A Critical Evaluation of the Next Generation Simulation (NGSIM) Vehicle Trajectory Dataset

Benjamin Coifman, PhD

Associate Professor
The Ohio State University
Joint appointment with the Department of Civil, Environmental, and Geodetic Engineering, and the Department of Electrical and Computer Engineering

Hitchcock Hall 470
2070 Neil Ave, Columbus, OH 43210
Phone: (614) 292-4282
E-mail: Coifman.1@OSU.edu

Lizhe Li

PhD Candidate
The Ohio State University
Department of Electrical and Computer Engineering

Abstract

A clear understanding of car following behavior and microscopic relationships is critical for advancing traffic flow theory. Without empirical microscopic data, plausible but incorrect hypotheses perpetuate in the vacuum. The Next Generation Simulation (NGSIM) project was undertaken to collect such data and the NGSIM data set has become the de facto standard, underlying the vast majority of empirically based advances of the past decade. But there has been a growing minority of researchers who have found unrealistic relationships in the NGSIM data. To date, the critical findings have almost exclusively come from the existing NGSIM database itself. Unfortunately, as this paper shows, the NGSIM errors are beyond anything that could be corrected strictly through cleaning or interpolation of the reported NGSIM data.

This paper takes the deepest evaluation yet of the NGSIM data. This research manually re-extracts the vehicle trajectories from a portion of the original NGSIM video to explicitly quantify NGSIM errors, e.g., piecewise constant speeds punctuated by brief periods of large acceleration exhibited by the NGSIM data were not evident in the newly extracted trajectories. This point is particularly troublesome for applications that rely on acceleration, e.g., most car following models. The magnitude of errors exhibit a dependency on speed, location and vehicle length. Examples are shown where a real vehicle stopped but the NGSIM trajectory does not and then overruns the location of the real leader. Needless to say, the re-extracted trajectories showed much cleaner speed-spacing relationships than the corresponding raw NGSIM trajectories. Finally, this work tracked the original NGSIM vehicles seen in one camera and added another 236 vehicles (11%) visible before/after the period of NGSIM tracking. As of publication, the manually re-extracted data from this paper will be released to the research community.

Keywords

Traffic flow theory, highway traffic, microscopic models, congested traffic, NGSIM, empirical data

Highlights

- NGSIM has become the de facto empirical microscopic traffic data set
- This study manually re-extracts vehicle trajectories from NGSIM video
- The raw NGSIM data exhibit trends not evident in the re-extracted trajectories
- The magnitude of NGSIM errors depends on speed, location and vehicle length
- As of publication the manually re-extracted data will be publically distributed

1. Introduction

Understanding microscopic traffic flow and car following behavior is critically important for advancing traffic flow theory, which in turn has broad reaching impacts on many aspects of surface transportation including control, management, design, and planning. To this end there is a great need for empirical microscopic traffic data to develop, validate and refine theories and simulations. While there have been numerous efforts to collect empirical microscopic traffic data over the years the fact remains that collecting such data is quite demanding, either requiring very sophisticated automated processing or labor intensive data reduction. So advances in traffic flow theory remain limited by the quantity and quality of empirical traffic data. Without these empirical microscopic data plausible but incorrect hypotheses perpetuate in the vacuum. By the early 2000's these needs were widely recognized and the Next Generation Simulation (NGSIM) project was undertaken to collect such data (Kovvali et al., 2007). To date the NGSIM data is the largest set of empirical microscopic traffic data available to the research community. The NGSIM data have become a de facto standard in the traffic flow community and over the past decade the vast majority of empirically based microscopic advances depend heavily on the NGSIM data.

There are four NGSIM data sets, two from freeways (I-80 and US-101) and two from arterial corridors (Lankershim and Peachtree). In each case the data set includes complete vehicle trajectories for all vehicles over the freeway or arterial from entrance to exit. While the NGSIM data sets are the largest available, they are still relatively small with each set consisting of a few thousand vehicles observed over a short distance (roughly 1/3 mile per set) and duration (under an hour per set).

Although the NGSIM data are widely used, since the release of the data sets there has been a growing minority of researchers who have found unrealistic relationships in the NGSIM data and now question the accuracy of the NGSIM trajectories, e.g., Thiemann et al. (2008), Hamdar and Mahmassani (2008), and Duret et al. (2008) found noise in the instantaneous position, speed, and/or acceleration data so they resorted to low pass filtering to smooth the data. Punzo et al. (2011) undertook an exhaustive search of various errors in the NGSIM data and their findings led to Montanino and Punzo (2015), which is one of the most thorough efforts to address NGSIM data quality. They studied the first 15 min of the I-80 data set to identify periods of unrealistic behavior in the trajectories. The work then used four steps to clean the NGSIM data: (1) recalculate the vehicle position over short stretches to eliminate accelerations with infeasible magnitude. (2) Apply a low pass filter to remove high frequency noise from the vehicle trajectories. (3) Interpolate conditions within a trajectory and across successive trajectories to eliminate physically impossible kinematics. (4) Then finally, apply a low pass filter one more time. While Montanino and Punzo's work is laudable, their data cleaning relies strictly upon the NGSIM vehicle trajectory database. Unfortunately, as this paper will show, the NGSIM errors are beyond anything that could be corrected strictly through such cleaning and interpolation of the reported NGSIM data.

The remainder of the paper is as follows. Section 2 examines numerous chronic errors in the NGSIM data, starting with a brief review of self-evident errors. Then to quantify the errors the section uses the original NGSIM video to re-extract the vehicle trajectories over a portion of the I-80 data set to delve deeper and explicitly quantify the associated errors. Section 3 presents the results of the study. The paper closes with a discussion and conclusions in Section 4.

2. Analysis

This section examines numerous chronic errors in the NGSIM data, with a primary focus on a portion of the I-80 data set. Section 2.1 reviews errors that are readily apparent in the NGSIM data without any external validation. Given the fact that there are errors so large in the NGSIM data set that they become self-evident, Section 2.2 seeks to quantify the errors even when they are not self-evident. To this end, after briefly reviewing the limitations of conventional image processing based vehicle tracking we manually re-extract vehicle trajectories from a portion of the video that was originally used by the NGSIM project to extract the vehicle trajectories in the first place. The raw NGSIM data are then compared to the newly re-extracted data on a number of dimensions of assessment.

2.1. We Hold These Errors to be Self-Evident

There are many chronic errors in the NGSIM trajectory data that have been largely overlooked by the research community. For example, the NGSIM trajectories track the front bumper of the vehicles. One can readily calculate the coordinates of the rear bumper by subtracting the vehicle length from the front bumper position. Fig. 1 shows a 40 s example of data from lane 3 in the 0400-0415 period of the I-80 data set. After accounting for the rear bumper locations there are two collisions evident: trajectory 1444 overruns trajectory 1441 around 490 s and trajectory 1497 overruns trajectory 1486 just before 500 s; and a near miss between trajectories 1456 and 1444 at 485 s. Across the entire I-80 data more than 13% of the trajectories overrun their leader like this. Clearly there were not 747 vehicular collisions during the 45 minutes of study. These large errors alone should call in to question the quality of the NGSIM trajectory data.

Thiemann et al. (2008) identified the collision problem depicted in Fig. 1 and Montanino and Punzo (2015) designed their processing to specifically eliminate these events from their cleaned vehicle trajectory data. Montanino and Punzo also observed an exceptional number of samples in the raw data that exhibited acceleration with magnitude on the order of 10 ft/s^2. Recognizing that such large acceleration is uncommon, their processing was designed to smooth the data in such a way to generally reduce the magnitude of acceleration. To illustrate this problem in the raw data, the solid curves in Fig. 2 show portions of the acceleration time series for four successive vehicles from Fig. 1 and these are typical of the trajectories from the raw I-80 NGSIM data set. The peaks in these curves correspond to the periods of high magnitude acceleration noted in Montanino and Punzo. In fact, the peaks in the raw NGSIM acceleration data are actually truncated from what they should be. If one takes the difference of the raw speed in the NGSIM I-80 trajectory data the peaks can actually be much larger than the NGSIM reported acceleration, as show by the dashed curves in Fig. 2. Briefly reviewing all four of the NGSIM data sets, each one has between 8% and 15% of samples with acceleration over 10 ft/s^2, I-80 and US-101 have the magnitude of acceleration truncated at 11.2 ft/s^2, Peachtree truncated at 12.27 ft/s^2, and Lankershim is unbounded, with a few samples that have a recorded acceleration up to 2000 ft/s^2.

The final self-evident problem is subtle, but is of great concern for many applications of the data. Observe the fact that the typical acceleration time series in Fig. 2 shows zero acceleration most of the time. While at first blush this feature may seem innocuous, the long span of zero acceleration mean that the trajectories exhibit piecewise constant speed. Fig. 3A-D show the raw time series speed corresponding to the four successive trajectories from Fig. 2 and the piecewise constant speeds are readily apparent. All four of the vehicle trajectories show a

constant, non-zero speed at or below 5 ft/s for at least 5 s. While low speeds are common in freeway queues, from our work with probe vehicles we know that very few drivers ever maintain a constant speed below 10 ft/s except when stopped. In contrast to the NGSIM time series speeds, Fig. 3E-F show typical time series speed from a probe vehicle in similar traffic conditions (from the publicly available collection of vehicle trajectories in Coifman et al., 2016); the smoothly varying curves from Coifman et al. show little resemblance to the stair-stepped curves from the raw NGSIM.

These low speed, constant, non-zero speeds in the raw NGSIM data persist in the cleaned data from Montanino and Punzo (2015), e.g., see Figure 3g in their paper. Continuing our brief review of all four of the NGSIM data sets, each one exhibits time series acceleration curves that are consistent with the piecewise constant speeds found in this study of the I-80 data. Suggesting that the general trends found in Fig. 2 and Fig. 3 are typical of the NGSIM data extraction in all of the data sets.

The piecewise constant speeds punctuated by brief periods of large acceleration as found in the NGSIM vehicle trajectory data sets (and retained in the cleaning by Montanino and Punzo) are unrealistic of actual driving behavior. This point is particularly troublesome for any application that relies on acceleration either as an input or an output, e.g., most car following models.

