TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–13

DOI: 10.1177/03611981211031884
journals.sagepub.com/home/trr
SAGE

Understanding the Lateral Dimension of Traffic: Measuring and Modeling Lane Discipline

Rafael Delpiano[1]

Abstract

There is growing interest in understanding the lateral dimension of traffic. This trend has been motivated by the detection of phenomena unexplained by traditional models and the emergence of new technologies. Previous attempts to address this dimension have focused on lane-changing and non-lane-based traffic. The literature on vehicles keeping their lanes has generally been limited to simple statistics on vehicle position while models assume vehicles stay perfectly centered. Previously the author developed a two-dimensional traffic model aiming to capture such behavior qualitatively. Still pending is a deeper, more accurate comprehension and modeling of the relationships between variables in both axes. The present paper is based on the Next Generation SIMulation (NGSIM) datasets. It was found that lateral position is highly dependent on the longitudinal position, a phenomenon consistent with data capture from multiple cameras. A methodology is proposed to alleviate this problem. It was also discovered that the standard deviation of lateral velocity grows with longitudinal velocity and that the average lateral position varies with longitudinal velocity by up to 8 cm, possibly reflecting greater caution in overtaking. Random walk models were proposed and calibrated to reproduce some of the characteristics measured. It was determined that drivers' response is much more sensitive to the lateral velocity than to position. These results provide a basis for further advances in understanding the lateral dimension. It is hoped that such comprehension will facilitate the design of autonomous vehicle algorithms that are friendlier to both passengers and the occupants of surrounding vehicles.

Traditionally, traffic flow theory has concentrated its analyses on the longitudinal dimension of vehicle movement. The main macroscopic traffic variables (velocity, flow, density) are all intrinsically linked with this focus. The need for both simplicity and scalability has led modelers to discretize the lateral dimension, reducing it to an integer variable representing a simple lane index number.

The lateral dimension has, however, an influence on vehicle traffic behavior that is not insignificant. During the execution of a lane change, a vehicle affects both its current lane and its target lane (*1*). Its lateral position affects the comfort of surrounding vehicles (*2*) as well as that of its own occupants (*3*) and is related to the relaxation phenomenon through anticipation (*4, 5*). Standard deviation of lateral position has been studied as a road safety (*6, 7*) and a performance (*8–10*) metric. Lane width (*11, 12*) and the consequent distances to nearby vehicles, road barriers and signs (*13–15*) affect vehicle flow capacity and velocity, in a phenomenon known as lateral friction.

Drivers tend to increase lateral distance when another vehicle is located directly parallel to them (*16*). This distance increases with velocity (*17*), perhaps caused by a subjective perception of safety risk. An autonomous vehicle (AV) overtaking a conventional one that does not take this into account and stays in the center of its lane may be seen as a danger both for its own occupants and those of the vehicle it is overtaking, depending on the latter's lateral position. This reduced safety perception could result in greater lateral friction, affecting traffic efficiency and potentially creating an actual unsafe situation.

Note that of the 38 collisions reported by Google/Waymo through October 2018 under California's AV testing regulations, there were two where the car operating in driverless mode was neither stationary nor hit from behind, implying that some responsibility on the part of the AV cannot be ruled out *a priori*. In both cases, the

[1]Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Santiago, Chile

Corresponding Author:
Rafael Delpiano, rdelpiano@uandes.cl

accidents were side impacts related to the cars' lateral positions (*18*).

In light of the foregoing, a clearer understanding is needed of the lateral dimension of traffic. This greater knowledge could facilitate a faster adoption of AVs and a more harmonious transition period during which both human and autonomous drivers will continue to share the roads.

There has been a growing interest for some years in comprehending and modeling the lateral behavior of vehicles and drivers. The first study to call attention to the importance of this issue was that of Gunay (*2*). It has since been followed by a series of works presenting two-dimensional models focused primarily on situations of poor lane discipline (*19–23*). However, these formulations either do not model lateral position choice or, if they do, neglect the existence of lanes.

The first studies modeling lane changing as a non-instantaneous process have also appeared in recent years (*1*). In certain cases they have also taken account of lateral position during the lane-changing maneuver itself (*24, 25*). To the author's knowledge, however, only one model, that developed by the author, has fully incorporated the lateral dimension into the context of multi-lane traffic (*4, 5, 26*). This formulation, called the two-dimensional microscopic traffic model (2D- μ TM), is based on a social force paradigm (*27, 28*) in which drivers are modeled as though their vehicles were subject to physical forces. Thus, a *lane force* is defined as representing the will of the driver to center their vehicle in the chosen lane. That study left for future work a proper refining or calibration of such force.

In the present article, *lane discipline* is defined as a set of decisions and actions that determine lateral position choice. Lane discipline can be measured by its effect (i.e., the lateral position) and its importance lies in the extent to which it weakens or strengthens the efficiency of vehicle traffic and drivers' safety risk perception. This study is confined to the lateral positions of vehicles that remain in their lanes, as opposed to those that are in the process of changing lanes. As part of the analysis, a number of metrics will be defined for quantifying lane discipline. They will then be employed to calibrate two models to be proposed.

For purposes of this study, vehicle trajectories were obtained from the data sets constructed by the Next Generation SIMulation (NGSIM) program of the U.S. Federal Highway Administration (*29–31*). Although they have been subject to some criticism for the noise in the measurements (*32*), they continue to be the standard source in this area of research and corrections to mitigate their deficiencies can be made.

The objective of this paper is twofold: to measure and model lane discipline. Specifically, it aims to model by means of refining and/or calibrating the aforementioned lane force, as well as to propose metrics for quantifying behavior in the lateral dimension, so that the models can be assessed. Unfortunately this cannot be done without first addressing the issue of the noise in NGSIM data.

Lane Discipline in the 2D- μ TM

In the aforementioned microscopic 2D social force model three forces were defined: (i) acceleration force f^a, which is exclusively longitudinal and represents the desire of drivers to accelerate up to their free-flow velocity; (ii) lane force f^l, which is exclusively lateral and controls lane discipline; and (iii) repulsive force f^r, which is bidimensional and represents driver caution, leading drivers to keep their distance from other vehicles, especially when they approach or are already close (*4, 26*).

Lane force is necessary to counteract the effect of the lateral component of the repulsive force. The latter is especially required when an overtaking vehicle attempts to maintain sufficient clearance from vehicles in neighboring lanes that are particularly wide or very close so as to avoid potentially negative impacts on surrounding traffic.

Lane force in the 2D- μ TM for the experiments in the author's earlier studies was defined as keeping a vehicle centered in position x_l, the center of lane l, as follows:

$$\mathbf{f}_i^l(t) = \begin{bmatrix} -(x_i(t) - x_l)k_1 - v_i^x(t)k_2 \\ 0 \end{bmatrix} \quad (1)$$

where the first component expresses the magnitude of the lateral acceleration aimed at centering vehicle i in position x_l, $v_i^x(t)$ and $x_i(t)$ are respectively the lateral velocity and position of vehicle i, and k_1 and k_2 are parameters measuring the intensity of the response per unit of lateral distance from the lane center (in units of time $^{-2}$) and per unit of lateral velocity (in units of time $^{-1}$). Note that social forces are not physical, so neither is mass. For simplicity vehicles are assumed a *social mass* of 1, so that $f = a$.

Following the nomenclature of the 2D- μ TM, in the present study lateral position, velocity, and acceleration will be represented by x, v_x and a_x, while longitudinal position, velocity, and acceleration will be represented by y, v_y and a_y. When the two dimensions are graphed, the lateral one will be on the horizontal axis and the longitudinal one on the vertical one. Finally, on these graphs the general forward direction of the vehicles is upwards.

As a working assumption in the earlier study, lane force played the dual role of keeping vehicles in their lanes and controlling the execution of lane changes. As regards the latter, k_1 and k_2 were assumed to be 0.25 and 1, respectively, which meant that a lane-changing vehicle

touches the central line of the destination lane after approximately 5 s, following Moridpour et al. (*1*).

