RELATIONSHIPS BETWEEN DISCRETIONARY LANE-CHANG AND TRAFFIC CONDITIONS

Minju Park  
Department of Civil and Environmental Engineering, KAIST,  
291 Daehak-ro, Yuseong, Daejeon, Korea  
E-mail: cozypmj80@kaist.ac.kr

Kitae Jang  
The Cho Chun Shik Graduate School for Green Transportation, KAIST,  
291 Daehak-ro, Yuseong, Daejeon, Korea  
E-mail: kitae.jang@kaist.ac.kr

Hwasoo Yeo* (Corresponding author)  
Department of Civil and Environmental Engineering, KAIST,  
291 Daehak-ro, Yuseong, Daejeon, Korea  
Phone: +82-42-350-3634  
Fax: +82-42-350-3610  
E-mail: hwasoo@kaist.ac.kr

ABSTRACT  
This paper evaluates the probability of discretionary lane-changing as functions of speed difference and density difference between adjoining lanes under congested traffic condition. NGSIM trajectory data were aggregated via mesoscopic approach and analyzed to obtain detailed information on traffic conditions near discretionary lane-changing maneuvers. We first constructed joint probability distribution of lane-changing in speed difference and density difference domain. The distribution shows that speed difference and density difference significantly influence lane-changing rate. The probability distribution is then quantified by using logistic regression. The result showed that drivers change the lane more frequently as target cell has higher speed and lower density than origin cell. It is also found that the relationship between the probability of lane-changes and speed difference is non-linear, as well as the relationship between the probability of lane-changing and density difference.
1. Introduction

Lane-changing is the lateral movement of vehicles across the roadway; and creates interactions between traffic in two adjacent lanes. Lane-changing is a frequent event in roadway traffic and also has significant impacts on traffic streams: i) increased lane-changing maneuvers in the vicinity of bottleneck reduces discharge rate of the bottleneck \([1, 2]\); ii) it synchronizes traffic flows across lanes \([3]\); iii) it triggers shock wave under congested traffic conditions and, therefore, initiates formation and evolution of oscillations (i.e., stop and go traffic) \([4-6]\). Due to its consequences, it is important to understand in what circumstances the drivers change their lanes, and how to model lane-changing phenomenon.

Macroscopic and microscopic approaches have been used to analyze the impact of LC on traffic streams. Macroscopic approach views vehicles collectively as a fluid stream, and describes them with aggregate variables such as traffic density, speed and density. Many studies have been carried out to understand various characteristics of lane-changing traffic based on kinematic wave (KW) theory \([7-12]\). However, these studies do not fully explain all the events related with lane-changing. Lane-changing is an event which happens discretely, so it is difficult to explain the event within macroscopic models which view the lane-changing as aggregated continuum flow. Using this macroscopic concept, it is impossible to understand the individual driver’s behavior and its influential effect to the following vehicle while drivers change the lane. They are limited by focusing only on the merging, weaving and lane-drop sections.

In microscope traffic models, many studies explain behaviors of lane-changing vehicle based on gap acceptance model. These models require many parameters to consider complex behavioral decision-making processes \([13, 14]\). Many researchers have tried to find a reasonable range of gap acceptance to change the lane, particularly at merging or weaving section \([15-17]\). These models can describe detailed lane-changing behaviors, but cannot directly explain the effect of lane-changing to the entire traffic environment at macroscopic level.

To overcome these problems, Laval \([18]\) developed a hybrid model of lane-changing traffic that combines macroscopic and microscopic models to represent heterogeneous traffic steams. He merged together a continuum KW stream with models describing the acceleration capabilities of each individual slow vehicle, but some input parameters required for lane-changing in this model were not empirically verified.

Lane-changes are initiated by driver’s desire to travel better driving conditions (discretionary), and/or to enter/exit the freeway (mandatory) \([13, 19-22]\). Therefore, the latter is dependent on drivers’ origin-destination pairs and on-/off-ramp locations. Meanwhile, the former is determined based on heterogeneous traffic conditions such as speed and density differences across lanes. However, it is not completely unveiled how difference in traffic conditions relates to discretionary lane-changing. There is lack of empirical evidence about
discretionary lane-changing, since it is difficult to identify motivation for the discretionary lane changing.