2.2. Manually Re-Extracting NGSIM Vehicle Trajectories

Given the self-evident problems within the NGSIM data sets discussed in Section 2.1, this section seeks to delve deeper and explicitly quantify the associated errors. In the process, this work also provides more accurate vehicle trajectories. The NGSIM project extracted the vehicle trajectories from orthorectified video recorded from tall buildings. The basic concept of extracting vehicle trajectories from orthorectified imagery recorded from a high vantage point is nothing new. It was fairly common in the 1960's and 70's, relying on labor-intensive data reduction techniques, e.g., Forbes and Simpson, M. (1968), Treiterer and Myers (1974). Turner-Fairbank Highway Research Center collected data at one frame per second from 18 locations in 1983 and used microcomputers to expedite the data reduction process (Smith, 1985; Smith and Mark, 1985). Unlike the earlier works, NGSIM relied on a primarily automated process to extract the vehicle trajectories.

Video image processing remains more of an art than a science. While automated tools can yield a good first cut, achieving the precision necessary for extracting sufficiently accurate empirical vehicle trajectories to develop traffic flow models remains a very challenging prospect. Without extensive human intervention, even the best image processing tools cannot overcome the difficulties of: projection errors, occlusions, shadows, the non-rectilinear shapes of real vehicles, vehicles that have colors similar to the pavement, etc. to the degree necessary for advancing traffic flow theory.

While little used by the research community, NGSIM also includes validation video that superimposes the extracted vehicle tracks on top of the original video. In the 1990's Coifman (1997) used orthorectified imagery to manually generate ground truth data from 13 shockwaves to validate the Roadwatch video image processing study (Coifman et al., 1998). The present work revives the manual data reduction and applies it on a larger scale to re-extract all of the longitudinal vehicle positions from the NGSIM I-80 camera 6 validation video during the 0400-0415 period. Fig. 4A shows a sample frame from the validation video. The NGSIM researchers superimposed a pink box for each vehicle on to the original video used for tracking. The pink

box bounds the given vehicle's coordinates as recorded in the NGSIM data set at that instant and the green number associated with the box is the vehicle's ID in the trajectory data set.[1]

The very first step of the analysis was synchronizing the video with the NGSIM data. In comparing the boxes in the validation video against the trajectory data it turns out that indeed the time stamp in the trajectory database (column 2, *Frame Number* of the original NGSIM data) does indeed correspond to the actual frame of the video. Furthermore, camera 6 spans longitudinally from 1,243 to 1,509 ft (as measured in column 6, *Local Y* of the original NGSIM data) and that each pixel is roughly 0.5 ft along the road.

The NGSIM vehicle tracking appears to have followed conventional practice, with the 2D image plane seen by the camera projected into the 2D ground plane of the roadway with the implicit assumption that all objects are indeed in the ground plane. Obviously 3D vehicles violate this ground plane assumption, resulting in projection errors. These projection errors are like shadows, the higher the feature is off of the ground, the further it projects away from its true ground coordinates. The projection errors also increase as the vehicle moves further from the camera; thus, the projection of the top front of the vehicle should seemingly move faster over ground than the actual vehicle, a point we will return to in the next section.

In the case of NGSIM, they made the challenges even greater; the original NGSIM video was shot in full frame rate high resolution, but before tracking the researchers downsampled the video to 640x480 pixels and 10 fps, with most of the frame falling outside of the freeway. In the case of I-80 camera 6, the study area occupies roughly 160x480 pixels in the raw video. With the lower resolution many distant vehicles are only 12x12 pixels. This low resolution is far too small to use robust feature tracking tools. The low resolution video was then projected into the ground plane leading to further distortion. The projection expands the small number of pixels occupied by the distant vehicles in the raw video to be comparable in size to the same vehicles in the near field, giving false confidence and masking the descritization errors of the original downsampling.

The NGSIM data extraction tracked the front of the vehicles, but for camera 6 the target vehicles are moving away from the camera. So at best only the top of the front of the vehicle was visible, or in the case of box trucks the box can completely occlude the actual front of the vehicle (as is the case for the two semitrailer trucks in Fig. 4A).

Given the fact that camera 6 views the rear of the vehicles, to avoid the large projection errors arising from the front of the vehicles this study manually tracks the bottom rear of the vehicle because that feature is always closest to the ground and should exhibit the least projection error. While much of the image processing literature considers shadows as a confounding factor to be eliminated, this work actually exploits the vehicle shadows in their own lane of travel since the shadows are inherently in the ground plane and do not suffer from projection errors. Of course, the shadows are lost when an intervening object blocks the sun from the vehicle (either a tall truck in the adjacent lane or a structure outside of the freeway), so during these periods we track the lowest observable feature on the vehicle through the shadowed period and align it with the shadow's track on either end of the shadowed period. Finally, Gaussian Kernel smoothing is applied to the position data for each trajectory to reduce the impact of data reduction noise, e.g., being limited to the resolution of discrete pixels.

Fig. 4B-D show a detail of lane 3 (3rd lane from the left) from the validation video for three frames. Both the validation video and NGSIM data set are recorded in 10 frames per

[1] This frame is copied directly from the NGSIM video and some of the colors might not be evident in a print version of this paper. Since we cannot change the original artwork any critical location information shown in this image will be presented in other figures.

second, numbered from the start of the video. As noted above, we found that the data set is time synchronized with the video and in this paper the time stamp in seconds is equal to the frame number divided by 10. We also verified that the bounding boxes in the video correspond to the positions recorded in the raw NGSIM data set.

To maintain consistency between the raw and re-extracted trajectories, when referring to the raw NGSIM trajectories the remainder of this work will use the rear of the raw NGSIM vehicles by subtracting the vehicle length from the longitudinal position (as per Fig. 1). In Fig. 4B the validation bounding boxes are visible for several vehicle trajectories. For clarity, the NGSIM recorded position of the rear of each vehicle is shown to the left of the image with the label "raw" and indexed by the NGSIM vehicle ID. The corresponding manually re-extracted positions from the current work are shown to the right for the same vehicles, preserving the original NGSIM vehicle ID and labeled "new". In Fig. 4B there are only a few small errors in the original raw NGSIM positions, e.g., the NGSIM position for vehicles 1441 and 1456 is ahead of the actual position by a few feet. According to our re-extracted, "new" data, at this frame vehicle 1456 has just come to a stop and the following vehicle 1463 will do the same within 1 s as a stop wave propagates upstream. Fig. 4C shows the same location 150 frames (15 s) later and corresponds to the same instant chosen for the example in Fig. 4A. According to the new data vehicle 1456 has just started moving and the next three vehicles upstream are stopped at this instant, including the immediate follower 1463. Yet the raw NGSIM trajectories recorded for 1456 and 1463 show these vehicles are moving at this instant and as evident in this figure, both of these trajectories have drifted far from the actual vehicle location, e.g., raw 1463 is a full vehicle length ahead of the vehicle's actual position. Fig. 4D shows the instant when vehicle 1463 finally catches up to its trajectory recorded in the raw NGSIM trajectory data. In this snippet 1463 is the only raw trajectory that is within 10 ft of the vehicle's actual location and two of the raw NGSIM trajectories have completely overrun the location of the actual leading vehicle. Trajectory 1478 has pulled a full vehicle length ahead of the vehicle's real position. The next trajectory, 1486, is only half a vehicle length ahead of the vehicle's true location, but it has been overrun by trajectory 1497 (one of the "collisions" that are self-evident in the raw NGSIM trajectories) and in fact at this instant trajectory 1497 is even ahead of the actual location of its lead vehicle, 1486. Similarly, the next trajectory, 1506, has overtaken the actual location of its lead vehicle, 1497, but in this case there is no "collision" of trajectories because raw 1497 is so far ahead of its actual vehicle in this frame.

The progression in the lower half of Fig. 4 showed just three frames. Fig. 5 shows the resulting trajectories (vehicle rear) for 10 successive vehicles in lane 3 over 40 s. The raw NGSIM trajectories are shown with dashed lines and the re-extracted trajectories with solid lines. Each trajectory is labeled with its ID and the three frames used in Fig. 4 are indicated with vertical lines in the time-space plane. The new re-extracted trajectories show that the six successive vehicles from 1456 to 1506 came to a stop within the camera 6 field of view in response to the upstream moving stop wave, while the corresponding raw NGSIM trajectories only show vehicle 1478 stopping. At these slow speeds the piecewise constant speed of the raw NGSIM trajectories are clearly evident, with only infrequent changes in speed.

2.2.1. Calculating speed

Even the re-extracted vehicle positions underlying the trajectories in Fig. 5 exhibit small instantaneous positioning errors (on the order of a foot) due to the relatively low resolution of the NGSIM video. While these errors are far below a perceptible threshold at the scale of Fig. 5, the

errors are amplified when taking the position difference in two successive frames and degrade the quality of conventional speed calculations if they are not addressed. To this end, the speed is calculated using the median difference over multiple time-steps as follows. First, for each trajectory the time series speed is separately calculated using a range of time steps via Equation 1, where i denotes the current frame, n is the time-step, and $x(i+n)$ is the vehicle's position in frame i+n. Thus, $\hat{v}_n(i)$ is the estimated average speed of the vehicle over $2 \cdot n$ frames (n/10 s) bounding the current frame. This work simultaneously evaluates Equation 1 with integer n from 1 to 7, yielding 7 different measures of speed at a given instant. The vehicle's median speed at frame i is then calculated by taking the median of all 7 time steps.

$$\hat{v}_n(i) = \frac{x(i+n) - x(i-n)}{2 \cdot nT} \quad (1)$$

This processing could be viewed as a non-linear low pass filter. It is important to recognize that any speed derived from image processing will require some form of low pass filtering or equivalent smoothing even if the given work does not explicitly call out the filtering. In other words, this need for low pass filtering is common to all image processing based vehicle tracking and is not unique to the current work. Regardless, since a low pass filter combines information across several instances, the windowing will obscure the exact instant that a vehicle comes to a stop or begins to move. To retain precise boundaries between moving and non-moving conditions this work classifies a given vehicle's median speed time series into three conditions: *stopped*, *almost-stopped*, and *moving*. When the median speed is below 0.3 ft/s (0.2 mph) the vehicle is classified as stopped, when the median speed is above 4 ft/s (2.7 mph) the vehicle is classified as moving and any median speed between the two thresholds is classified as almost-stopped. All vehicles have at least one period where speeds are classified as moving. Some vehicles will come to a stop, in which case there will be a period of moving, followed by a period of stopped, followed by another period of moving. Between each moving and stopped period there is a brief transition through a period classified as almost-stopped. This cycle will repeat if the vehicle stops multiple times. For each contiguous period classified as stopped the final speed is set to zero. For each contiguous period classified as moving, a Savitzky-Golay filter is applied to the median speed to smooth out any remaining noise. Finally, in the almost-stopped periods to ensure a smooth transition the speed is interpolated using a cubic spline between the moving speed on one end of the almost-stopped period and the zero stopped speed on the other end of the period.