The combined effect of these assumptions was that the modeled vehicles tended to be overly centered in their lanes. A more refined definition of lane force that would generate more realistic lane discipline results was left for later analysis. Concretely, what is needed is a new functional form for $\mathbf{f}_i^l(t)$ and/or different values for its parameters. In this paper, some ways for achieving these necessary improvements in the context of single-lane trajectories are proposed (i.e., excluding lane changes).

While refining and/or calibrating the lane force is a rather small step toward understanding human driving behavior in the lateral dimension, it is a very necessary one. In previous work, the author has shown that real-world lateral position choices are consistent with the sometimes conflicting goals of keeping centered in a lane and being at a safe distance from other vehicles to avoid risks. There is a need initially to understand each of these to assess how they interact and to grasp the big picture.

NGSIM Data Sets and the Error Problem

As part of the NGSIM program, vehicle trajectory databases were constructed from video images. Three of these data sets, known as Prototype, I-80 and US-101 (the latter two named after the freeways they were collected from), contain data on uninterrupted traffic. All three include information on lateral and longitudinal vehicle positions over time for the road segment under study during the observation period. The main characteristics of each data set, including the video duration, time step between positional data, and road segment length, are shown in Table 1.

The data themselves were extracted from video images captured by a series of cameras trained on successive sections of the segment studied and have been analyzed by various researchers (*33*). The trajectories were constructed with the help of the image recognition technology available at the time, which introduced noise into the measurements. Later corrections were made but they were not always consistent from one data set to the next and not all of the noise was successfully eliminated.

According to Punzo et al. (*32*), the noise includes an abundance of outliers, suggesting the noise has a heavy-tailed distribution. Coifman and Li (*34*) describe this distribution and demonstrate the apparent introduction of bias in the longitudinal dimension from projection errors which "increase with height of the vehicle and distance away from the camera." Moreover, "the original NGSIM data extraction did a poor job at the transition from one camera to the next." This appears to be the case also for the lateral dimension, as will be shown. The noise tends naturally to be magnified in calculations of finite differences in velocities and accelerations, resulting in implausible values, especially for the latter. Note that lateral acceleration was calculated using finite differences 2, since the data sets do not provide these data.

$$a_x(t) = \frac{v(t + dt) - v(t)}{dt} \tag{2}$$

A series of corrections are suggested in Punzo et al. (*32*), including a low-pass filter, a likelihood analysis of inter-vehicle spacings and acceleration variations, and a spectral analysis assuming that frequencies above 1 Hz are implausible. All of these corrections are concerned exclusively with the trajectories' longitudinal dimension, however, taking no account of the lateral coordinates. In the present paper, therefore, the more widely available original data will be used, whose two dimensions are free of any differential treatment.

With a view to partially mitigating the effects of the noise, preprocessing of the data was undertaken. This began by taking a subsample every five time steps. A new set of velocities was then calculated based on finite differences. To eliminate lane-change effects, all vehicle observations for which the vehicle was not at that instant in the same lane it was observed in 5 s before or after were discarded. Since this procedure also excluded vehicles close to the start or end of the road segment, only those located more than 100 m from either end were retained. This also served to avoid introducing velocity biases. Furthermore, only the first five lanes in each data set were considered given that the sixth one includes acceleration and deceleration lanes, and non-constant width.

It should also be noted here that the original data appear in fact to have been subjected to some corrections. In particular, two of the data sets show signs of velocity adjustments. As shown in Figure 1, there is a number of obviously artificial peaks at multiples of 5 ft/s.

Table 1. The Three NGSIM Uninterrupted Traffic Data Sets and Their Main Characteristics

Name	Location	Date	Start	Duration	Time step	Length
Prototype	I-80 Freeway, CA, USA	December 2003	2:35 p.m.	30 min	1/15 s	899 m
I-80	I-80 Freeway, CA, USA	April 2005	4:00 p.m.	45 min	0.1 s	503 m
US-101	US-101 Freeway, CA, USA	June 2005	7:50 a.m.	45 min	0.1 s	640 m

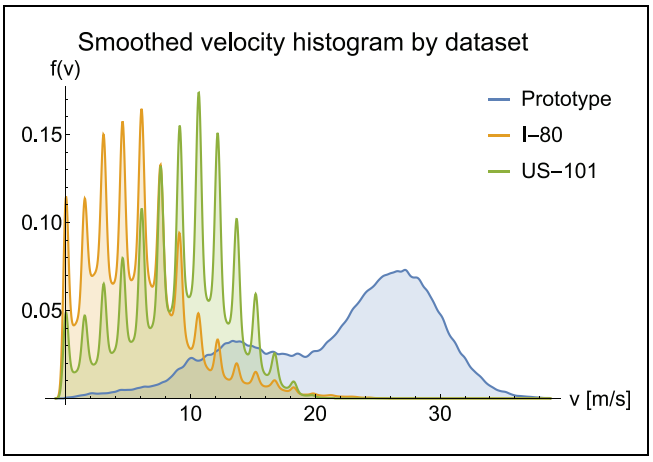

Figure 1. Smoothed velocity histogram by dataset. Density estimated using kernel density estimation (KDE). Velocities calculated for finite differences.

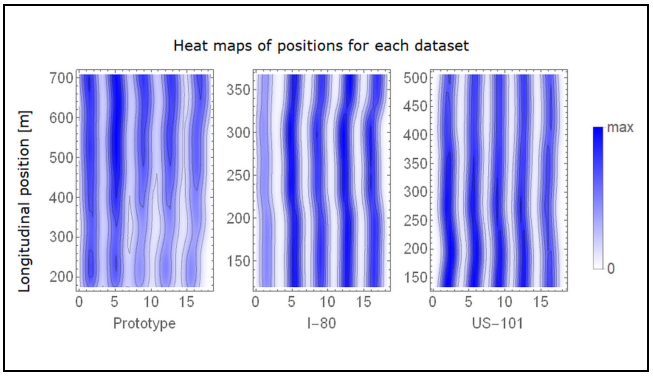

Figure 2. Heat maps of vehicle positions for each data set. The bias in lateral positions may be caused in part by camera perspective effects and the use of video capture from different cameras.

With regard to the bias from perspective and the use of multiple cameras, significant discrepancies are observed in the mean lateral position between different longitudinal positions of a given data set (after eliminating vehicles changing lanes). This is illustrated by a wide difference of more than 57 cm in mean lateral position (t -statistic >150, p -value $<10^{-12000}$) at 70 m of longitudinal distance in the I-80 dataset. Such differences are observed to a greater or lesser degree in all three data sets, as seen in the heat maps of the observed vehicle positions shown in Figure 2. The effect varies from lane to lane, thus affecting the measurement of the separation between vehicles in contiguous lanes.

More evidence pointing to this may be appreciated in Figure 3, which tracks lateral position in the I-80 data set. For each lane l, the graph in the figure shows the mean lateral position E $[x]$ relative to the lane center as a

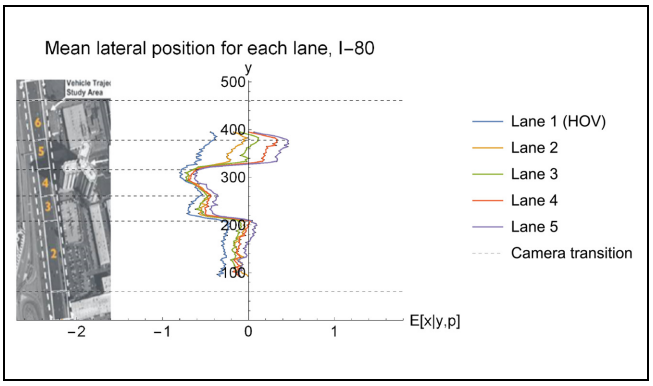

Figure 3. Mean lateral vehicle position as a function of longitudinal position for each lane. Abrupt trend changes occur at coverage transitions between successive cameras. Based on Federal Highway Administration data set (*30*).

function of the longitudinal position y. The road segment graphed is divided into six sections in the photo, each one covered by a different camera. At each transition from one camera to the next, there is an abrupt change in trend.

