Freeway lane-changing: some empirical findings

Marcel Sala*a, Francesc Sorigueraa

*UPC – BarcelonaTECH, Jordi Girona 1-3, Building B1, Office 114, 08034 Barcelona (Spain)

Abstract

Lane changing activity is thought to play an important role in the capacity degradation of congested freeways. However, proofs of this negative impact are scarce due to the difficulties in obtaining suitable data. In this paper, the lane changing activity in the B-23 freeway accessing the city of Barcelona is analyzed. Lane changes (LC) were video recorded in six different stretches from where loop detector measurements were also available. The obtained database allowed finding a consistent relationship between LC activity and congestion. LC peaks in all analyzed sections when they become congested. This is particularly intense at the traffic breakdown, between congested and free flowing conditions. As an example, it is observed that LC activity peaks just downstream of a fixed bottleneck where free-flowing conditions are recovered. In addition, data show that the larger the lane changing rates, the smaller the maximum observable flows, supporting the hypothesis that LC is a key contributor to a capacity drop. In spite of all these findings, this research highlights the difficulty in obtaining a suitable database to definitively answer most of the research questions regarding freeway lane-changing. The spatial coverage of measurements is one of the major drawbacks. To this end, a careful planning of the data collection is necessary in order to obtain meaningful conclusions.

© 2018 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Selection and peer-review under responsibility of the scientific committee of the XIII Conference on Transport Engineering, CIT2018.

Keywords: Lane changing; freeway traffic; capacity drop; empirical traffic database; freeway congestion.

1. Introduction

Freeway congestion is a recurrent phenomenon around the world. When demand exceeds the capacity, congestion arises. The paradox is that when traffic breaks down, the available supply (i.e. the capacity) is reduced. This harmful phenomenon is known as the capacity drop, and has been found repeatedly around the world (Banks, 1991; Hall and...
Agyemang-Duah, 1991; Cassidy and Bertini, 1999; Cassidy and Rudjanakanoknad, 2005; Chung et al., 2007; Patire and Cassidy, 2011; Oh and Yeo, 2012; Srivastava and Geroliminis, 2013; Yuan et al., 2015). The typical capacity drop in active freeway bottlenecks ranges from 3% to 18% (Oh and Yeo, 2012).

While the existence of the capacity drop has been extensively demonstrated, the traffic mechanism behind it is still under debate. In the literature two traffic characteristics closely related to capacity reductions have been explored: lane changing and vehicle sluggish acceleration when leaving a queue.

Lane changing has been found to cause a capacity reduction near bottlenecks in different scenarios. In Cassidy and Rudjanakanoknad (2005) systematic lane-changing from the shoulder to faster lanes caused traffic breakdown. Patire and Cassidy (2011) found a significant flow reduction after lane-changing increases due to the existing speed variation between lanes. Also, the massive lane-changing happening at speed drops generates disturbances which end up queueing all lanes (Hatakenaka et al., 2006). In contrast, on a congested freeway, when a HOV lane is activated it smoothens traffic even in the adjacent general purpose lanes, reducing the amount of lane-changing and increasing the freeway capacity (Menendez and Daganzo, 2007; Cassidy et al., 2010).

Vehicles’ sluggish acceleration was found to be related to lane-changing in Laval and Daganzo (2006) and Leclercq et al. (2011), where lane-changing vehicles near a bottleneck were found to create larger gaps at the arriving lane due to their limited acceleration when moving from a slow to a faster lane. This leads to a total throughput reduction. These findings are in accordance with Yuan et al. (2015), where it is reported that the slower the queue speed, the greater the capacity drop.

However, the negative effects of lane-changing are not limited to being a main capacity drop contributor. Also, they trigger traffic oscillations (Mauch and Cassidy, 2002; Wang and Coifman, 2008) and stop and go waves (Ahn and Cassidy, 2007). Moreover, lane changes are found to globally increase delay in Coifman et al. (2006), as the time saved in the exiting lane is smaller than the induced delay in the arriving lane.