The solid curves in Fig. 6 show the resulting final speed for the re-extracted trajectories from four successive vehicles in Fig. 5 shown previously in Fig. 2-3. These time series speeds look much more like the probe vehicle data in Fig. 3E-F than they look like the raw NGSIM data in Fig. 3A-D that are reiterated here with dashed curves for reference. While the piecewise constant speeds of the raw NGSIM trajectories roughly follow the average re-extracted speeds, the final re-extracted speeds offer far greater fidelity, e.g., according to the new re-extracted data all four of the vehicles come to a stop while the raw NGSIM data shows only one of these four vehicles coming to a stop.

2.2.2. Calculating acceleration

With the clean speeds, acceleration can be calculated using a simpler approach than what was used to find the speeds. First, for each trajectory the time series acceleration is calculated using a single time step via Equation 2, where n is set to 1 and $v_{revised}(i+1)$ is the final speed for frame i+1. Then the acceleration time series is smoothed according to the median speed

classification for the given period using the same filtering strategy: setting acceleration to zero during each contiguous stopped period, applying a Savitzky-Golay filter during each contiguous moving period, and a cubic spline interpolation across each contiguous almost-stopped period.

$$\hat{a}_1(i) = \frac{v_{revised}(i+1) - v_{revised}(i-1)}{2 \cdot T} \quad (2)$$

The solid curves in Fig. 7 show the resulting acceleration for the four re-extracted trajectories shown in Fig. 6. For reference, the plots in this figure reiterate the raw NGSIM speeds with dashed curves that were shown previously in Fig. 2. In each plot there is very little similarity between the two time series. The new acceleration curve is rarely zero except when stopped and the extreme spikes that characterized the raw NGSIM acceleration are completely absent.

2.2.3. Extracted Data

This process of manually extracting the longitudinal vehicle trajectories followed by calculating time series speed and acceleration was repeated for all lanes and all of the non-motorcycle vehicles visible in the I-80 camera 6 for the 0400-0415 time period. More precisely, while most of the motorcycles were actually tracked in re-extraction, some were not tracked due to the extra challenges involved as discussed below.

In the process of examining every vehicle two further problems with the raw NGSIM data became apparent in which vehicles that are laterally separate will sometimes be assigned to the same lane at the same instant and longitudinal location; thus, in the raw NGSIM data occasionally vehicles in the same lane will seemingly overtake one another the absence of a tracking error. First, as already observed by Montanino and Punzo (2015), motorcycles often did not exhibit lane adherence as they "split lanes," i.e., running between the normal lanes of travel. Since all vehicles were assigned to a discrete lane, during an overtaking these motorcycles will coexist with other vehicles in a given lane and longitudinal location. Second, several raw NGSIM trajectories assigned to lane 6 (the right most lane) were observed to actually be traveling illegally in the right-hand shoulder. According to the NGSIM documentation these vehicles should have been assigned to (virtual) lane 9 to indicate that they were driving on the shoulder, but in the raw NGSIM database they were all assigned to lane 6. As such, the current work has reassigned these shoulder traveling vehicles to lane 9. Consumers of the NGSIM database need to be careful in the vicinity of the motorcycles and the shoulder running vehicles. There is no clear way of identifying when the motorcycles were splitting lanes and when they were traveling within lane, but even when they are within lane there is a good chance that the following driver might exhibit different behavior than they would for a larger vehicle. For users of the original NGSIM database (including the other camera views and time periods) the lane 9 vehicles erroneously assigned to lane 6 might be evident by their lateral position on the roadway.

The readers are cautioned not to rely too heavily upon the lateral position reported in the raw NGSIM database though. While we did not make any corrections for the lateral position beyond assigning vehicles to lane 9, it is readily apparent that the bounding boxes in the NGSIM validation video often did a poor job of laterally tracking the target vehicle, e.g., a close examination of Fig. 4A will reveal that for some vehicles the bounding box followed the left side of the vehicle (including the projection error to the left side of the road due to the viewing angle) and for other vehicles the bounding box followed the right side of the vehicle (with no projection error, but sometimes the bounding box falls further to the right due to inclusion of part of the shadow).

As noted above, the process of manually extracting the longitudinal vehicle trajectories was repeated for all lanes and all of the non-motorcycle vehicles visible in the I-80 camera 6 for the 0400-0415 time period. This set of 2283 vehicles is larger than the 2047 non-motorcycles tracked in NGSIM. In the original NGSIM study they only tracked vehicles that entered the corridor on the mainline in camera 1 or on-ramp in camera 2 after the start time and subsequently exited the corridor in camera 7 before the end time. In other words, they did not track any vehicles that were already between cameras 1 and 7 at the start of the video, nor any of the vehicles that were still between cameras 1 and 7 at the end of the video. The present study proceeded to longitudinally track all 181 of the non-motorcycle vehicles ahead of and 55 behind the NGSIM platoon seen in the camera 6 video; yielding an additional 236 vehicles. For these previously untracked vehicles ahead and behind of the NGSIM platoon, they are sorted by lane with no further lateral refinements.

As of the publication of this paper the entire set of newly extracted data will be made publically available to the research community via Coifman (2017).

3. Results

While it would be impossible to show in this paper the individual results for the over 2000 vehicles in the data set in a format like Fig. 5, a validation video has been posted at Coifman (2017) showing the newly extracted vehicle positions with blue bars aligned with the rear bumper, superimposed on the original NGSIM validation video with their pink bounding boxes used for the original validation. The remainder of this section presents aggregate results for the raw NGSIM data set.

Fig. 8 shows the resulting speed spacing relationship for the four successive vehicles from Fig. 6. The darker points show the raw NGSIM trajectories, characterized by many samples at a given speed before an abrupt jump to a new speed, consistent with the piecewise constant speed from above; whereas the lighter points show the same vehicles over the same time period after re-extracting the trajectories. These four vehicles from the on-going example are typical of the entire data set. Taking all of the non-motorcycles in the raw NGSIM 0400-0415 trajectory set, Fig. 9A shows the probability distribution of the raw NGSIM position minus the manually re-extracted position by lane and all lanes combined together for camera 6. The spread of the peaks reflects the discrepancy between the raw and newly extracted data, e.g., as evident by the vertical differences in Fig. 5. The raw NGSIM positions tend to lag the re-extracted positions, as evident by the modes to the left of the origin in Fig. 9A. If all lanes exhibited the same non-zero mode it would simply indicate a fixed offset between the two coordinate systems, but the mode is more negative in lanes 1 and 6 (the far left and far right lanes, respectively). These two lanes are characterized by higher speeds on average. Fig. 9B shows the mean position error as a function of the re-extracted speed, once more by lane and all lanes combined together. Indeed, at higher speeds the raw NGSIM position falls behind the re-extracted position. Whereas at lower speeds, the raw NGSIM position pulls ahead of the re-extracted positions, consistent with the trends seen in Fig. 4-5.

As noted in Section 2.2, the NGSIM image processing appears to have used conventional projection in to the 2D ground plane of the roadway with the (implicit) assumption that all objects are indeed in the ground plane. The resulting projection errors increase with height of the vehicle and distance away from the camera. Fig. 9C shows the mean position error as a function of longitudinal position along the segment. It turns out that the original NGSIM data extraction

did a poor job at the transition from one camera to the next. These errors occur in all lanes, as evident by the large deviations within the upstream and downstream ends of the camera 6 field of view (i.e., the first and last 50 ft of Fig. 9C). Using the vehicle lengths recorded in the NGSIM data set, Fig. 9D shows the mean position error by vehicle length and in this case the position error increases from the shortest vehicles to vehicles about 40 ft long and then declines slightly up to the longest vehicles. This error is consistent with the original NGSIM automated processing tracking the front of the vehicle since maximum vehicle height roughly increases with vehicle length up to about 40 ft and then levels off as length increases further (Lee and Coifman, 2012). In other words, in the range of camera 6 the raw NGSIM data appears to exhibit biases in position along the road due to projection errors from the camera orientation and to the height of the vehicle as a result of the original NGSIM processing following the front of the targets. For locations upstream of the camera location (NGSIM cameras 1-3) the error may reverse and project upstream if they tracked the top front of the vehicles, or it may go away if they tracked the bottom of the front of the vehicles, similar to how this work tracked the bottom of the vehicle end closest to the camera.