Given the above evidence of bias, the data preprocessing included, for purposes of comparison, a lateral position correction in which for each lane and neighboring longitudinal position, the locally obtained mean is subtracted from the original lateral position data. The resulting corrected lateral position is denoted $x^* = x - \mathrm{E}[x|y, l]$, from which the corrected lateral velocity v_x^*, the corrected lateral acceleration a_x^*, and so forth, are then calculated using finite differences.

The corrected lateral position data do not, however, totally replace the original data in the analyses. (i) On one hand, the mean as calculated has a standard error that oscillates around 2.5 cm, meaning that it is possible for the real mean to be as much as 5 cm away in either direction; and (ii) forcing the means to zero results in a loss of information in that it implies the real means are, in fact, zero. But this is not necessarily the case so it would have to be proved, and there are documented situations where it is not true (*35*).

The possibility of error measurement distortions or overcorrections must also be considered when performing experiments with such data and drawing conclusions from them. Some examples of this are focusing on the means or medians of large quantities of data, thus ignoring noise in the tails; using longer time intervals for calculation finite differences; and focusing on trends rather than exact numerical results.

In any case, the metrics and methodology proposed here are applicable to any possible source of data on lateral and longitudinal positions. As new data sets become available, the results presented in this study can be updated and refined.

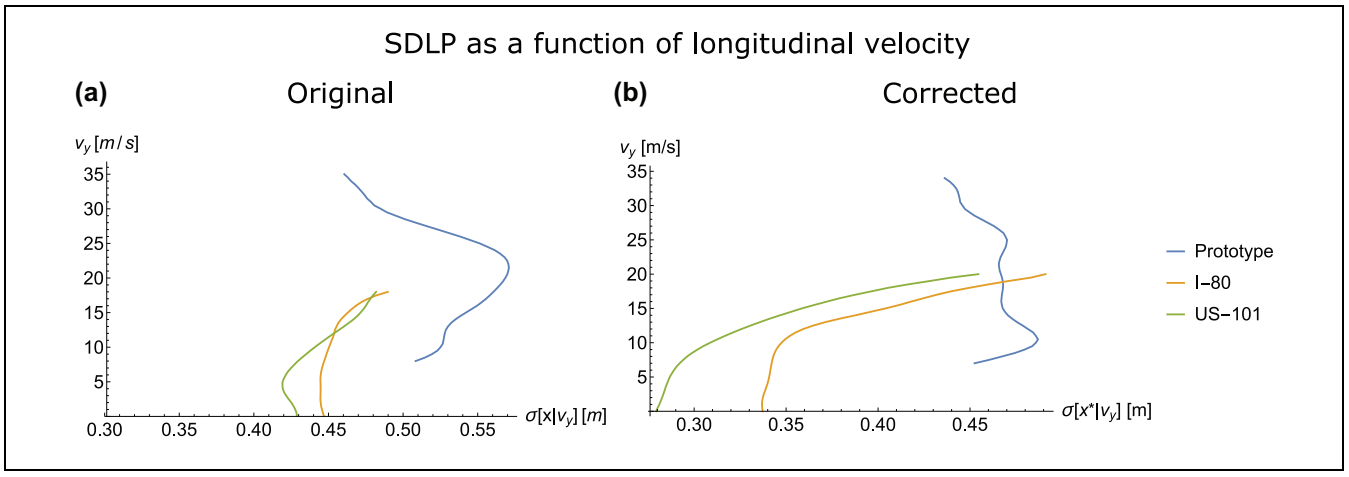

Figure 4. Standard deviation of the lateral position as a function of longitudinal velocity: (*a*) original NGSIM data and (*b*) corrected data.

Characterization of Lane Discipline

In what follows a series of metrics for lane discipline measured from the NGSIM data sets are described. Some of these characteristics are general sample moments of the kinematic variables (means and standard deviations) whereas others are the relations between some of these moments and other kinematic variables.

A key metric for present purposes is the standard deviation of lateral position (SDLP). Its relevance lies in its suitability to describe drivers' departure from the center of their lane under similar lane and vehicle width conditions, as well as its use in the literature as a road safety metric means. The standard deviations of lateral position for the three NGSIM data sets, both in their original and corrected versions (the former calculated directly from the NGSIM data) are given in Figure 2. Using the original data (labeled SDLP), the values range from 44 cm to 52 cm. With the corrected data (SDLP *), the standard deviations are 7 to 13 cm lower. In both cases, the higher values for Prototype may be the result of greater velocities or less attentive driver behavior given that the data were collected outside of peak hours.

The evolution of SDLP as a function of longitudinal velocity is shown in Figure 4 (vertical axis). It was obtained by estimating the conditional variance V $[x|v_y]$ using kernel density estimation (KDE). For each data set the most representative range of velocities is graphed. As can be seen, for all three sets, both original and corrected, SDLP increases with increasing velocity until about 20 m/s, with the exception of the corrected Prototype data where its trend stops rising at around 10 m/s.

Note that unless otherwise indicated, the same method for calculating conditional moments from the density function using KDE is applied to the other relationships between the kinematic variables presented here.

Another useful set of characteristics are those describing the profiles of lateral velocity v_x and its relation to longitudinal velocity. These will be identified by constructing histograms and calculating the lateral velocity means and standard deviations. Given the coexistence of a majority of lateral velocities that are very small with a not-insignificant proportion of them that are several orders of magnitude larger, it will be convenient also to express lateral velocity on a logarithmic scale. We therefore define a variable denoted log speed as the logarithm of the absolute value of lateral velocity. Since any velocities numerically equal to zero will be excluded from this logarithmic variable, it is worth checking what percentage of the results they constitute.

The means and standard deviations of the lateral velocities and log speeds are shown in Figure 3 for the three data sets. As may be expected, the lateral velocity means are very small.

The approximate distribution of log speed for each data set (original and corrected) is set out in Figure 5. Note that the velocities are concentrated between 0.01 and 0.1 m/s. The increasing relation between the standard deviation of lateral velocity and longitudinal velocity is evident in Figure 6. Moreover, the relationship seems to be fairly close to linear. This would be consistent with a steering angle inversely proportional to longitudinal velocity.

The mean lateral position is biased to the left for Prototype and I-80 (−26 and −24.2 cm respectively), and close to zero for US-101 (0.67 cm). For all three, the original data exhibit statistically significant deviations from the lane center, but become zero after correction. As regards the much smaller value for the US-101 uncorrected case, the author cannot say whether or not it was subjected to correction, but note that this database is the most recent of the three. Also, with the available

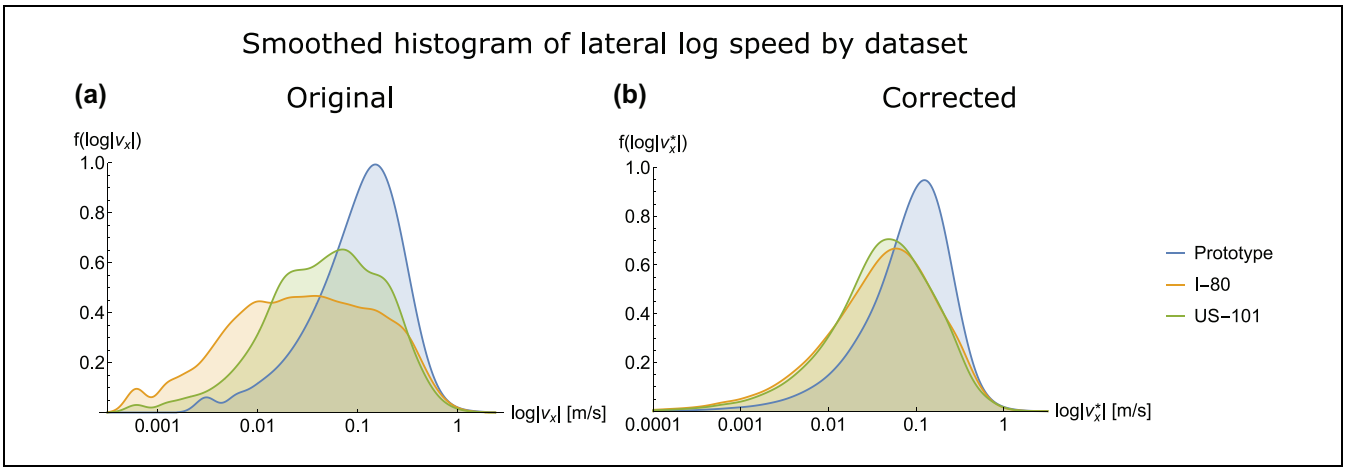

Figure 5. Smoothed histogram of lateral log speed by data set: (*a*) original NGSIM data and (*b*) corrected data.

