 2333, 2342, 2348, 2351, 2360, 2371, 2378, 2383, 2391, 2396, 2399, 2411, 2419, 2427, 2426, 2447, 2453, 2457, 2461, 2471, 2493, 2511, 2524, 2540, 2583
Lane 2	315, 324, 346, 372, 393, 397, 404, 412, 416, 435, 472, 475, 478, 678, 688, 698, 703, 705, 715, 754, 1583, 1592, 1594, 1601, 1609, 1618, 1636, 1662, 1675, 1691, 1700, 1704, 1712, 1718, 1737, 1741, 1766, 1780, 1805, 1823, 1889, 1895, 1903, 1928, 1966, 2000	431, 468, 672, 718, 755, 1476, 1483, 1490, 1515, 2004, 2353	431, 468, 672, 718, 755, 1476, 1483, 1490, 1515, 2004, 2353	8, 13, 18, 22, 24, 29, 33, 40, 45, 50, 78, 125, 129, 165, 169, 174, 232, 236, 250, 255, 260, 263, 267, 272, 277, 283, 289, 302, 308, 722, 760, 763, 765, 769, 773, 777, 782, 817, 822, 826, 832, 842, 855, 857, 875, 882, 918, 948, 986, 991, 998, 1009, 1014, 1024, 1035, 1038, 1044, 1050, 1060, 1069, 1072, 1080, 1098, 1110, 1116, 1158, 1178, 1181, 1185, 1200, 1209, 1217, 1248, 1256, 1263, 1268, 1276, 1282, 1291, 1294, 1312, 1323, 1331, 1337, 1341, 1348, 1369, 1376, 1386, 1394, 1413, 1417, 1426, 1435, 1443, 1451, 1525, 1533, 1539, 1566, 1725, 1864, 2229, 2233, 2241, 2267, 2300, 2318, 2325, 2327, 2334, 2340, 2364, 2375, 2379, 2397, 2404, 2415, 2418, 2428, 2439, 2449, 2456, 2469, 2473, 2479
Lane 3	474, 479, 523, 529, 588, 597, 627, 632, 639, 650, 663, 674, 677, 689, 714, 724, 740, 744, 752, 766, 831, 876, 950, 1582, 1584, 1589, 1602, 1613, 1692, 1703, 1721, 1739, 1748, 1751, 1797, 1804, 1820, 1847, 1855, 1860, 1868, 1876, 1883,	786, 789, 848, 943, 1663, 1668, 1735, 1760, 2155	32, 180, 365, 392, 458, 463, 467, 605, 717, 809, 815, 819, 828, 860, 915, 937, 1595, 1728, 1834, 1845, 1965, 1999, 2016, 2258	7, 19, 27, 30, 39, 58, 62, 66, 93, 103, 105, 109, 119, 124, 130, 133, 136, 156, 177, 230, 233, 241, 256, 265, 280, 284, 287, 295, 311, 322, 357, 361, 395, 402, 427, 432, 438, 447, 534, 547, 993, 1000, 1045, 1048, 1056, 1062, 1075, 1083, 1102, 1109, 1114, 1118, 1123, 1184, 1201, 1205, 1210, 1216, 1243, 1250, 1259, 1264, 1274, 1283, 1290, 1301, 1319, 1327, 1387, 1397, 1405, 1407, 1411, 1415, 1421, 1427, 1448, 1454, 1485, 1489, 1499, 1528, 1537, 1540, 2006, 2160, 2193, 2202, 2242, 2286, 2308, 2316, 2322, 2330, 2335, 2344, 2347, 2362, 2363, 2382, 2389, 2398, 2405, 2410, 2423, 2436, 2443, 2450, 2474, 2506, 2515, 2523, 2527, 2535, 2543, 2550, 2562, 2571, 2580

	1890, 1944, 1993, 2013, 2046, 2055, 2065			
Lane 4	480, 501, 526, 533, 538, 543, 561, 566, 636, 649, 675, 690, 701, 711, 719, 727, 742, 747, 1553, 1581, 1586, 1596, 1616, 1629, 1640, 1664, 1787, 1806, 1833, 1846, 1857, 1925	548, 619, 631, 679, 713, 938, 949, 954, 1546, 1549, 1621, 1666, 1701, 1707, 1714, 1744, 1792, 1796, 1838, 1851, 1861, 1885	390, 493, 497, 528, 603, 801, 803, 811, 824, 839, 1507, 2023, 2236, 2558	14, 34, 38, 46, 57, 77, 86, 107, 111, 117, 122, 127, 134, 138, 162, 167, 173, 178, 185, 189, 225, 238, 257, 264, 270, 285, 288, 292, 298, 309, 313, 334, 353, 360, 382, 387, 394, 396, 400, 487, 598, 615, 779, 818, 827, 877, 887, 944, 989, 999, 1004, 1071, 1085, 1103, 1107, 1113, 1119, 1121, 1130, 1251, 1253, 1277, 1284, 1287, 1297, 1317, 1333, 1349, 1363, 1390, 1399, 1403, 1408, 1414, 1419, 1422, 1444, 1449, 1469, 1482, 1481, 1529, 1536, 2056, 2071, 2140, 2157, 2168, 2187, 2195, 2203, 2208, 2215, 2222, 2265, 2277, 2302, 2317, 2324, 2331, 2336, 2345, 2350, 2357, 2361, 2414, 2424, 2431, 2441, 2448, 2454, 2460, 2466, 2475, 2508, 2516, 2521, 2553, 2570, 2575, 2581, 2590