Fig. 10A shows the corresponding probability distribution of the raw NGSIM speed minus the manually re-extracted speed by lane and all lanes combined together. Once more, the spread of the peaks reflects the discrepancy between the raw and newly extracted data, e.g., as evident by the vertical differences in Fig. 6. The raw NGSIM speeds are slightly higher than the re-extracted trajectory speeds in lane 1, but not the other lanes. Fig. 10B shows the mean speed error as a function of the re-extracted speed. The vast majority of speeds in lane 1 were above 40, while the remaining lanes show that as speeds drop below 20 ft/s that the raw NGSIM exhibits increasingly positive errors. While the magnitude of the mean speed error is fairly small, in most cases being under $|2\ ft/s|$, the mean relative error grows to over 100% at the lowest speeds. Such large relative errors in the raw NGSIM data are likely to be problematic for studies that focus on stop and go conditions or the stop waves passing through the segment. Fig. 10C shows the mean speed error as a function of longitudinal position along the segment and the trends are consistent with the errors discussed earlier, with all lanes showing large speed errors at the upstream and downstream ends of the segment. Between 1300 and 1500 ft all lanes show that the speed error slowly becomes more positive the further downstream one goes, consistent with the projection errors growing as a target moves further away from the camera. The error in lane 1 is consistently a little higher than the other lanes, likely because this lane is traveling faster, but lane 1 still exhibits the same general trend as the other lanes. Fig. 10D sorts the speed error by vehicle length and once more the error increases from the shortest vehicles to about 40 ft long vehicles and then declines slightly up to the longest vehicles. Here too, the trend is likely related to vehicle heights increasing the projection errors.

4. Discussion and Conclusions

One of the biggest problems with microscopic traffic flow models is the lack of large quantities of in situ data for development, calibration, and validation. Empirical microscopic traffic data have historically been difficult to collect, and as a result, advances of microscopic traffic flow theory (e.g., car following behavior) were very slow in coming. The Next Generation Simulation (NGSIM) project sought to alleviate the data needs by providing empirical vehicle trajectory data. The NGSIM data sets form the basis for most of the empirically based advances in microscopic traffic models since the data sets were released roughly ten years ago.

Although NGSIM delivered one of the largest microscopic traffic data sets to date, this paper has shown that NGSIM failed to deliver accurate data. The analysis began with a short review of previously known problems that are self-evident in the NGSIM data set: after accounting for vehicle length trajectories often over run their leaders seemingly resulting in "collisions of trajectories", and that the NGSIM acceleration often exhibits unrealistically large magnitudes. Our analysis went on to show another self-evident error, that the NGSIM speeds exhibit unrealistic piecewise constant behavior. It appears as if at low speeds the NGSIM video image processing frequently linearly interpolates the trajectories between two points in space observed many seconds apart. As shown in Fig. 3, this piecewise constant speed of the raw NGSIM data was not evident under similar conditions in an independent empirical microscopic data set (Coifman et al., 2016). The fact that these large errors in the NGSIM speed data could persist without being detected by the research community after detailed, microscopic study by so many researchers is evidence of just how uncommon empirical microscopic traffic data sets really are; even many traffic flow experts lack firsthand experience to spot the inconsistencies. Given the self-evident position, speed, and acceleration errors, the raw NGSIM data sets run the risk of undermining any microscopic model fit to those data.

This work sought to explicitly quantify the errors in the NGSIM data by manually re-extracting all of the longitudinal vehicle positions from the NGSIM I-80 camera 6 validation video during the 0400-0415 period. This manual re-extraction revived some useful old approaches, but was not in itself revolutionary. Rather, the extracted data and validation of the NGSIM data are the important outcomes.

The analysis steps were presented in the context of four successive vehicles that were typical of the camera 6 vehicles. It was shown that several raw NGSIM vehicle trajectories continued moving and pulled ahead of the actual vehicle locations as a stop waved propagated upstream through those vehicles, resulting in *collisions of trajectories* that were observed in the review of self-evident errors. Speeds were then calculated from the re-extracted positions. The re-extracted speeds offer far greater fidelity than those reported in the NGSIM data set, e.g., all four of the subject vehicles actually came to a stop but only one of the corresponding raw NGSIM trajectories did so. Next, accelerations were calculated for the re-extracted trajectories. Comparing the raw NGSIM acceleration against the new re-extracted trajectories there is very little similarity between the two time series. While the raw NGSIM exhibited mostly zero acceleration punctuated by brief periods of large acceleration, the newly extracted acceleration curves are rarely zero except when the vehicle is stopped, furthermore, the large peaks that characterized the raw NGSIM acceleration are completely absent. This error in the raw NGSIM data is particularly troublesome for any application that relies on acceleration either as an input or an output, e.g., most car following models.

The re-extracted trajectories from the present study showed much cleaner speed-spacing relationships than the corresponding raw NGSIM trajectories, e.g., Fig. 8. Taking the entire I-80 camera 6 0400-0415 data set collectively it was shown that the raw NGSIM position and speed data exhibit a dependency on speed, location, and vehicle length.

The readers can judge for themselves by watching our revised validation video,[2] superimposing our re-extracted longitudinal vehicle positions on top of the original NGSIM validation video. In the video the blue lines show the re-extracted longitudinal distance and the pink bounding boxes from the original video show the NGSIM position, similar in style to Fig. 4. It is worth noting that NGSIM only tracked those vehicles that completely passed through all 7

[2] As of publication the revised validation video can be found at Coifman (2017)

cameras in the segment during the filming period. So, all of the vehicles that were already in the segment at the start and all of the vehicles that were still in the segment at the end were not tracked in the NGSIM data set. The present work longitudinally tracked these vehicles while in the camera 6 field of view, adding another 236 trajectories to the data set (an increase of 11% in the number of vehicles).

The present work is not the first attempt to correct the raw NGSIM data. Montanino and Punzo (2015) is probably the most thorough preceding work and they did a laudable effort to clean the NGSIM data, but their cleaning relies strictly upon the NGSIM vehicle trajectory database. Unfortunately, as our paper has shown, the NGSIM errors are beyond anything that could be corrected strictly through such cleaning and interpolation of the reported NGSIM data. While Montanino and Punzo's cleaning eliminated most if not all infeasible events in the raw NGSIM data, like almost all consumers of the raw NGSIM data, they underestimated the full extent of the errors in the original NGSIM tracking data. Even after eliminating the infeasible maneuvers many erroneous but feasible maneuvers remain in their cleaned data. The depth of the errors in the original NGSIM data only become evident from careful and labor intensive comparison against the source video. Fig. 11 compares the raw NGSIM and Montanino and Punzo cleaned data against the manually re-extracted data. Montanino and Punzo improved the position of many vehicles (e.g., vehicles 1456 and 1497 in Fig. 11E); however, many position errors remain in their cleaned data (e.g., vehicles 1463, 1478 and 1486 in Fig. 11E). Fig. 11B shows that vehicle 1486 had a positioning error that drifted from -7 ft to +18 ft in Montanino and Punzo's cleaned data. Fig. 11C-D show that in spite of reducing the acceleration spikes, Montanino and Punzo often preserved the unrealistic but feasible piecewise constant speeds from the raw NGSIM data.

Some of the errors in Montanino and Punzo's cleaned data simply reflect an unavoidable tradeoff between two compromises. They explicitly set out to ensure that their cleaned data were internally consistent, i.e., that the integrated acceleration over time is consistent with the velocity and the integrated velocity over time is consistent with the position. This internal consistency is critical for many car following models. Because the original vehicle positions were measured with error, these errors will be amplified when differencing to calculate speed and acceleration. It appears that Montanino and Punzo chose to minimize the errors in the speed vector. As a result, to preserve internal consistency the recalculated position will be subject to errors (e.g., as per Fig. 11). Given the unavoidable measurement errors from the relatively low resolution NGSIM video it is impossible to achieve internal consistency without inducing errors on at least two of the three time series: position, speed, and acceleration. In contrast, we chose to minimize the errors individually on position, speed, and acceleration; but doing so comes at the expense of no longer maintaining internal consistency. Fortunately, if internal consistency is critical for a given application the user can use the speed from our re-extracted data and then recalculate position and acceleration by summing and differencing the speed, respectively.[3]

Returning to the real problem at hand, there remains a need for additional empirical microscopic data sets, e.g., Coifman et al. (2016) used herein. As of publication, the re-extracted data and validation video from this paper will be released to the research community via

[3] In the final cleaning the position data were manually corrected up to 1ft to better align with the video, which would induce transient speed jumps when differencing the final position data. These final adjustments were not made to the speed data, hence, one should use the speed rather than position to generate internally consistent data. Using the data as recorded, when summing speed to derive position the average absolute error by vehicle ranges from 0 to 0.9 ft, with an average across vehicles of 0.063 ft. Similarly, when summing acceleration to derive speed the average absolute error by vehicle ranges from 0 to 1.25 ft/s, with an average across vehicles of 0.021 ft/s.

Coifman (2017). These data include longitudinal position, speed, and acceleration, indexed by time and the NGSIM vehicle ID if NGSIM tracked the given vehicle or a new unique vehicle ID for the additional vehicles we tracked.

Finally, this paper focused strictly the process of extracting the data within a single camera. It turns out that the original NGSIM data extraction did a poor job at the transition from one camera to the next. These errors are evident in Fig. 9C and Fig. 10C, the position and speed errors show systematic spikes at the upstream and downstream ends of the camera 6 field of view. On-going work seeks to address these errors at the camera boundaries and expand the re-extraction to multiple cameras at the I-80 site.

Acknowledgements

This material is based in part upon work supported in part by the National Science Foundation under Grant No. 1537423. The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. This report does not constitute a standard, specification or regulation.

The authors would like to thank Doug Thornton, Lan Wu, and Gus Fragasse for their input and feedback in the early stages of this exploration, and the anonymous reviewers for their input and feedback in preparing this paper.

References

Coifman, B. (1997) *Time Space Diagrams for Thirteen Shock Waves,* University of California - PATH.

Coifman, B. (2017) Data Sets, http://www.ece.osu.edu/~coifman/documents/

Coifman, B., Beymer, D., McLauchlan, P., Malik, J. (1998) A Real-Time Computer Vision System for Vehicle Tracking and Traffic Surveillance, *Transportation Research: Part C,* vol 6, no 4, pp 271-288.

Coifman, B., Wu, M., Redmill, K., Thornton, D. (2016) "Collecting Ambient Vehicle Trajectories from an Instrumented Probe Vehicle- High Quality Data for Microscopic Traffic Flow Studies" *Transportation Research Part C,* Vol 72, pp 254-271.

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 2088*, p. 188-197.

Forbes, T., Simpson, M. (1968) Driver-And-Vehicle Response In Freeway Deceleration Waves, *Transportation Science*, Operations Research Society of America, Volume 2, Issue 1, pp 77-104.