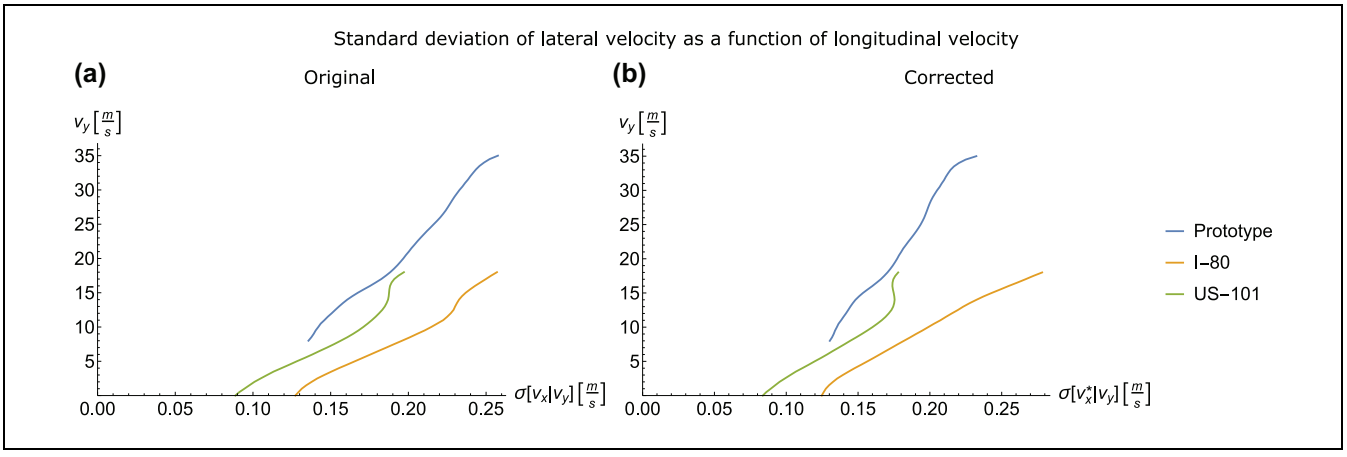

Figure 6. Standard deviation of lateral velocity as a function of longitudinal velocity: (*a*) original NGSIM data and (*b*) corrected data.

information it is not possible to determine whether the off-center results for Prototype and I-80 are caused exclusively by perspective bias or some objective effect of driver preference or perception.

Finally, as a first approximation to a direct measurement of the lane force parameters formulated in the 2D-μ TM (see 4), the lateral acceleration a_x is characterized as a function of lateral position and lateral velocity. Once again, the conditional expected values E $[a_x|x]$, E $[a_x|v_x]$ are used. For the present analysis the methodology is different. As with the velocities, there are many minor lateral accelerations and a considerable number of major ones. To study their relationship with other variables, however, there is a need to separate the positive values from the negative ones, so in this case the use of the absolute value will be avoided. Likewise, constant bandwidth KDE is not applied, since it cannot be adapted to the aforementioned characteristics. What is utilized instead is a moving average for a fixed quantity of data, sorting the independent variable by value.

The resulting relationships for each data set, original and corrected, are shown in Figures 7 and 8. As may be observed, in all cases the relationship is reasonably linear, especially near the origin. Deviations from this trend, as steeper slopes, occur at about two standard deviations from the mean (see 3). In the case of velocity, a region of average acceleration close to zero observable in the original data disappears in the corrected data.

The dashed trend lines added to the graphs were obtained by linear regression. Those describing the relationship between acceleration and lateral positions using the original data (Figure 7, at left) cut the horizontal axis between 20 and 30 cm, which is consistent with the off-center mean lateral position.

The linear regression parameter estimates are summarized in Figure 4. All of the slopes are more pronounced for the corrected data, suggesting that perspective bias appears to attenuate the relationships.

ith these results, simple linear models can be constructed for E $[a_x|x, v_x]$. As an example, the I-80 data set gives

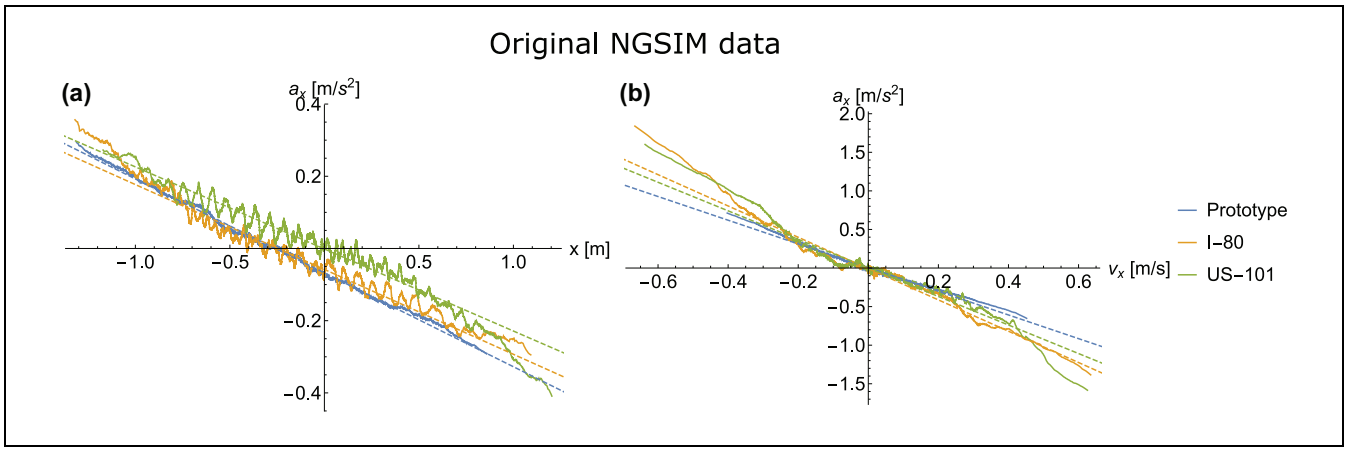

Figure 7. Relationship between lateral acceleration, lateral velocity, and distance from lane center with original coordinates, by data set: (*a*) mean lateral acceleration versus position, (*b*) mean lateral acceleration versus velocity.
Dashed lines represent the corresponding linear regressions.