In spite of all these laudable findings, the lack of reliable empirical data has been a recurrent problem to increase the knowledge on lane-changing behavior and its effects. Several researchers have designed ad-hoc experiments to validate their models (Sun and Kondyli, 2010; Sun and Elefteriadou, 2012). In light of this data scarcity, big efforts have been made to construct reliable databases to support the research community. Take as an example the NGSIM project (FHA, 2005; FHA 2006). Unfortunately, the freeway traffic trajectories database resulting from the project, suffers from large errors regarding lateral vehicle position (Punzo et al., 2011). Thus, directly estimating lane changing from NGSIM database implies big errors. In order to correct this issue, Montanino and Punzo (2015) did a meticulous data filtering job to improve lane assignment in the NGSIM dataset. Still, some errors remain as Coifman and Li (2017) point out. This later work finds some of the error sources, like vehicles (mostly motorcycles) traveling on the hard shoulder, or in between lanes.

The present paper contributes in filling this void by exploring a comprehensive freeway traffic database obtained from a variable speed limit experiment on the B-23 freeway accessing the city of Barcelona (Soriguera and Sala, 2014). The database includes measurements from different types of traffic detectors, placed with high density on the analyzed freeway stretch. This allows a detailed characterization of flow, occupancy and average speed on the freeway. In addition, videos recorded by freeway surveillance cameras are also available for some zones, from which lane-changes are extracted by using a semi-automatic video processing method (Sala et al. 2018). This makes the database a particularly adequate and unique source of empirical data to assess freeway lane-changing behavior. Soriguera et al. (2017) analyzes several different traffic variables including lane changing in a single location. In the current paper the database has allowed providing empirical proofs to some key aspects regarding lane-changing in multiple locations, like its contribution to the capacity drop due to the peaking of lane-changing rates at traffic state transitions between congestion and free-flowing traffic.

The reminder of the paper is organized as follows. In section 2 the traffic database used to obtain some empirical findings regarding lane-changing is presented. This also includes a description of the methods used to process and aggregate the data. Section 3 presents the main findings and shows the lane-changing peaking in congestion and at bottleneck locations. How the inter-lane speed differences affect lane-changing and the relationship between the lane-changing rates and the average lane flow is also discussed. Finally, Section 4 outlines and highlights some of the conclusions obtained from the analysis.
2. The B-23 database

This section contains a description of the database supporting the empirical findings presented in the paper. The data was collected in the context of a dynamic speed limit experiment on the B-23 freeway accessing the city of Barcelona from the south-west, only on the inbound direction (Figure 1). The experiment took place on the last 13 Km stretch of the freeway before entering the city in 7 different days (namely Day 1 to 7). In order to ensure similar demands and traffic conditions, data was only collected with good weather (clear skies) on Tuesdays, Wednesdays and Thursdays during June, 2013. See (Soriguera and Sala, 2014) for a complete description of the experiment.

![Fig. 1. B-23 schematic layout and lane numbering](image)

Figure 1 shows the video surveillance zones (in red) and the traffic detectors close to them (blue vertical lines). In all cases, traffic detectors are between 25 and 300 m apart from the video surveillance zones. Traffic direction goes from left to right in Figure 1. Each on- or off-ramp representation means that there is at least one ramp of this type between detectors.

Traditional traffic data (e.g. flow, average speed and detector occupancy as a proxy for the freeway density) are obtained from traffic detectors. Either using the traditional double loop detectors or the newer non-intrusive devices, they are point detectors, in the sense that their measurement zone takes less than 10 meters of the freeway. Thus, the spatial coverage of the measurements is extremely limited, and spatial variables can only be indirectly derived from the measurements of consecutive detectors on the highway. This surveillance scheme is not suitable for measuring lane changes which need to be observed over space and time. To that end, video camera surveillance is used. Detection zones were set over the camera coverage along the line dividing the lanes. Lane changes were counted if at least 50% of the maneuver happened within the time-space region of the detection zone. Details about the retrieval of lane-changing data from video recordings, the different techniques used and their errors are provided in (Sala et al. 2018). Note that motorbike lane changes were discarded, as they travel quite often in-between lanes, especially in congestion, and generally represent a source of errors in lane-changing data (Coifman et al., 2017). Six different camera locations were available for recording (see Figure 1), but due to some technical limitations of the traffic management center, only three could be recorded simultaneously.

2.1. Lane changing database

Lane changes are obtained from video recordings. For each detection zone, a database of individual lane-changing maneuvers, \( c(p, h, f) \), is constructed. The attributes of each lane-change \( (c) \) are: \( (p) \), the neighboring pair of lanes between which the change took place; \( (h) \) the height in the detection zone (i.e. the spatial position within the detection zone); and \( (f) \), the video frame (i.e. the time of occurrence). This lane-changing database can be aggregated by section (all pairs of lanes), space and/or time, if necessary.