Hamdar, S., Mahmassani, H. (2008) "Driver Car-Following Behavior: From Discrete Event Process to Continuous Set of Episodes," *Proc. of the 87th Annual Meeting of the Transportation Research Board.*

Kovvali, V., Alexiadis, V., Zhang, L. (2007). "Video-Based Vehicle Trajectory Data Collection," *Proc. of the 86th Annual TRB Meeting*, TRB.

Lee, H., Coifman, B., (2012) "Side-Fire LIDAR Based Vehicle Classification," *Transportation Research Record 2308,* pp 173-183.

Montanino, M., Punzo, V. (2015) Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns, *Transportation Research Part B*, Vol 80, pp 82-106.

Punzo, V., Borzacchiello, M., 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*, Vol 19, No 6, pp 1243-1262.

Smith, S. (1985) *Freeway Data Collection for Studying Vehicle Interactions*, Technical Report FHWA/RD-85/108, FHWA, USDOT

Smith, S., Mark, E. (1985) Creation of data sets to study microscopic traffic flow in freeway bottleneck sections, *Transportation Research Record 1005*, pp 121-128.

Thiemann, C., Treiber, M., Kesting, A. (2008) "Estimating acceleration and Lane-Changing Dynamics from Next Generation Simulation Trajectory Data," *Transportation Research Record 2088*, pp 90-101.

Treiterer, J., Myers, J. (1974) The Hysteresis Phenomenon in Traffic Flow, *Proc., Sixth International Symposium on Transportation and Traffic Theory*, (D. J. Buckley, ed.), pp 13-38.

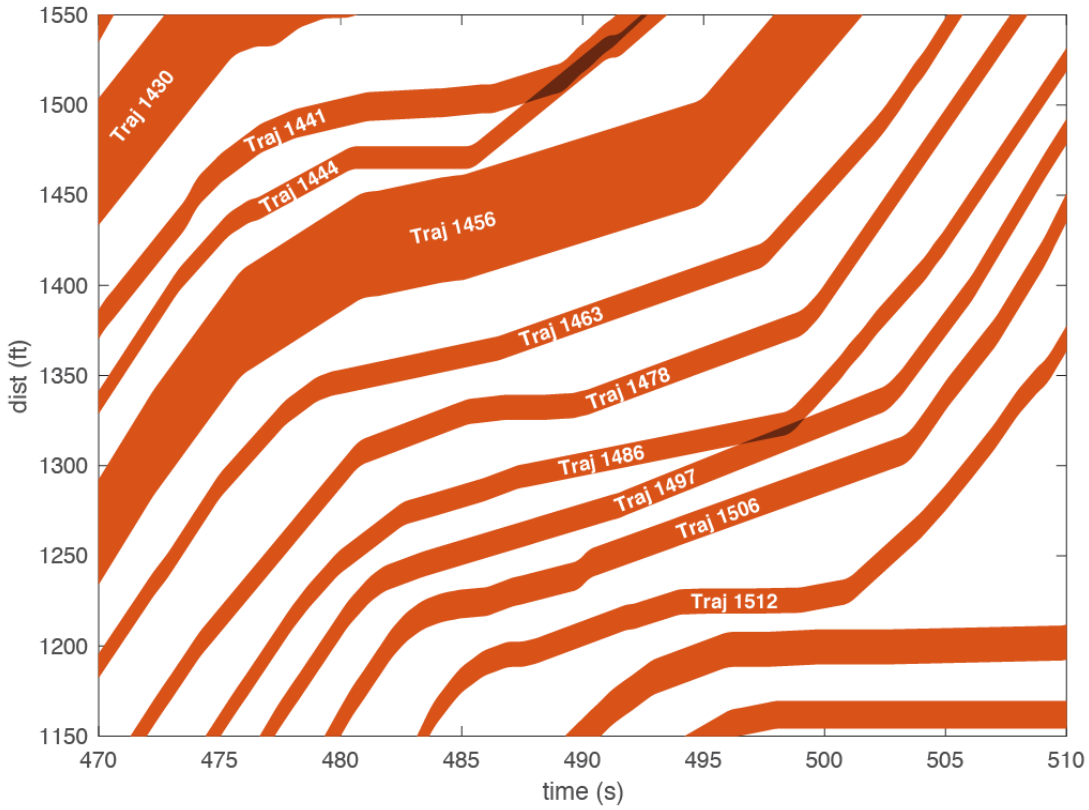

Figure 1, An example of the full vehicle trajectories computed from the NGSIM front bumper location (downstream end) and vehicle length (upstream end) for each target vehicle. There are two "collisions" evident in this figure, shown with a darker color: Trajectory 1444 overruns Trajectory 1441 at about 490 s and Trajectory 1497 overruns Trajectory 1486 at about 498 s.

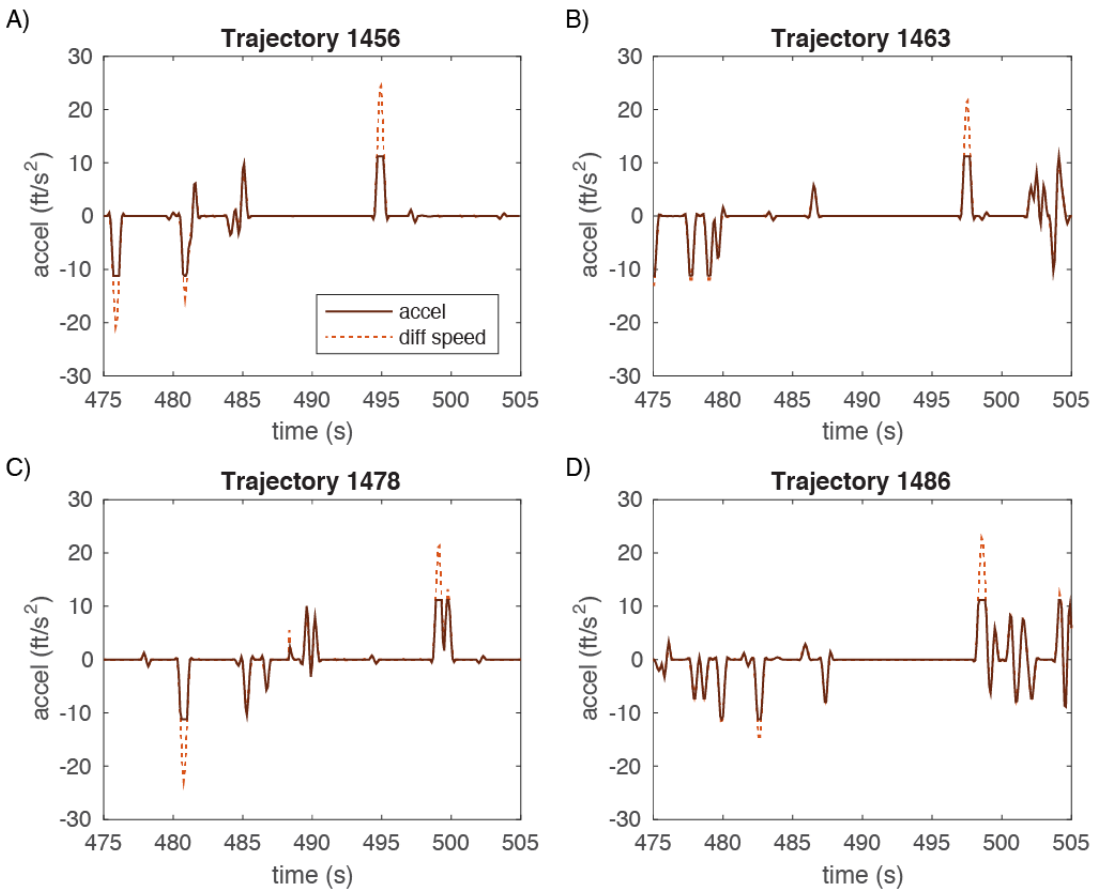

Figure 2, Portions of the acceleration time series (solid curves) for four successive vehicles from Fig. 1 and calculated acceleration from the difference in successive speeds (dashed curves). These time series are typical of the NGSIM I-80 data set.

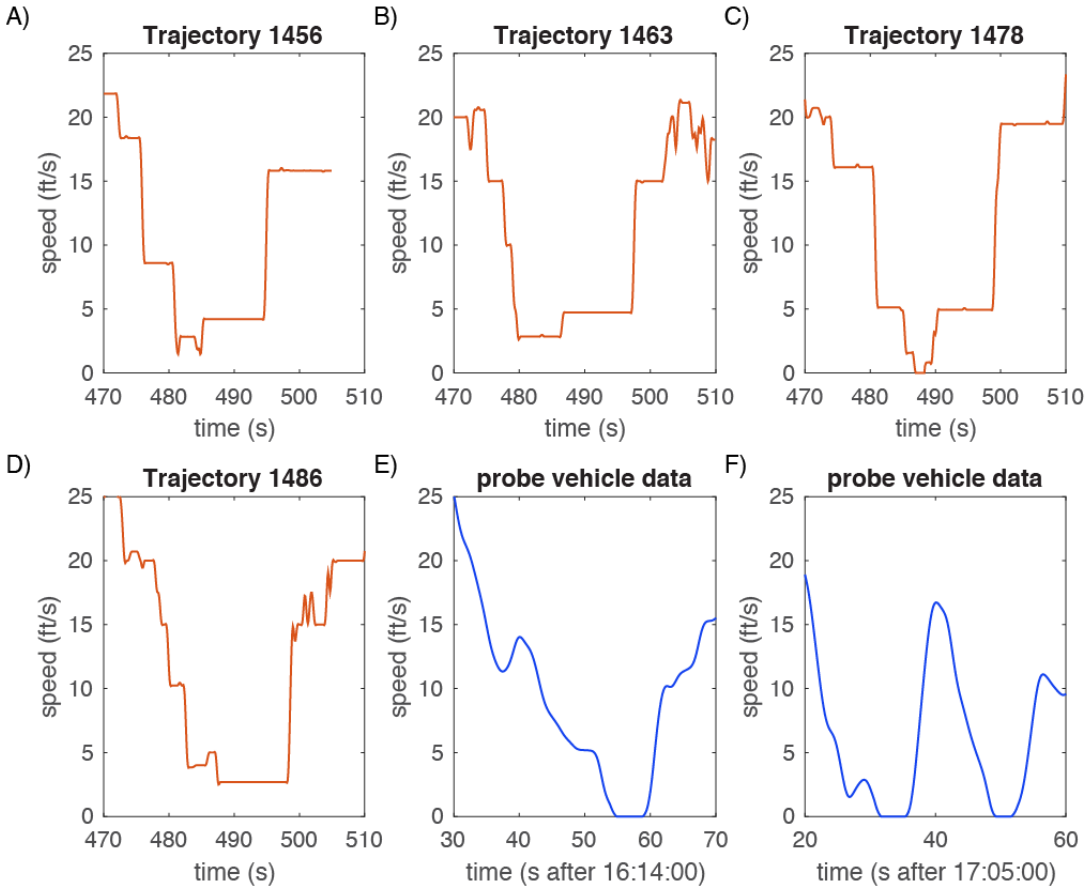

Figure 3, (A)-(D) Time series speed corresponding to the vehicles in Fig. 2. These time series are typical of the NGSIM I-80 data set. (E)-(F) For reference typical time series speed from a separate instrumented probe vehicle study under similar traffic conditions (Coifman et al., 2016).