In some paper [18, 23, 24], discretionary lane-changes are assumed to be triggered by speed differences between adjacent lanes. They assumed that the probability of lane-changing proportionally increases as speed difference increases. However, the linear relation between increasing lane-changing probability and speed difference is not verified from the data. Recently, Kan et al. [25] presented a systematic investigation of driver’s motivation during discretionary lane changes movements on a multi-lane freeway section. Nonetheless, they did not suggest any reasonable method for identifying discretionary lane-changing. Knoop et al. [26] studied the traffic conditions, speed difference and density values in origin lane and target lane, which lead to a number of discretionary lane changes in free flow condition. This study has a limitation by showing the number of lane-changing influenced by only one factor, speed difference or density.

This study focuses on traffic conditions (speed and density) around discretionary lane-changing vehicle under congestion using microscopic data (NGSIM). Since it is hard to properly distinguish between mandatory and discretionary lane changing with NGSIM data, in this research, discretionary lane changing is defined by reasonable data filtering process.

This paper introduces cell-based model that simulate realistic traffic conditions around lane changing vehicles. The ultimate purpose of this research is to compare and analyze relative conditions (relative speed and relative density) using real data in order to better understand lane changing behavior and the likely causes and environments of such behavior. In addition, to examine the quantified lane-changing probability, logistic regression will be conducted with collected data.

In the remainder of this paper, section 2 introduces data which is used for analysis. We explain NGSIM data and method for discretionary lane changing extraction. Section 3 explains data analysis by mesoscopic approach to consider surrounding traffic conditions of lane-changing vehicle. Section 4 then shows the joint probabilities distribution of lane changing according to the distribution of the speed difference and the density difference between origin and target cells and between origin and previous cells. And to quantify traffic conditions surrounding a lane changing, we use the logistic regression analysis. Section 5 presents the conclusions.

2. Data Preparation

1) Data Description
The Next Generation SIMulation trajectory data [27] were analyzed to obtain detailed information on traffic conditions near discretionary lane-changing maneuvers. NGSIM data transcribes the vehicle trajectory data from the video. The vehicle trajectory data provide the precise location of each vehicle within the study area every one-tenth of a second. Although
NGSIM provides two sets of trajectory data collected from I-80 and US-101, the latter was used in the paper because I-80 site includes a HOV lane that possibly influences the lane-changing behavior. This dataset consists of a total of 45 minutes of data under congestion, segmented into three 15-minute periods: 7:50 a.m. to 8:05 a.m.; 8:05 a.m. to 8:20 a.m.; and 8:20 a.m. to 8:35 a.m.

2) Discretionary Lane Changing Extraction
First, lane number 5, 6, 7, and 8 will be removed for US 101 data. 7th an 8th lanes stand for the on-ramp and the off-ramp respectively. 5th lane is the rightmost lane connecting both on-ramp and off-ramp. And 6th lane is the additional lane for smooth connection of off-ramp and on-ramp to the road. These lanes may be greatly influenced by the mandatory lane changing.

Second, lane changing from left to right is removed. We can assume that mandatory lane changing takes place from left to right because the off-ramp is located by right side of roads. On the contrary, discretionary lane changing can be assumed as all lane-changes from right to left. It is possible that discretionary lane changing is included in left to right lane changing, but it is difficult to track all of lane changing respectively to check whether it is discretionary or not. In this paper, it is assumed that all lane changes from right to left are discretionary lane changes. In this paper, the number of lane changing on the first and the last two cells is excluded due to incomplete data. Total* means the number of lane changing except for the number of lane changing on the first and the last two cells. And also, the lane changings by motorcycle and by the data error in wrong video tracking are removed from the number of lane changing. Because most motorcycles run on the center of two lanes in zigzags under congestion, it makes many lane changings in a short time. We checked directly these errors with NGSIM video.

Table 1 Comparison of the number of lane changing between NGSIM report and extracted data

<table>
<thead>
<tr>
<th></th>
<th>NGSIM report</th>
<th>Extracted Data</th>
<th>Extracted Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Total*</td>
<td>Total*</td>
</tr>
<tr>
<td>Number of LC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>07:50-08:05</td>
<td>327</td>
<td>327</td>
<td>254</td>
</tr>
<tr>
<td>08:05-08:20</td>
<td>228</td>
<td>227</td>
<td>191</td>
</tr>
<tr>
<td>08:20-08:35</td>
<td>256</td>
<td>253</td>
<td>212</td>
</tr>
<tr>
<td>Total</td>
<td>811</td>
<td>807</td>
<td>657</td>
</tr>
</tbody>
</table>

Total*: the number of extracted LC except for the number of LC in the first and the last two cell
Total**: the number of extracted LC except for the lane changing by motorcycle and the data error in wrong video tracking