Note that the previous lane-changing measurements are influenced by: i) the distance encompassed by each detection zone and ii) the traffic flow. For a given stretch under stationary traffic conditions, the longer the detection zone, the more lane-changes are expected to be observed. In addition, the higher the traffic flow is (i.e. veh/unit time), the larger the number of candidates to change lane. To face these issues, the Lane Change Probability (LCP) is defined in Equation 1. This is simply a normalized version of the lane-changing rate, where \( d \) is the distance encompassed by the detection zone and LC and \( q \) are respectively the measured number of lane changes and the
flow in the sampling interval, $\Delta t$. For type I detection zones the LCP is computed using the arithmetic mean of the flows from both detectors.

$$\text{LCP} \ [\text{dimensionless}] = \frac{\int_{t_1}^{t_2} q \, dt}{(t_2 - t_1) \left[ \frac{\text{veh}}{\text{km}} \right]}$$

(1)

2.2. Data available from traffic detectors

Traffic detectors measure flow ($q_{tl}$), speed ($v_{tl}$) and occupancy ($\rho_{tl}$). They provide a read out each minute ($t$), for each lane ($l$). These measurements can be aggregated or averaged for the whole section and for extended periods of time. Flow is additive if considering multiple lanes or extended periods of time. The average occupancy in this extended measurement region is simply obtained as the arithmetic average of single measurements. In contrast, in order to obtain the average speed, each speed measurement needs to be weighted by its respective measurement.

3. Lane changing: some empirical findings

Several interesting relationships between traffic states and lane changing activity are unveiled if adequate data processing is applied to the presented database. Data processing includes the construction and plot of oblique cumulative curves. These curves were introduced by Cassidy and Windover (1995) and they allow observing much richer detail than by using the usual cumulative curves. The difference in oblique curves resides in subtracting to the cumulative curve a constant value over time. This value is chosen to be close to the mean. By eliminating the large cumulative average, the plot shows the subtle variations of the variable instead of the fairly constant increase. In the following subsections, detailed insights obtained from the temporal, spatial and corridor wide analyses using oblique cumulative curves are presented.

3.1. Lane changing peaks in congested periods

Lane-changing peaks in congested periods. In some detection zones this phenomenon is particularly intense when transitioning between free flow and congested regimes. This fact is supported by the evidences presented in Figure 2. For the sake of brevity only few examples are presented, but the same behavior has been observed in every single congestion episode, with just a couple of apparent counterexamples that will be discussed later. Plots on Figure 2 consist of three cumulative oblique curves, $N(t)$, $T(t)$ and $L(t)$. $N(t)$ is the cumulative flow (i.e. the total number of vehicles that have crossed the section since the beginning of observation). $T(t)$ represents the same concept but considering the occupancy (i.e. the total cumulative time all vehicles spent on the detector since the beginning of observation). $L(t)$ represents the same concept but considering the occupancy (i.e. the total cumulative time all vehicles spent on the detector since the beginning of observation). Finally, $L(t)$ is again the same concept but considering lane-changing maneuvers for all the detection zone, including all lane pairs. Congestion is observed as an increase of the slope of the $T$-curve without an equivalent increase in the slope of the $N$-curve. In order to ease observations in Figure 2, congestion is shaded with light grey. Clearly, lane-changing rates increase in congestion, as shown by the increase of the slope of the $L$-curve during these periods.
One apparent counterexample is shown in Figure 3a. The figure shows the same previous oblique cumulative curves but for detection zone 2305 and considering the traffic data from the downstream detector (i.e. 10 ETD). Three peaks in the lane-changing rate are observed (i.e. around 8:00; between 8:20 and 9:10; and between 9:40 and 9:50 approximately). In contrast, data from detector 10 ETD (Figure 3a) exhibits free-flowing conditions during the whole observation period. The issue is that while the downstream detector location is not congested, actually there is congestion within detection zone 2305. A bottleneck is caused by the off-ramp (see Figure 1) in the zone. This ramp ends at a roundabout which during peak periods reaches capacity and queues are created. These queues reach the freeway mainline and spills back into the rightmost lanes, eventually congesting all the freeway trunk. This can be observed in Figure 3b, equivalent to Figure 3a but considering the data from the upstream detector (i.e. 11 ETD). It can be seen that congestion reaches detector 11 ETD three times which approximately match the three surges in the lane-changing rates. Peaking times between T(t) and L(t) slightly differ due to the spatial differences between the bottleneck and the 11 ETD detector locations.