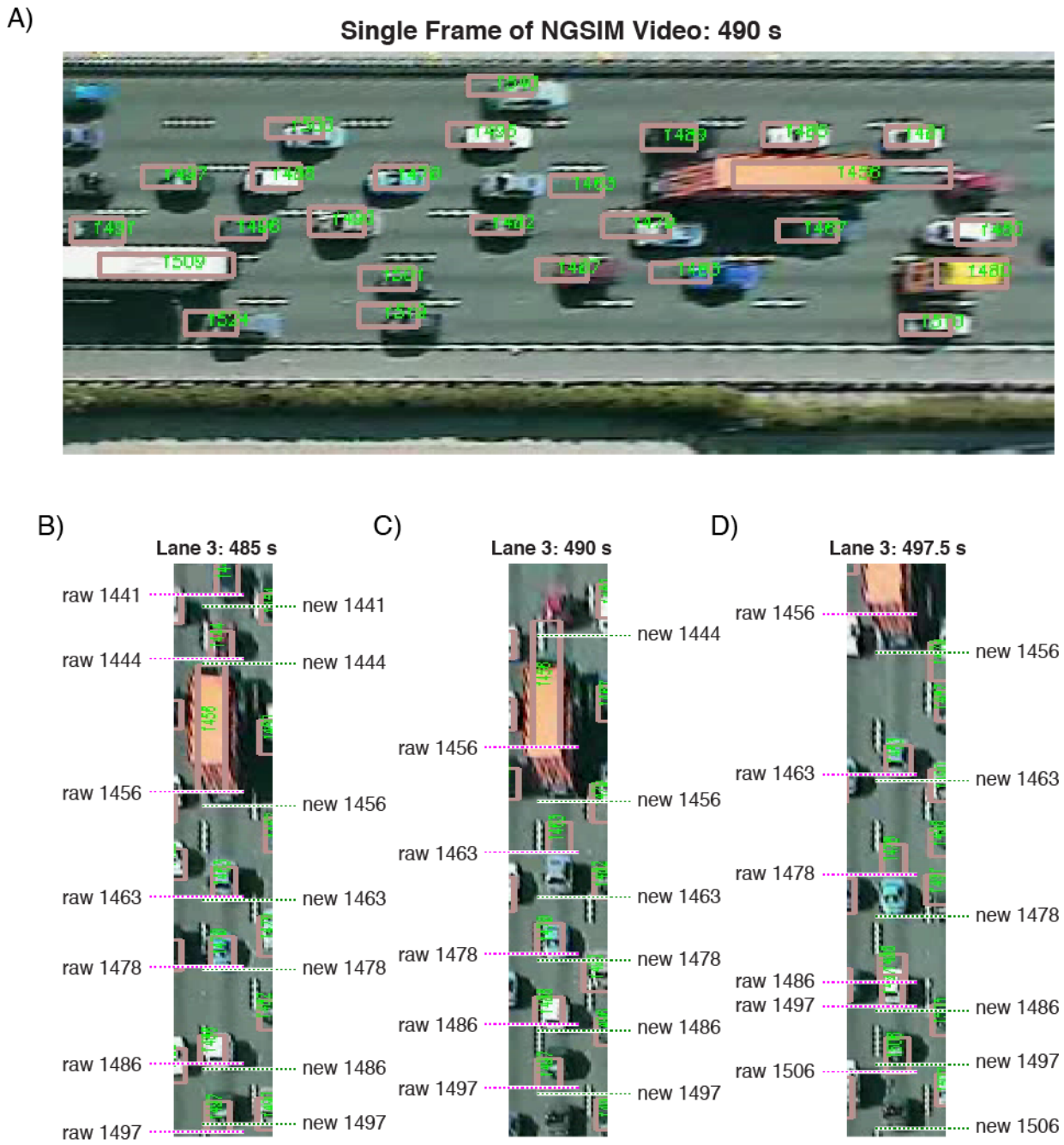

Figure 4, (A) A typical frame from the NGSIM validation video for I-80 camera 6 at 490 s. Traffic is flowing from left to right in this image. The NGSIM researchers superimposed a pink bounding box for each vehicle. A Detail of lane 3 (3rd lane from the left) from the validation video for three frames (B) 485 s, (C) 490 s, and (D) 497.5 s. For each vehicle the rear bumper location in the NGSIM data set is shown on the left (corresponding to the upstream end of the bounding box) and is indexed by the vehicle ID, while the new rear bumper location calculated in our study is shown on the right.

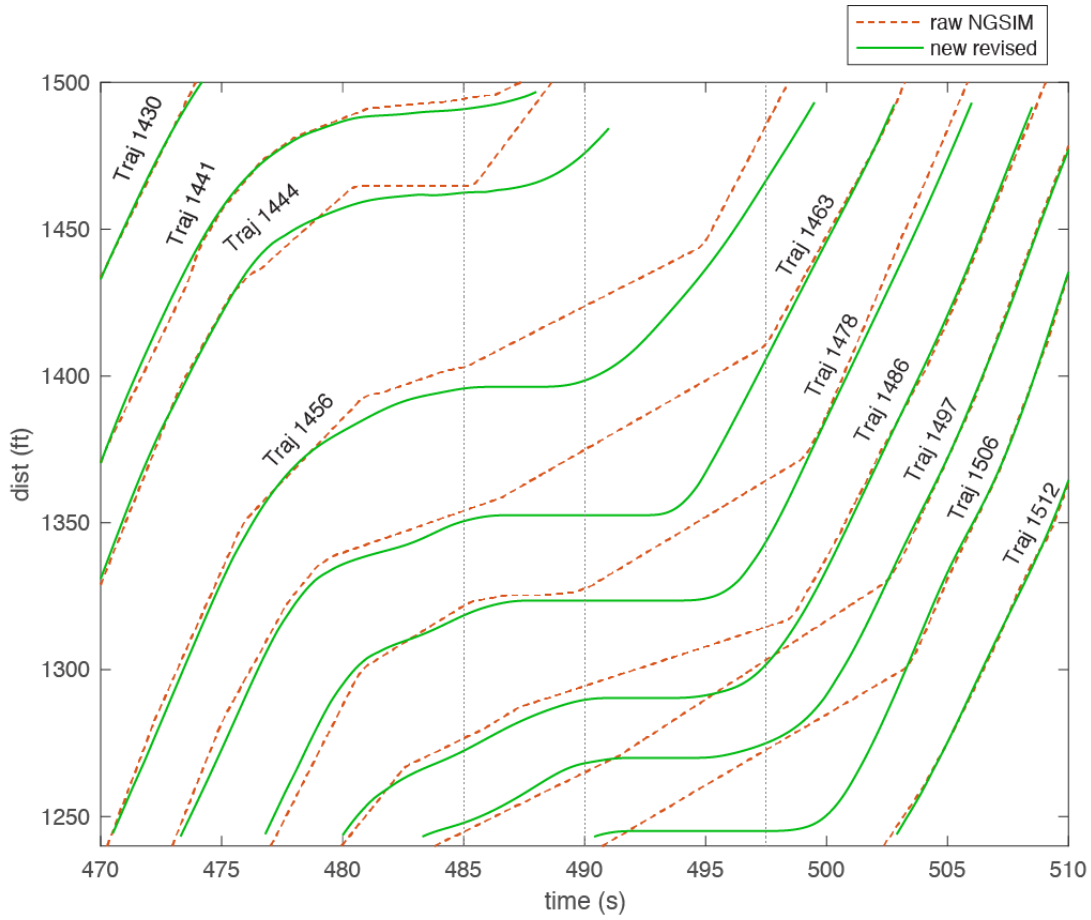

Figure 5, The resulting trajectories (vehicle rear) for 10 successive vehicles in lane 3. The raw NGSIM trajectories are shown with dashed lines and the new trajectories from this work with solid lines. Each trajectory is labeled with its ID and the three frames used in Fig. 4 are indicated with vertical lines.