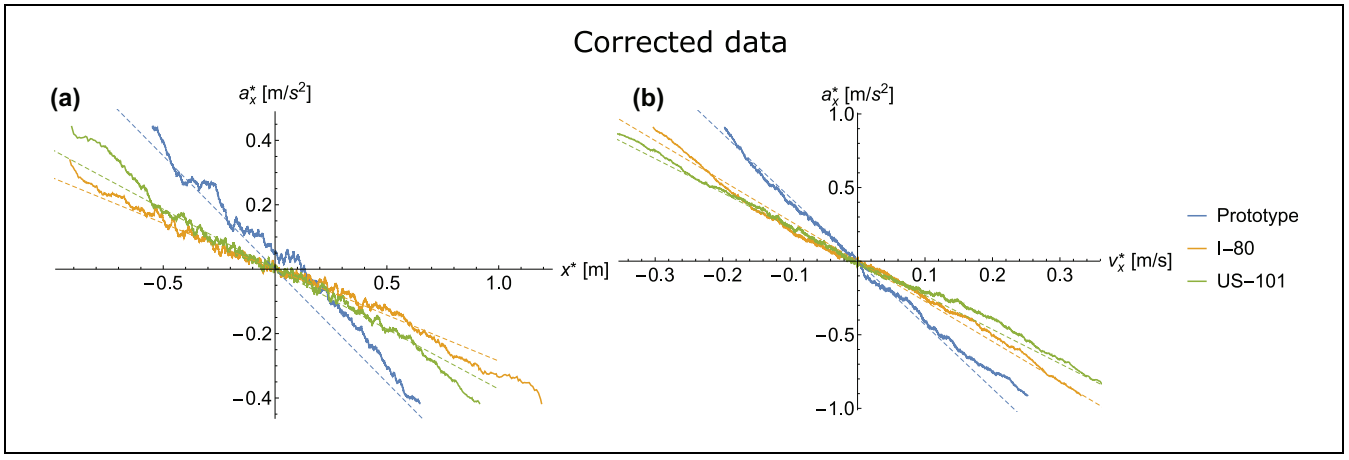

Figure 8. Relationship between lateral acceleration, lateral velocity, and distance from lane center with corrected coordinates, by data set: (*a*) mean lateral acceleration versus position, (*b*) mean lateral acceleration versus velocity.
Dashed lines represent the corresponding linear regressions.

Table 2. Standard Deviation of Lateral Position (SDLP) from Original and Corrected (*) Data, Relative to Lane Center, by Data Set

Dataset	Prototype	I-80	US-101
SDLP (cm)	52.8	44.9	43.8
SDLP (cm)	45.9	35.4	31.1

$$\mathrm{E}[a_x|x, v_x] = -0.29x - 2.73v_x. \quad (3)$$

The standard deviation of the residual error for each data set is 1.67 m/s ^{2} for Prototype, 0.59 for I-80, and 0.54 for US-101. The error is calculated as the difference between actual acceleration and the value predicted by the data sets' linear models in the form given in Equation 3 with the coefficients obtained from the corrected data which is in (Table 4). The distributions of the errors have heavier tails than a standard normal distribution.

Models

In this section two models are proposed that reproduce some of the lane discipline characteristics discussed in the previous section. The idea is to refine the aforementioned lane force concept by calibrating the k_1 and k_2 parameters in (*4*). With this in mind, an attempt is made to separate

Table 3. Mean and Standard Deviation (SD) of Lateral Velocity and Lateral Log Speed

	Original NGSIM data					Corrected data			
	v_x [m/s]		$\log_{10}\lvert v_x\rvert$		Velocity Zero	v_x^* [m/s]		$\log_{10}\lvert v_x^*\rvert$	
	Mean	SD	Mean	SD		Mean	SD	Mean	SD
Prototype	0.006	0.20	−1.02	0.46	0.76%	−0.003	0.18	−1.13	0.52
I-80	−0.002	0.16	−1.55	0.73	5.15%	$8.19 \cdot 10^{-4}$	0.15	−1.42	0.67
US-101	0.002	0.15	−1.33	0.58	2.15%	$8.18 \cdot 10^{-5}$	0.14	−1.40	0.63

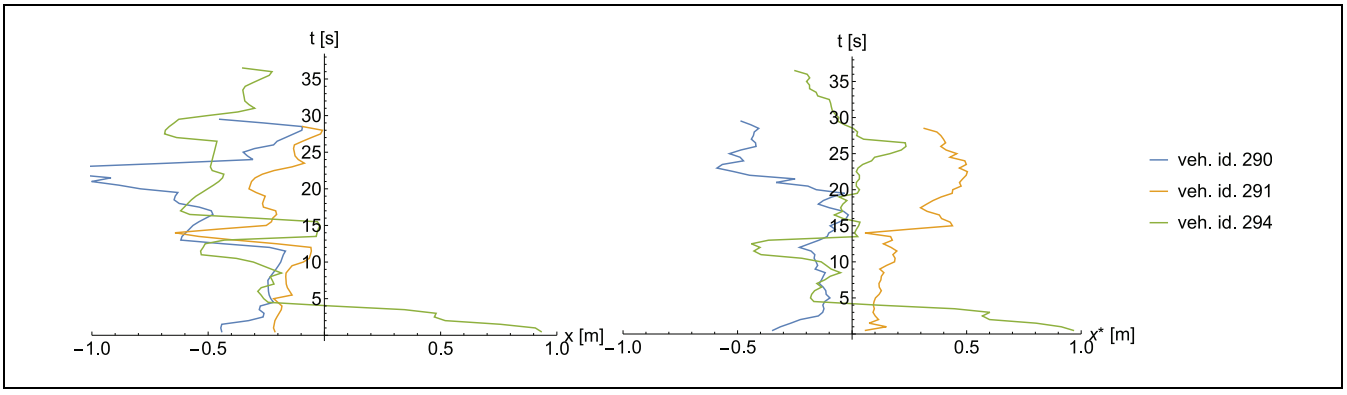

Figure 9. Selected actual vehicle trajectories using original (*left*) and corrected (*right*) data.

the predictable response to the lateral position and velocity from other influences such as the various stimuli attributable to neighboring vehicles, perception errors, tolerance thresholds, and so forth.

The models therefore include a deterministic term for lane force and a random term to represent whatever effects lane force does not cover. For the sake of simplicity the influence of longitudinal velocity is left out of the analysis. Thus, the proposed formulations are concerned only with the lateral axis. Both take the form of a random walk (*36*).

The models' quality is evaluated by how well they fit the metrics proposed in the previous section. The specific metrics used are SDLP, standard deviation of lateral velocity, and the mean and standard deviation of lateral log speed. The evolution of the lateral positions (using the original and corrected data sets) of three arbitrarily selected vehicles from the first 15 min segment of the I-80 data set (original and corrected) is shown in Figure 9.

The models range from the very simple and analytically tractable to the relatively complex but more accurate in reproducing reality. The simplest one has just two parameters while the most complex one requires five.

For each model, 4,000 random walks were run from scratch, each consisting of 100 time steps after a 50-time step warm up period. Both numbers were selected to approximately match the number of trajectories and data points of the NGSIM datasets.

First-Order Autoregressive Model

The first model is a first-order autoregressive formulation, AR(1). Though extremely basic it has a known analytic solution, equivalent in discrete time to the Ornstein-Uhlenbeck process (*37*). It is written as follows:

$$v_x(t) = -kx(t) + \xi(t), \xi(t) \sim \mathcal{N}(0, \nu^2) \tag{4}$$

The model has only two parameters: k, for the deterministic part, and ν, for the random part. $\mathcal{N}$ represents the normal distribution. Applying Euler's method of approximation with a step size of 0.5 s (to be consistent with the subsample of the original data), gives

$$\begin{aligned} x(t+dt) &\approx x(t) + dtv_x(t) \\ &= x(t) + \frac{1}{2}(-kx(t) + \xi(t)) \\ &= \left(1 - \frac{k}{2}\right)x(t) + \frac{\xi(t)}{2}; \end{aligned} \tag{5}$$

now let $\varphi = 1 - \frac{k}{2}, \varepsilon_t = \frac{\xi(t)}{2}$ and $\sigma = \frac{\nu}{2}$, then

$$x(t+dt) = \varphi x(t) + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{6}$$

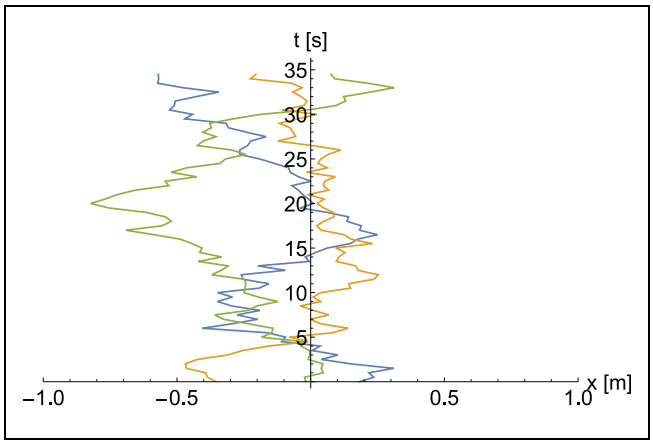

Figure 10. Selected trajectories generated by the first-order autoregressive model.