Fig. 3. Bottleneck at detection zone 2305. LC data was collected on detection zone 2305. Traffic data source: (a) detector 10 ETD; (b) detector 11 ETD. All data was collected on 4th June 2013.

Given this context with partial congestion upstream of the off-ramp in detection zone 2305, it is interesting to analyze the spatial distribution of lane-changing within the zone. This is illustrated in Figure 4, consisting on a speed contour plot with all the lane-changing activity on top. The spatial variation of the speed is computed by linear interpolation between the upstream and downstream detectors. The linear interpolation is just to give the reader an approximate idea of the speed value within the detection zone.

Fig. 4. Lane-changing spatio-temporal distribution in detection zone 2305. (a) Two rightmost (i.e. fast) lanes; (b) Two leftmost (i.e. slow) lanes. Note: The spatial speed variation is the result of linear interpolation between the upstream (i.e. 11 ETD) and the downstream (i.e. 10 ETD) detectors. Speed values are computed as the mean between the adjacent lanes considered. Data collected on 4th June 2013 (i.e. Day 1).

Observing Figure 4a, which shows the behavior of the two left-most (i.e. fast) lanes, lane-changing appears to be homogeneously distributed in space; regarding the time distribution, a slight increase is observed with slower speeds (i.e. congestion). However, in the rightmost (i.e. slow) lanes (see Figure 4b) the situation is very different: there exists a strong concentration of lane changing after the bottleneck location (i.e. off-ramp). Again, and with more intensity, the lane-changing activity is higher when the bottleneck is active (i.e. upstream congestion). Note also that during congested periods, the lane-changing increase is extended to just upstream of the off-ramp.

This behavior shown in Figure 4b can be explained by a combination of factors. First, the increase in lane-changing rates just before the off-ramp in congested conditions could be due to some vehicles leaving the rightmost
lane towards the middle lane in order to avoid the growing queue in the former. Other vehicles perform the opposite maneuver, moving from the middle to the rightmost lane and cutting in the queue in order to take the off-ramp at the last moment avoiding part of the congestion. These late lane-changing vehicles do not obey the solid line explicitly forbidding this maneuver. Unfortunately, the relative amounts of each type of maneuver cannot be determined as the method does not provide the direction of the lane-changes. This weaving clearly contributes to a capacity drop in the section. After the off-ramp, the solid line ends, and the voids left in the rightmost lane by those vehicles who have exited at the off-ramp are filled by vehicles coming from other lanes. To some extent, this behavior is expected to happen after all types of bottlenecks, where the recovering of free-flowing condition implies a lesser density and an increase of lane changing rates in order to accommodate vehicles according to their desired free-flow speeds. Obviously, this situation is magnified in contexts where the congestion upstream of the bottleneck is concentrated in one or few lanes (like the situation analyzed here).

Figure 5 further illustrates the spatial distribution of lane-changing activity in detection zone 2305, in this case considering all days in the database. In Figure 5, lane-changes are spatially aggregated in 10 meters sections. The previously described behavior is confirmed, and the lane changing between lanes 1 and 2 (i.e. leftmost) is quite homogeneous over the 290 m long detection zone. In contrast, lanes 2-3 (i.e. rightmost) show a much greater lane-changing rate just after the off-ramp until the end of the measuring zone (i.e. about 70 meters downstream of the off-ramp). In this area, lane-changing is four times more intense at the peak than anywhere else on this detection zone. Note from Figure 5 that the increase of lane-changing just upstream or just at the off-ramp, observed in Figure 4b during congested periods, is much subtler. This is due to the fact that congested periods represent only a fraction of the time considered in Figure 5.

3.2. Lane changing probability vs. maximum flow

Figure 6a plots the average flow with respect to the lane changing probability for different detection zones. Note that the average flow corresponds to the average sectional flow of one of the traffic detectors in the detection zone, although it has been normalized per lane in order to account for sections with a different number of lanes. In turn, the lane-changing probability (LCP) is computed as the average number of lane-changes for each travelling vehicle in one Km (see Equation 1). This means that theoretically, LCP can be larger than one because vehicles could change lanes more than once in one km. All values in Figure 6a are computed over 5 min. time aggregations.