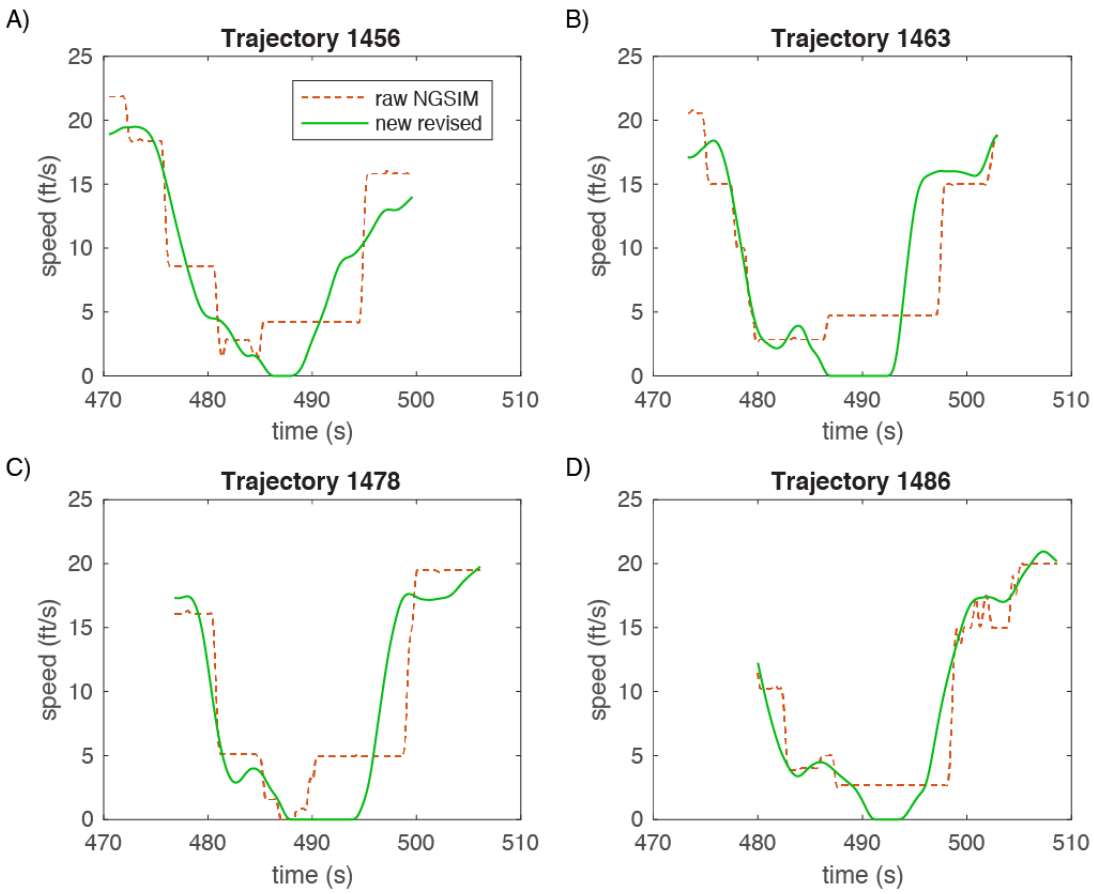

Figure 6, Reiterating the raw NGSIM time series speed from Fig. 3 (dashed curves) and the corresponding time series for the newly extracted trajectory shown with a solid curve.

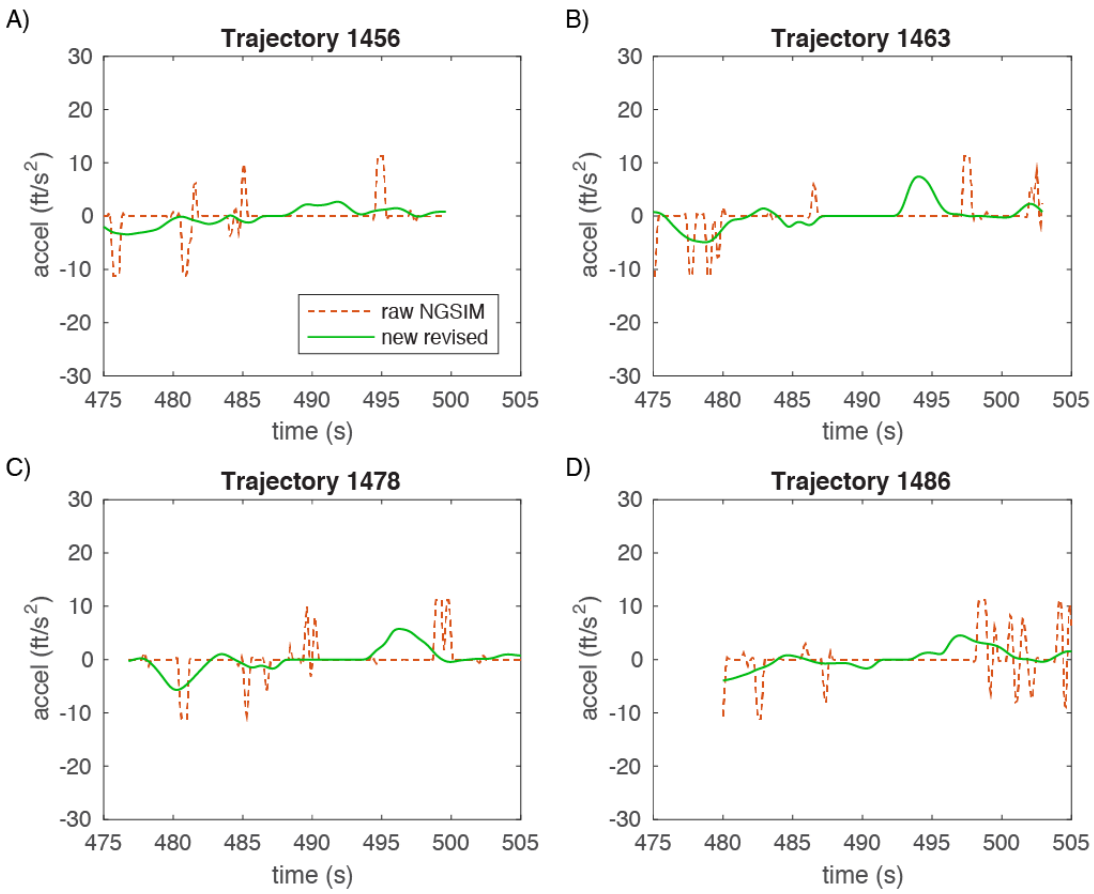

Figure 7, Reiterating the raw time series acceleration from Fig. 2 (dashed curves) and the corresponding time series for the newly extracted trajectory shown with a solid curve.

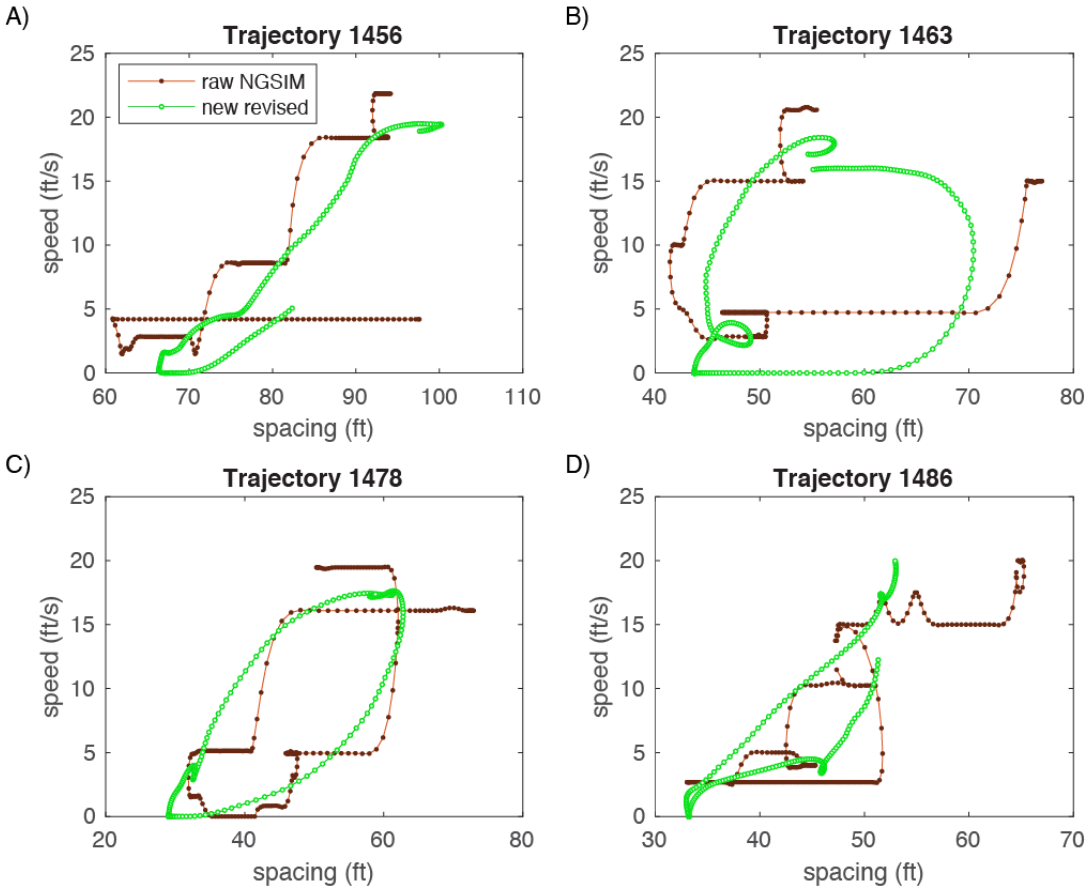

Figure 8, The resulting speed-spacing relationship for the four vehicles in the on-going example in the raw NGSIM and the same plot for the newly extracted trajectories. The contrast between the two curves in each plot showe the impact of the combined errors evident in Fig. 5 and Fig. 6. Note that the horizontal range varies from plot to plot to show the greatest horizontal detail of the respective curves.