Lane 5	810, 816, 1659, 1681, 1788, 1897, 1915, 2002, 2018, 2094	792, 850, 1617, 1623, 1782, 1987	527, 544, 686, 712, 743, 821, 838, 1168, 1676, 1689, 1770	21, 25, 56, 64, 70, 75, 90, 113, 118, 121, 126, 131, 137, 142, 151, 160, 179, 186, 217, 253, 258, 271, 276, 281, 286, 293, 310, 318, 321, 327, 333, 356, 377, 383, 388, 391, 399, 401, 445, 537, 580, 582, 657, 673, 680, 741, 749, 1124, 1144, 1148, 1153, 1187, 1193, 1197, 1257, 1292, 1307, 1310, 1351, 1388, 1401, 1412, 1416, 1420, 1450, 1480, 1492, 1494, 1497, 1508, 1548, 1552, 1555, 1557, 1590, 1597, 1605, 1734, 1841, 1836, 2047, 2171, 2179, 2183, 2188, 2198, 2209, 2321, 2341, 2392, 2401, 2442, 2444, 2478, 2497, 2499, 2507, 2545, 2552, 2596

Table C.4 Vehicle IDs and corresponding trajectory groups in US-101-3 data.

Lane ID	Trajectory groups *			
	FaADFS	**ADFS**	**FaADF**	**ADF**
Lane 1	97, 243, 249, 257, 264, 269, 276, 283, 293, 299, 309, 310, 349, 355, 357, 365, 376, 380, 385, 391, 396, 401, 409, 446, 705, 709, 826, 929, 1058, 1171, 1179, 1194, 1197, 1201, 1416, 1495	238, 261, 315, 340, 371, 725, 759, 767, 786, 790, 793, 804, 814, 821, 832, 914, 923, 1074, 1080, 1099, 1206, 1391, 1399, 1610, 1636	11, 29, 102, 105, 353, 406, 416, 424, 430, 441, 563, 684, 757, 772, 798, 812, 846, 850, 861, 900, 919, 935, 956, 968, 980, 986, 994, 1063, 1127, 1159, 1165, 1213, 1223, 1230, 1245, 1264, 1459, 1461, 1486, 1512, 1548, 1551, 1565, 1572, 1580, 1603, 1704, 1883	14, 24, 32, 38, 41, 47, 55, 62, 88, 110, 429, 436, 457, 465, 478, 487, 496, 497, 517, 523, 528, 550, 570, 574, 577, 586, 590, 604, 612, 617, 622, 627, 633, 643, 651, 662, 667, 688, 695, 780, 837, 853, 867, 878, 884, 888, 893, 908, 974, 976, 990, 1000, 1005, 1008, 1013, 1021, 1025, 1045, 1050, 1055, 1071, 1093, 1122, 1133, 1137, 1146, 1148, 1155, 1237, 1242, 1270, 1279, 1283, 1289, 1295, 1312, 1316, 1318, 1326, 1332, 1335, 1340, 1347, 1351, 1366, 1375, 1378, 1383, 1388, 1396, 1475, 1517, 1528, 1534, 1537, 1542, 1557, 1584, 1588, 1593, 1605, 1615, 1618, 1633, 1645, 1650, 1662, 1665, 1676, 1694, 1712, 1718, 1723, 1726, 1732, 1737, 1752, 1756, 1762, 1769, 1774, 1778, 1788, 1793, 1797, 1808, 1844, 1854, 1859, 1864, 1871, 1877
Lane 2	186, 222, 236, 239, 246, 253, 258, 262, 270, 280, 300, 308, 328, 334, 1004, 1009, 1023, 1032, 1046, 1052, 1064, 1070, 1075, 1084, 1094, 1102, 1126, 1160, 1172, 1177, 1212, 1218, 1222, 1228, 1232, 1236, 1241,	206, 212, 217, 244, 265, 345, 438, 450, 883, 992, 995, 1041, 1149, 1251, 1299, 1329, 1334, 1462, 1487, 1525, 1576, 1581, 1589, 1594, 1757, 1789, 1860, 1878, 1900, 1902, 1909, 1917, 1932	27, 181, 198, 608, 618, 777, 815, 969, 977, 1014, 1018, 1079, 1278, 1469, 1506, 1509, 1533, 