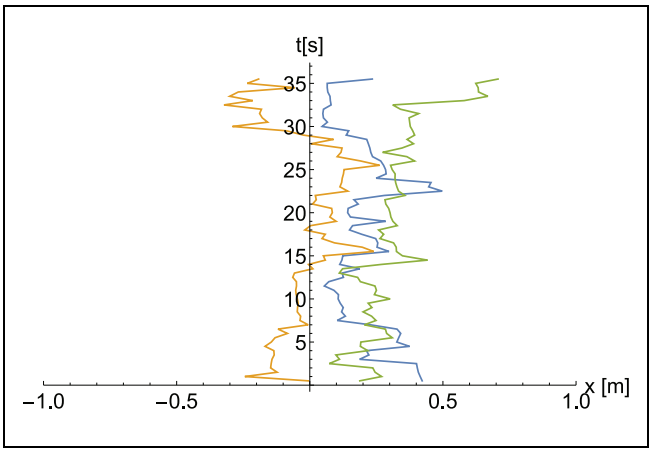

Figure 11. Selected trajectories generated by the mixture-distributed *doblog* model.

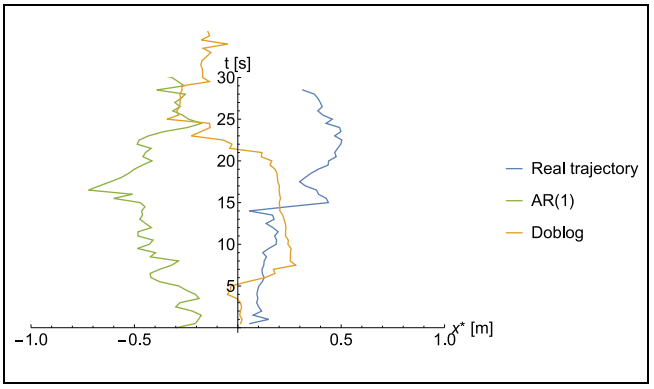

Figure 12. Comparison of sample real and modeled trajectories. The first-order autoregressive trajectory zigzags excessively, while the real and *doblog* trajectories alternate relatively calm segments with sudden corrections.

which fits the definition of an AR(1) model where the constant term is zero. This allows us to determine *a priori* using algebraic methods the stationary distribution of the positions, which will be normal with mean zero and variance:

$$V[x(t)] = \frac{\sigma^2}{1-\varphi^2} = \frac{\nu^2}{(4-k)k} \tag{7}$$

If $x(t) \sim \mathcal{N}(0, \frac{\nu^2}{(4-k)k})$ and $\xi(t) \sim \mathcal{N}(0, \nu^2)$, then by Equation 4 and the properties of the normal distribution,

$$v_x(t) \sim \mathcal{N}\left(0, \frac{4\nu^2}{4-k}\right) \tag{8}$$

From Equation 7 and Equation 8, and given the standard deviations for lateral position and lateral velocity, values for the parameters can be calculated. Using the corrected values for data set I-80 from Tables 2 and 3, gives the following system of equations:

$$\begin{aligned} \frac{\nu}{\sqrt{(4-k)k}} &= 0.354 \\ \frac{2\nu}{\sqrt{4-k}} &= 0.15 \end{aligned} \tag{9}$$

whose positive solutions are $k \approx 0.0449$ and $\nu \approx 0.149$.

The trajectories resulting from this model exhibit the standard deviations for lateral position and velocity it is desired to reproduce. However, because of the model's simplicity there is little ability to control the other characteristics. The mean and standard deviation of log speed can be obtained algebraically and are approximately -1.099 and 0.4824, respectively. The former value depends linearly on the standard deviation of velocity while the latter value is invariant with respect to the parameters.

Examples of trajectories generated by this model are plotted in Figure 10. As can be seen, they appear to oscillate more abruptly than the actual trajectories in Figure 9 that alternate relatively calm periods with sudden corrections. This is consistent with the normal distribution of the modeled velocities, which implies a lower frequency of very low or zero velocities and very high velocities than the true distribution. As a result, it indicates a frequency of intermediate velocities that is clearly greater than the real case (see also Figure 13 for the corresponding distribution of lateral log speed). This suggests that a more complex model with heavier tails for the distribution of velocities is needed so that more of the characteristics observed in the previous section can be reproduced.

Second-Order Model with Mixture-Distributed Random Term (doblog)

A simple way of achieving the above-described combination of low and high accelerations is to combine two distributions, one whose standard deviation is high and the

Table 4. Linear Regressions of Moving Averages for Lateral Acceleration with Respect to Distance to Lane Center and Lateral Velocity

	Original				Corrected	
	x vs a_x		v_x vs a_x		x vs a_x	v_x vs a_x
Data set	Slope	Intercept	Slope	Intercept	Slope	Slope
Prototype	−0.26	−0.07	−1.53	0.01	−0.70	−4.33
I-80	−0.24	−0.06	−2.03	−0.01	−0.29	−2.73
US-101	−0.23	≈0.00	−1.85	≈0.00	−0.37	−2.32

other for which it is low. With this approach in mind, the following model is formulated:

$$\begin{aligned} a_x(t) = & -k_1 x(t) - k_2 v_x(t) + \lambda(t), \\ & \lambda(t) \sim p \times \mathcal{L}(0, s) + (1-p) \times \mathcal{L}(0, u), \end{aligned} \tag{9}$$

In other words, the random part has a probability p of following one logistic distribution with shape parameter s, and a probability $1-p$ of following the other logistic distribution with shape parameter u. This model is called *doblog*, an abbreviation for its double nature from two different logistic distributions.

To calibrate *doblog*, the NSGA-II algorithm (*38*) for multi-objective optimization is used, with objective functions that fit the model-derived results to the direct-data results for each of the four metrics. All four objectives can be satisfied. The results of the parameter calibration are as follows:

$$k_1 = 0.07381, k_2 = 2.4785,$$

$$p = 0.4824, s = 0.2308, u = 0.01316$$

Selected trajectories generated by the model are shown in Figure 11. They reveal a greater frequency of low accelerations than was the case with the previous model, which is more consistent with the true trajectories in Figure 9. In other words, a pattern of alternating relatively calm periods and sudden corrections more alike to the real trajectories is observed.

Finally, Figure 12 compares sample real and modeled trajectories. The AR(1) trajectory zigzags excessively, while the real and *doblog* trajectories alternate relatively calm segments with sudden corrections. The histogram in Figure 13, in turn, compares the log speed frequencies for both proposed models—and an additional second-order one without the mixture-distributed term, excluded here for space reasons—with those of the real (corrected) I-80 data shown in Figure 5.

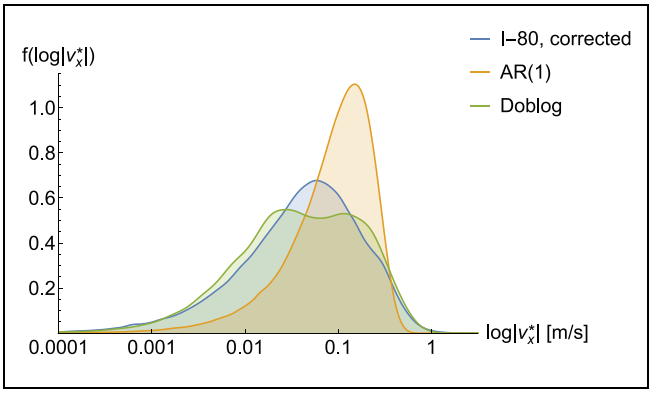

Figure 13. Smoothed histogram of log speed for proposed models and corrected I-80 data set.