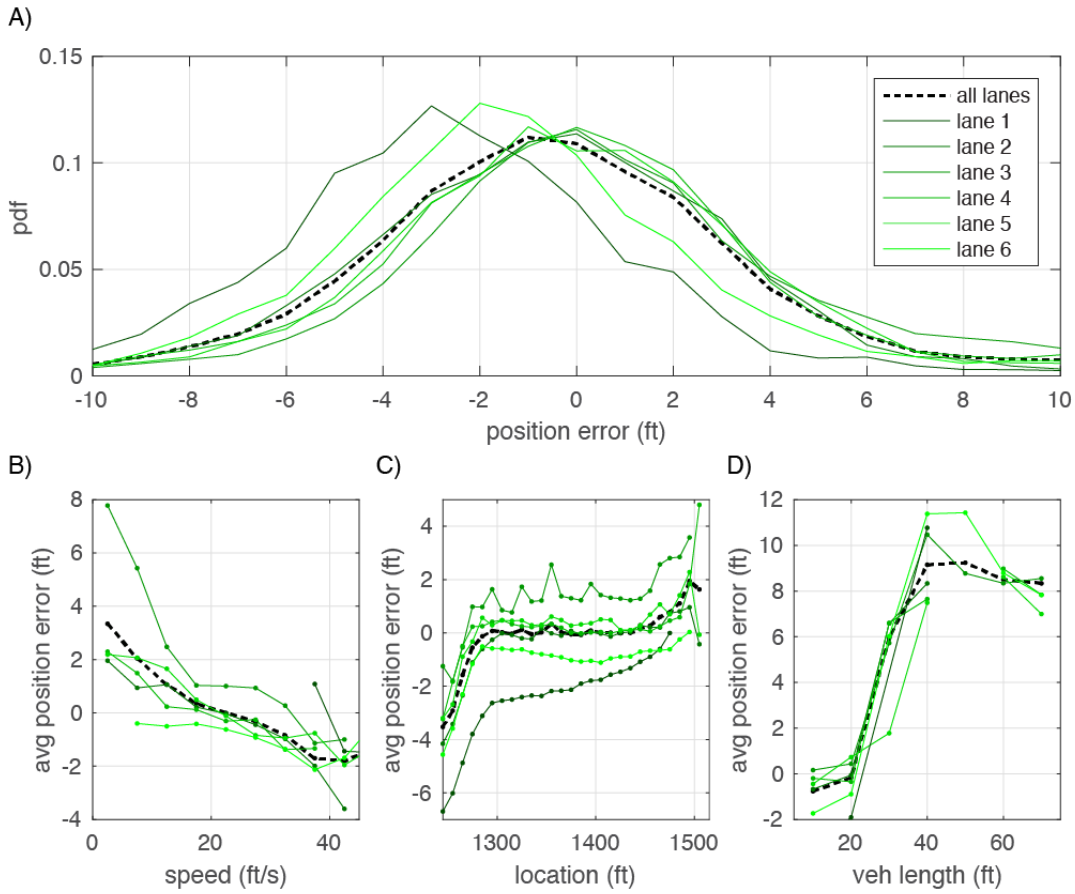

Figure 9, (A) The probability distribution function of the raw NGSIM position minus the manually re-extracted position by lane and all lanes combined together. (B) Mean position error as a function of the re-extracted speed. (C) Mean position error as a function of location. (D) Mean position error by vehicle length.

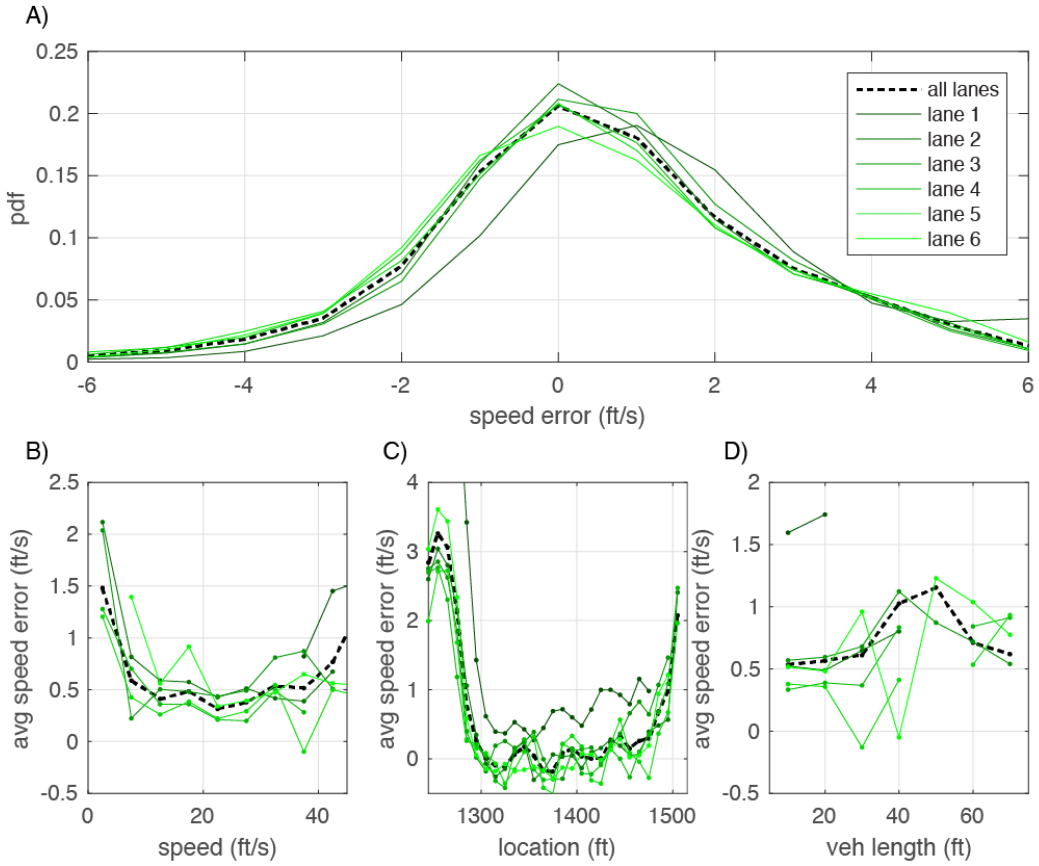

Figure 10, (A) The probability distribution function of the raw NGSIM speed minus the manually re-extracted speed by lane and all lanes combined together. (B) Mean speed error as a function of the re-extracted speed. (C) Mean speed error as a function of location. (D) Mean speed error by vehicle length.

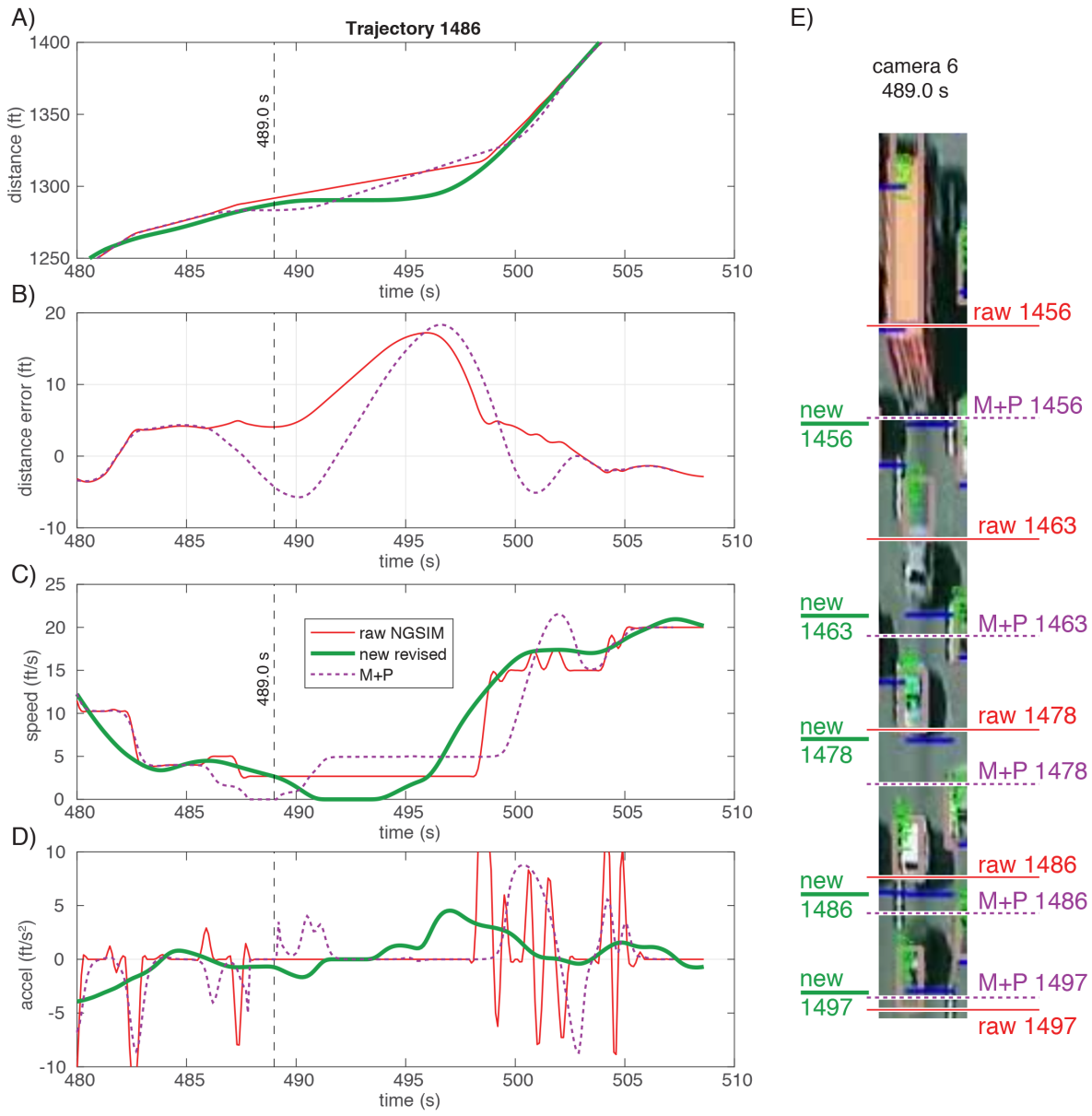

Figure 11, An example showing the raw NGSIM data and the cleaned data from Montanino and Punzo against the manually re-extracted data for vehicle 1486 (A) in the time space plane, (B) after subtracting off the manually re-extracted data, (C) time series speed, and (D) time series acceleration. While (E) shows the position of 5 vehicles from the three different data sets against a frame from the revised validation video where the blue bars show the re-extracted rear of vehicles (note that some bars from the adjacent lanes are evident, in each lane the bars are shown to span the width of the lane and many of the lane 3 vehicles are projected in to the ground plane of lane 2 to the left).