Results show an upper bound to the maximum observable lane changing probability for a given flow (see Figure 6a). Different traffic conditions in the detection zone can exist for the same flow level, leading to higher or lower lane-changing rates. However, the maximum lane-changing rate is bounded, meaning that if the lane-changing exceeds this maximum rate, this will be translated into a reduction of the sectional flow. As discussed previously, lane changes are a disruptor in traffic flow, implying that the largest flows can only be achieved with small to none lane-changing. Note from Figure 6a that no data is available for small flows because data was collected during the rush hour. Therefore, inferences for flows under 1000 veh/h·lane must be taken with extreme caution.

Figure 6b shows the maximum expected number of lane-changes per lane, hour and km according to the maximum lane-changing probability defined in Figure 6a by linear regression. This number needs to be multiplied by the number of lanes in the analyzed section in order to obtain the maximum number of lane-changes/h/km. The resulting parabolic model exhibits a maximum of 1062 lane changes/l/h/Km when the circulating flow is 1130 veh/h/lane. The model considers a maximum capacity of 2265 veh/h·lane with zero lane-changes. This model could
be useful for lane-changing control applications, as it determines the maximum acceptable number of lane-changes to ensure a desired throughput of the freeway section.

Fig. 6. (a) Lane-changing probability versus average flow. Note. All data computed over 5 min. aggregation period; (b) Model for the maximum number of lane-changes per km between any pair of lanes as a function of the average flow per lane.

4. Conclusions

Lane-changing activity is thought to be one of the major disruptors in freeway traffic, and a key contributor to traffic instabilities. The empirical research presented in the paper supports this disruptor role of lane-changing in freeways. It has been found, for multiple locations and days, that lane-changing activity peaks during congestion. Even more, the largest concentration of lane-changes has been found to be located around shock waves which imply a transition between traffic states. Note that a bottleneck at a fixed location is a particular case of a shock wave with zero speed. In fact, this was the most severe case found, with large peaks of lane-changing rates around diverging bottlenecks. Although no direct evidence has been found of lane-changing activity being the trigger of congestion episodes, a clear relationship between the maximum lane change probability and the average lane flow has been observed. The larger the flow, the smaller the maximum lane-changing probability. Note that the lane changing probability is defined as the number of lane-changes per km over the travelling flow. With this definition, if the number of lane-changes are kept constant, the lane-changing probability would decrease by definition. However, the assumption of constant lane-changes with increasing flow actually represents a decrease of the unitary lane-changing activity. This means that results hold, in spite of the definitions and variables used. These findings match the prevailing theory in the literature arguing that lane-changes are an important traffic disruptor which can trigger congestion. In consequence, largest flows can only be achieved with few to none lane-changing, as there is no room for disruptions in such high flows. Therefore, lane-changing must be limited in peak periods, or otherwise can potentially make traffic unstable leading to a capacity reduction.

The analyses performed in the paper also allowed showing that empirical research in traffic is very sensitive to the quality and suitability of data. Some examples of the shortcomings and flawed conclusions that can arise if using inadequate data or data treatment processes have been highlighted in the paper. Especially conflictive is the location of detectors with respect to bottlenecks. In fact, the spatial representability of point detectors is a recurrent problem in traffic research. Using detailed trajectories data, which could also be retrieved from video recordings, would not only solve the spatial representability, but it would also increase the richness and detail of the available information. Nevertheless, the advantage of the simpler database used here is the much shorter time needed to retrieve the data from the video while keeping the error in the lane-changing measurement low (Sala et al., 2018). Note that automatically retrieved trajectory data from video recordings still has large errors regarding lane-changing, even if using the most powerful imaging techniques available to date.

5. Acknowledgments

Authors acknowledge the collaboration of Mr. Adria Torres and the comments of Mrs. Margarita Martinez. The database construction would not have been possible without the collaboration of the Servei Català de Trànsit. This research has been partially funded by the Spanish Ministry of Science and Innovation (TRA2016–79019-R/COOP).
6. References


Soriguera, F., Martinez, I., Sala, M., Menendez, M., 2017. Effects of low speed limits on freeway traffic flow. Transportation Research Part C 77, 257-274.