Conclusion

An analysis was conducted of the lateral dimension of vehicle behavior in highway traffic. The study focused on vehicles observed to be in the same lane 5 s before and after the instant of observation. The data consisted of high-resolution vehicle trajectories contained in three data sets constructed by the NGSIM program of the U.S. Federal Highway Administration.

Evidence was found in these trajectories that is consistent with perspective bias caused by the location of the cameras that filmed them. A simple solution was proposed for mitigating the problem that comes, however, at the cost of losing information on other possible biases such as an unconscious driver preference for a particular side of the lane.

A series of metrics were proposed to characterize lane discipline and its effects on lateral position choice. These metrics are (i) SDLP, also used in other areas of transport engineering; (ii) the mean of lateral position; (iii) standard deviation of lateral velocity; (iv) the mean and standard deviation of lateral log speed (i.e., the logarithm of the absolute value of velocity); (v) and the average relation between lateral position, lateral velocity, and lateral acceleration. In the case of SDLP and lateral velocity, their behavior as a function of longitudinal velocity was also studied. The proposed metrics should be applicable to any future vehicle traffic data set that records lateral and longitudinal positions over time.

Before the lateral position data were corrected, it was found that in two of the data sets the mean values were more than 20 cm to the left of the lane center. Although

part of this deviation seems not to be real but rather the effect of camera perspective bias, the existence of some other concurrent cause such as drivers' perception bias as to their true position and that of the lane center must also be considered given that the steering wheel is located to one side of the vehicle.

It was also discovered that SDLP consistently increased with longitudinal velocity, at least up to some critical level. The standard deviation of lateral velocity, on the other hand, increased almost totally linearly with longitudinal velocity.

Two models were proposed with differing levels of complexity and ability to reproduce certain characteristics observed in the data. The modeling focused exclusively on the lateral dimension, ignoring the effects of other factors such as longitudinal velocity. The proposed models were (i) an AR(1) model with two parameters that reproduces the standard deviations of lateral position and lateral velocity but not the peculiarities of the latter's distribution; and (ii) a second-order model with a random term that follows a mixture distribution and has five parameters, substantially improving the reproduction of the lateral velocities distribution.

Simulations with the *doblog* model resulted in calibrated values for the k_1 and k_2 parameters of approximately 0.07 and 2.5, respectively, to match the I-80 dataset lane discipline measurements, not too far apart from those obtained from linear regression, at 0.29 and 2.73 (per Table 4). This suggests that the correct calibration of lane force in the 2D- μ TM would have values in the same general vicinity. Naturally, since both methods yield different results, with non-negligible implications in modeled lane discipline, further research is needed to understand why there are such discrepancies or which of the two methods is the most accurate. A possible explanation could lie in (i) the assumption that lane force is linear or (ii) the effect of modeling all other factors (including lateral interaction with other vehicles and perception biases) as a random term independent from position, which could be inducing some distortion in the model.

The calibration results are not compatible with this lane force being used in changing lanes, which would take more than 40 s under the definition of lane change specified in Delpiano (*26*), meaning that, contrary to the assumptions made at that time, either lane-changing needs to be considered a separate process from lane keeping or the lane force needs to be nonlinear. Further research is needed to explore these alternatives.

The idea behind the use of a mixture distribution is consistent with a simple notion that deserves some comment here and would no doubt constitute an interesting topic for future research. The basic idea is that lateral accelerations have diverse causes and thus their characteristics may reasonably be suspected to vary accordingly. These causes include: (i) in the case of very small accelerations, inaccuracies in driver perception or vehicle alignment; (ii) driver corrections of lateral position to stay centered; (iii) and external stimuli such as road-surface imperfections, overtaking by vehicles that are particularly wide or off lane center, or the approach of nearby vehicles in neighboring lanes. Note, however, that, as can be seen in Figure 13, the *doblog* model still does not perfectly fit the data, but with the addition of further distributions, the goodness-of-fit would likely be improved. Incorporating the effects of longitudinal velocity into the models remains a task for future research.

Finally, with regard to driverless vehicle technology, safety risk perception and initial trust have been proposed as important factors in its adoption (*39*). According to Bellem et al. (*3*), safety risk perception depends in part on user comfort, which is determined in turn by the way AVs are operated—including the lateral dimension. Making AVs have a more naturalistic lane discipline by letting them move away from the lane's center when appropriate could contribute to a smoother adoption. As the authors note, "[T]he preferred driving style [of an AV] does not necessarily correspond to how a human driver would drive the vehicle manually" (*3*). Yet failing to take into account what is of value in the human style of driving may be counterproductive from the viewpoint of traffic efficiency (*40*) as well as traffic safety (*41*). In turn, it is a common concern that too strictly centered AVs could cause harm to infrastructure by concentrating the traffic load at specific points. AVs with naturalistic lane discipline could therefore be beneficial for roads by distributing the load.

Author Contributions

The author confirms sole responsibility for the following: study conception and design, data processing, analysis and interpretation of results, and manuscript preparation.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The author acknowledges support from FONDECYT through grant Proyecto Fondecyt Iniciación No. 11201137.

ORCID iD

Rafael Delpiano https://orcid.org/0000-0002-3366-5098.