3. Data Analysis Method

These data are microscopic because they record every vehicle’s movement within freeway sections in millisecond interval. However, microscopic data have limitations because the data contain noise from statistical fluctuations and are unable to capture realistic ranges of traffic situations surrounding lane-changing maneuvers (drivers do not make decisions at this level of microscopic space and time). On the other hand, macroscopic approach has not been successfully in modeling lane-changing behavior because this approach fails to explain the phenomena regarding vehicles deceleration and accelerations [7-11].
Due to the limitations in both macroscopic and microscopic approaches, hybrid models were introduced to simulate realistic traffic conditions [18]. However, the hybrid approach in those studies has not been applied to analyze lane-changing behavior and its surrounding traffic conditions. In this study, mesoscopic analysis method is used to consider surrounding traffic conditions.

In this approach, the roadway is segmented into N homogeneous sections (cells) with the same length for each lane as shown in Figure 1. Cell length is determined as moving distance of vehicle during the perception-reaction time of 1.5sec [28] on free-flow speed of 60mph. One cell length is 40m (131.2ft). This distance may be the reasonable distance that drivers scan traffic conditions while they are driving under congestion. In every time interval, traffic data within each cell were aggregated to smooth statistical noise in the microscopic data while keeping the necessary information. We used these processed data to analyze lane-changing behavior.

When the vehicle within the $i^{th}$ cell in lane $l$, cell $(i, l)$, change the lane to cell $(i, l-1)$, we collect the traffic data from the origin cell, cell $(i, l)$, and target cells, cell $(i, l-1)$. Since drivers often make decisions proactively by scanning not only origin and target cells but also cell in front of origin cell. Hence, data from cell $(i+1, l)$ are also collected. The collected data include acceleration rate, speed and density. To evaluate the probability of lane-changing, logistic regression analysis is conducted: dependent variable is binary (lane-changing or no lane-changing) and explanatory variables are collected traffic data from surrounding cells.

From the extracted data, we can compute speed difference and density difference between origin $(i, l)$ and target cells $(i, l-1)$ both lane-changing and non-lane-changing events.

$\Delta v^L = v^L_{i,l} - v^L_{i,l-1}$, $\Delta k^L = k^L_{i,l} - k^L_{i,l-1}$

$\Delta v^{nL} = v^{nL}_{i,l} - v^{nL}_{i,l-1}$, $\Delta k^{nL} = k^{nL}_{i,l} - k^{nL}_{i,l-1}$

By the equation (1), speed difference and density difference can be found for all the cells in both cases of lane-changing and non-lane-changing events. Figure 2 show CDFs of speed difference between origin $(i, l)$ and target cells $(i, l-1)$ for all data and lane changing data, respectively. In Figure 2(a), the data which includes both the lane-changing and non-
lane-changing events follows almost normal distribution, so it is symmetric. Meanwhile, for the data which includes only lane-changing event, the distribution is skewed towards negative difference in Figure 2(b). It shows that discretionary lane changing has tendency to happen, if the speed of the target cell is greater than the one of origin.

Figure 3 show CDFs of density difference between origin \((i, l)\) and target cells \((i, l-1)\) for all data and lane changing data, respectively. Density difference is discrete data, but this result for all data shows almost symmetric with density 0 in Figure 3(a). Meanwhile, the distribution of density difference for Lane Changing data in Figure 3(b) is skewed towards positive difference. It means that discretionary lane changing has tendency to happen, if the density of the target cell is lower than the one of origin.

![Diagram](image)

**Figure 2** Cumulative distribution function of (a) Speed difference for all data and (b) Speed difference for LC data between origin \((i, l)\) and target cells \((i, l-1)\) for US-101 data

![Diagram](image)

**Figure 3** Cumulative distribution function of (a) Density difference for all data and (b) Density difference for LC data between origin \((i, l)\) and target cells \((i, l-1)\) for US-101 data

Superscript L and nL indicate lane-changing and non-lane-changing cases,
respectively. $\Delta v$ is speed difference between cell($i, l$) and cell($i, l-1$). $\Delta k$ is density difference between cell($i, l$) and cell($i, l-1$). $v_{i,l}$ is speed in cell($i, l$). $k_{i,l}$ is density in cell($i, l$). With the count data, we could compute probability of lane-changing as follows:

$$P(v_n < \Delta v^L < v_{n+1}) = \frac{\sum I(v_n < \Delta v^L < v_{n+1})}{\sum I(v_n < \Delta v^L < v_{n+1}) \times k_{i,l} \times 0.1 + \sum I(v_n < \Delta v^{NL} < v_{n+1}) \times k_{i,l} \times 0.1}$$

where, $I = \begin{cases} 1 & \text{if } (v_n < \Delta v^L < v_{n+1}) \\ 0 & \text{otherwise} \end{cases}$