1538, 1547, 1568, 1573, 1585, 1700, 1888, 1926	10, 22, 92, 95, 101, 443, 477, 482, 488, 493, 498, 518, 522, 545, 557, 562, 566, 584, 589, 603, 611, 625, 632, 642, 646, 661, 675, 689, 696, 704, 707, 736, 748, 756, 760, 766, 773, 783, 795, 799, 805, 827, 831, 868, 873, 877, 886, 894, 899, 902, 922, 926, 934, 948, 960, 987, 1309, 1443, 1543, 1552, 1558, 1604, 1613, 1616, 1637, 1644, 1647, 1651, 1663, 1670, 1677, 1691, 1705, 1710, 1714, 1719, 1733, 1738, 1741, 1868, 1884

	1246, 1254, 1297, 1322, 1336, 1354, 1360, 1365, 1370, 1374, 1379, 1384, 1482, 1513, 1562, 1779, 1873, 1915			
Lane 3	213, 237, 255, 259, 266, 271, 289, 292, 314, 320, 326, 356, 393, 1260, 1290, 1294, 1298, 1303, 1315, 1348, 1352, 1357, 1367, 1373, 1411, 1417, 1423, 1497, 1508, 1514, 1536, 1539, 1544, 1553, 1590, 1595, 1602, 1904, 1923	245, 281, 286, 296, 304, 317, 351, 378, 765, 781, 958, 970, 1275, 1570, 1614, 1896	126, 136, 158, 178, 190, 220, 277, 2001, 761, 808, 882, 988, 1224, 1233, 1240, 1243, 1320, 1323, 1599, 1747, 1758, 1782, 1796, 1814, 1818, 1865, 1881, 1889	21, 26, 34, 40, 45, 64, 70, 86, 91, 131, 141, 166, 173, 184, 199, 207, 224, 227, 367, 404, 420, 453, 459, 462, 481, 486, 509, 519, 529, 546, 560, 628, 634, 638, 649, 677, 680, 686, 690, 693, 703, 706, 712, 715, 720, 730, 733, 743, 749, 778, 787, 910, 915, 932, 1035, 1039, 1047, 1051, 1054, 1062, 1065, 1091, 1095, 1101, 1106, 1204, 1209, 1215, 1229, 1620, 1634, 1639, 1643, 1652, 1666, 1671, 1675, 1692, 1696, 1702, 1709, 1715, 1841, 1861, 1899
Lane 4	119, 156, 174, 197, 202, 211, 218, 223, 241, 247, 251, 254, 260, 267, 272, 275, 284, 961, 979, 985, 989, 1006, 1048, 1053, 1060, 1067, 1077, 1114, 1185, 1203, 1227, 1259, 1266, 1276, 1338, 1342, 1387	25, 85, 90, 114, 125, 240, 1011, 1183, 1397, 1402	37, 297, 301, 305, 323, 382, 419, 434, 454, 485, 489, 606, 614, 708, 750, 927, 954, 993, 1040, 1069, 1300, 1317, 1441, 1503, 1530, 1532, 1638	31, 48, 52, 66, 331, 388, 421, 428, 461, 463, 470, 479, 508, 538, 623, 630, 636, 648, 681, 685, 691, 713, 718, 723, 726, 754, 758, 764, 769, 775, 779, 784, 788, 792, 794, 797, 802, 806, 892, 898, 903, 931, 1498, 1515, 1520, 1526, 1550, 1555, 1566, 1574, 1579, 1591, 1596, 1601, 1607, 1617, 1621, 1625, 1648, 1653, 1661, 1672, 1683, 1688, 1701, 1708, 1735, 1754, 1759, 1772, 1791, 1804, 1815, 1819, 1825, 1836, 1842, 1847, 1862, 1867, 1870, 1885, 1891, 1894, 1901, 1905, 1911, 1930
Lane 5	268, 332, 422, 983, 1002, 1007, 1038, 1066, 1072, 1097, 1163, 1217, 1219	252, 997, 1012, 1034, 1092	73, 117, 134, 235, 389, 398, 418, 431, 455, 475, 855, 918, 1043, 1181, 1193, 1235, 1325, 1339	36, 42, 71, 87, 93, 154, 171, 203, 208, 215, 221, 2000, 256, 298, 383, 460, 552, 607, 620, 624, 655, 687, 719, 724, 728, 732, 751, 776, 782, 785, 800, 803, 807, 813, 818, 824, 829, 858, 941, 978, 1262, 1265, 1277, 1291, 1296, 1301, 1372, 1395, 1398, 1403, 1413, 1425, 1426, 1432, 1460, 1479, 1522, 1535, 1567, 1583, 1592, 1609, 1622, 1624, 1627, 1635, 1640, 1641, 1654, 1659, 1669, 1693, 1730, 1749, 1816, 1820, 1824, 1826, 1831, 1837, 1848, 1863, 1921