References

1. Moridpour, S., M. Sarvi, and G. Rose. Modeling the Lane-Changing Execution of Multiclass Vehicles Under Heavy Traffic Conditions. *Transportation Research Record: Journal of the Transportation Research Board*, 2010. 2161: 11–19.
2. Gunay, B. Car Following Theory with Lateral Discomfort. *Transportation Research Part B: Methodological*, Vol. 41, No. 7, 2007, pp. 722–735. https://doi.org/10.1016/j.trb.2007.02.002.
3. Bellem, H., B. Thiel, M. Schrauf, and J. F. Krems. Comfort in Automated Driving: An Analysis of Preferences for Different Automated Driving Styles and their Dependence on Personality Traits. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 55, 2018, pp. 90–100. https://doi.org/10.1016/j.trf.2018.02.036.
4. Delpiano, R., J. C. Herrera, J. Laval, and J. E. Coeymans. A Two-Dimensional Car-Following Model for Two-Dimensional Traffic Flow Problems. *Transportation Research Part C: Emerging Technologies*, Vol. 114, 2020, pp. 504–516. https://doi.org/10.1016/j.trc.2020.02.025.
5. Delpiano, R., J. C. Herrera, and J. A. Laval. Lateral Interactions between Vehicles Help Explain Capacity Drop. Presented at 97th Annual Meeting of the Transportation Research Board, Washington, D.C., 2018.
6. Verster, J. C., and T. Roth. Standard Operation Procedures for Conducting the On-the-Road Driving Test, and Measurement of the Standard Deviation of Lateral Position (SDLP). *International Journal of General Medicine*, Vol. 4, 2011, pp. 359–371. https://doi.org/10.2147/IJGM.S19639.
7. Alm, H., and L. Nilsson. The Effects of a Mobile Telephone Task on Driver Behaviour in a Car Following Situation. *Accident Analysis & Prevention*, Vol. 27, No. 5, 1995, pp. 707–715. https://doi.org/10.1016/0001-4575(95)00026-V.
8. Lindheimer, T. E., K. Fitzpatrick, R. Avelar, and J. D. Miles. Effect of Geometric Factors on Lateral Position of Vehicles in Freeway Buffer-Separated Managed Lanes. *Transportation Research Record: Journal of the Transportation Research Board*, 2017. 2616: 10–18.
9. Rosey, F., and J.-M. Auberlet. Trajectory Variability: Road Geometry Difficulty Indicator. *Safety Science*, Vol. 50, No. 9, 2012, pp. 1818–1828. https://doi.org/10.1016/j.ssci.2012.04.003.
10. Louwerens, J. W., A. B. M. Gloerich, G. De Vries, K. A. Brookhuis, and J. F. O'Hanlon. The Relationship between Drivers' Blood Alcohol Concentration (BAC) and Actual Driving Performance during High Speed Travel. *Alcohol Drugs Traffic Safety*, Vol. 86, 1987, pp. 183–186.
11. Case, H. W., S. F. Hulbert, G. E. Mount, and R. Brenner. Effect of a Roadside Structure on the Lateral Placement of Motor Vehicles. *Highway Research Board Proceedings*, Vol. 32, 1953, pp. 364–370.
12. Avelar, R., K. Fitzpatrick, K. Dixon, and T. Lindheimer. The Influence of General Purpose Lane Traffic on Managed Lane Speeds: An Operational Study in Houston, Texas. *Transportation Research Procedia*, Vol. 15, 2016, pp. 548–560. https://doi.org/10.1016/j.trpro.2016.06.046.
13. Taragin, A. Driver Behavior as Affected by Objects on Highway Shoulders. *Highway Research Board Proceedings*, Vol. 34, 1955, pp. 453–472.
14. May, A. D. Friction Concept of Traffic Flow. *Highway Research Board Proceedings*, Vol. 38, 1959, pp. 493–510.
15. Salini, S., S. George, and R. Ashalatha. Effect of Side Frictions on Traffic Characteristics of Urban Arterials. *Transportation Research Procedia*, Vol. 17, 2016, pp. 636–643. https://doi.org/10.1016/j.trpro.2016.11.118.
16. Delpiano, R., J. C. Herrera M, and J. E. Coeymans A. Characteristics of Lateral Vehicle Interaction. *Transportmetrica A: Transport Science*, Vol. 11, No. 7, 2015, pp. 636–647. https://doi.org/10.1080/23249935.2015.1059377.
17. Budhkar, A. K., and A. K. Maurya. Characteristics of Lateral Vehicular Interactions in Heterogeneous Traffic with Weak Lane Discipline. *Journal of Modern Transportation*, Vol. 25, 2017, pp. 74–89. https://doi.org/10.1007/s40534-017-0130-1.
18. California Department of Motor Vehicles. Report of Traffic Collision Involving an Autonomous Vehicle (OL 316), 2018. https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/autonomousveh_ol316+.
19. Jin, S., D. Wang, P. Tao, and P. Li. Non-Lane-Based Full Velocity Difference Car Following Model. *Physica A: Statistical Mechanics and its Applications*, Vol. 389, No. 21, 2010, pp. 4654–4662. https://doi.org/16/j.physa.2010.06.014.
20. Jin, S., D.-H. Wang, and X.-R. Yang. Non-Lane-Based Car-Following Model with Visual Angle Information. *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2249: 7–14.
21. Maurya, A. Comprehensive Approach for Modeling of Traffic Streams with No Lane Discipline. *Proc., 2nd International Conference on Models and Technologies for Intelligent Transportation Systems*, Leuven, Belgium, 2011.
22. Schönauer, R., M. Stubenschrott, W. Huang, C. Rudloff, and M. Fellendorf. Modeling Concepts for Mixed Traffic. *Transportation Research Record: Journal of the Transportation Research Board*, 2012. 2316: 114–121.
23. Kanagaraj, V., and M. Treiber. Self-Driven Particle Model for Mixed Traffic and Other Disordered Flows. *Physica A: Statistical Mechanics and its Applications*, Vol. 509, 2018, pp. 1–11. https://doi.org/10.1016/j.physa.2018.05.086.
24. Jin, W.-L. A Kinematic Wave Theory of Lane-Changing Traffic Flow. *Transportation Research Part B: Methodological*, Vol. 44, No. 8-9, 2010, pp. 1001–1021. https://doi.org/10.1016/j.trb.2009.12.014.
25. Yang, D., L. Zhu, B. Ran, Y. Pu, and P. Hui. Modeling and Analysis of the Lane-Changing Execution in Longitudinal Direction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. PP, No. 99, 2016, pp. 1–9. https://doi.org/10.1109/TITS.2016.2542109.
26. Delpiano, R. *Modelo Microscópico de Tráfico en Dos Dimensiones Basado en Fuerzas Sociales*. PhD thesis. Pontificia Universidad Católica de Chile., Santiago de Chile, 2015.
27. Helbing, D., and P. Molnár. Social Force Model for Pedestrian Dynamics. *Physical Review E*, Vol. 51, No. 5, 1995, pp. 4282–4286. https://doi.org/10.1103/PhysRevE.51.4282.

28. Helbing, D., and B. Tilch. Generalized Force Model of Traffic Dynamics. *Physical Review E*, Vol. 58, 1998, pp. 133–138. https://doi.org/10.1103/PhysRevE.58.133.
29. Federal Highway Administration. Prototype Dataset. 2004. http://ngsim-community.org/index.php?option=com_docman&task=cat_view&gid=33&Itemid=34.
30. Federal Highway Administration. *Interstate 80 Freeway Dataset*. FHWA-HRT-06-137. Federal Highway Administration, Washington, D.C., 2006.
31. Federal Highway Administration. *US Highway 101 Dataset*. FHWA-HRT-07-030. Federal Highway Administration, Washington, D.C., 2007.
32. Punzo, V., M. T. Borzacchiello, and B. Ciuffo. On the Assessment of Vehicle Trajectory Data Accuracy and Application to the Next Generation SIMulation (NGSIM) Program Data. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 6, 2011, pp. 1243–1262. https://doi.org/10.1016/j.trc.2010.12.007.
33. He, Z. Research Based on High-Fidelity NGSIM Vehicle Trajectory Datasets: A Review, 2017, available at: https://www.researchgate.net/publication/319256552_Research_based_on_high-fidelity_NGSIM_vehicle_trajectory_datasets_A_review?channel=doi&linkId=5cd3fe22299bf14d95820ba6&showFulltext=true
34. Coifman, B., and L. Li. A Critical Evaluation of the Next Generation Simulation (NGSIM) Vehicle Trajectory Dataset. *Transportation Research Part B: Methodological*, Vol. 105, 2017, pp. 362–377. https://doi.org/10.1016/j.trb.2017.09.018.
35. Gunay, B., and D. Woodward. Lateral Position of Traffic Negotiating Horizontal Bends. *Proceedings of the Institution of Civil Engineers - Transport*, Vol. 160, No. 1, 2007, pp. 1–11. https://doi.org/10.1680/tran.2007.160.1.1.
36. Shumway, R. H., and D. S. Stoffer. *Time Series Analysis and its Applications: With R Examples*. 3rd ed. Springer, New York, 2010.
37. Uhlenbeck, G. E., and L. S. Ornstein. On the Theory of the Brownian Motion. *Physical Review*, Vol. 36, No. 5, 1930, pp. 823–841. https://doi.org/10.1103/PhysRev.36.823.
38. Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2, 2002, pp. 182–197. https://doi.org/10.1109/4235.996017.
39. Zhang, T., D. Tao, X. Qu, X. Zhang, R. Lin, and W. Zhang. The Roles of Initial Trust and Perceived Risk in Public's Acceptance of Automated Vehicles. *Transportation Research Part C: Emerging Technologies*, Vol. 98, 2019, pp. 207–220. https://doi.org/10.1016/j.trc.2018.11.018.
40. Anthony, S. E. The Trollable Self-Driving Car. *Slate*, 2016. http://www.slate.com/articles/technology/future_tense/2016/03/google_self_driving_cars_lack_a_human_s_intuition_for_what_other_drivers.html.
41. Reid, C. Nissan Driverless Car Guilty of "Close Pass" Overtake of UK Cyclist, 2017. https://www.bikebiz.com/business/nissan-driverless-car-guilty-of-close-pass-overtake-of-uk-cyclist.