By multiplying the denominator with $k_{i,l}$ (the number of vehicles within a cell) and time step (0.1sec), it can be expressed as Vehicle Hours Traveled (VHT). Then, above equation represents the number of lane changing per VHT. If density in target cell is 0, target cell’s speed is regarded as zero, resulting in the positive relative speed. The speed of origin cell with zero density is assumed to be free flow speed (60mi/h=88ft/s). From the collected data, average speed and density were computed across samples within each cell at each time step. From the extracted data, we can compute speed difference and density difference between origin ($i, l$) and target cells ($i, l-1$) and between origin ($i, l$) and previous cells ($i+1, l$) both lane-changing and non-lane-changing events because drivers proactively evaluate one time step later.

4. Result

1) Descriptive Statistics Analysis
Figure 4 shows joint probabilities of lane changing as functions of speed difference and density difference between origin ($i, l$) and target cells ($i, l-1$) and between origin ($i, l$) and previous cells ($i+1, l$). In Figure 4, Blue color indicates low lane-changing probability while red color presents high lane-changing probability.

In Figure 4 (a), most of discretionary lane changes are on the upper-left quadrant. This means when the vehicles change their lanes, drivers make decisions based on the speed and the density of target cell (next cell) with respect to their own speed and density. The drivers may seek to travel faster in the target lane and have enough space at the time of lane changing.

In Figure 4 (b), lane changings are evenly distributed over a wide area, implying that the previous cell’s density or speed is not a significant factor for drivers’ lane-changing decisions. Figure 4 represents joint probabilities distribution. The evaluation of quantitative relations between lane-changing and speed or density is included in the following logistics regression analysis.
2) Logistic Regression Analysis

To quantify the relation between LC and surrounding traffic conditions, we use the logistic regression analysis. Logistic regression is a type of regression analysis used for predicting the outcome of a categorical criterion variable based on one or more predictor variables. Binary logistic regression refers to the instance in which the criterion can take on only two possible outcomes (lane changing or non-lane changing). The multiple logistic regression model has the form

\[ y_i = \ln \left( \frac{P_i}{1 - P_i} \right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + e_i, e_i \sim \text{Normal}(0, \sigma^2) \tag{3} \]

The formula for the probability itself is

\[ p_i = \Pr(Y = 1|x) = \frac{e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)}}{1 + e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)}} \]

\[ + e_i^* = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p)}} \]

Odd ratios provide a measure of describing the strength of the partial relationship between an individual predictor and the predicted event. The odds ratio is computed quite simply as \( e^{\beta_i} \). Exponentiating a beta parameter provides the multiplicative effect of that predictor on the odds, controlling for the other variables. The farther \( \beta_i \) falls from 0, the stronger the effect of the predictor \( x_i \) in the sense that the odds ratio falls farther from 1. If \( \beta_i \) is positive, odds ratio is greater than 1 and \( p_i \) (success: \( y=1 \), event) increases. For example, if \( \beta_i = 0.7 \), the odds ratio is approximately 2.01 (because \( e^{0.7} = 2.01 \)). This means that the probability that \( Y \) equals 1 is twice as likely an increase of \( x \) by one unit. So there
is a positive relationship between $x$ and $Y$. An odds ratio of 1.0 indicates there is no relationship between $x$ and $Y$.

In this paper, we consider 4 variables: Speed difference and Density difference between origin $(i, l)$ and target cells $(i, l-1)$, Speed difference and Density difference between origin $(i, l)$ and previous cells $(i+1, l)$.

When an explanatory variable is categorical we use dummy variables to contrast the different categories. For each variable we choose a baseline category (reference value) and then contrast all remaining categories with the base line. If an explanatory variable has $k$ categories, we need $k-1$ dummy variables to investigate all the differences in the categories with respect to the dependent variable. In this paper, density differences are considered as dummy variables, because they have discrete values like zero or one. The reference value is when density difference is zero.

Table 2 estimated the models about the different sets of variables and evaluated the performance. Coefficients show the odd ratio results from logistic regression analysis according to variables. Statistical significance of coefficients is represented with symbol *s. Three *s mean the significant level of 99%, two *s are 95%, and one * is 90%.

Model 1 provides the result analyzed with speed difference between origin $(i, l)$ and target cells $(i, l-1)$. This result shows that there is a negative relationship speed difference ($\Delta v = v_{i,l} - v_{i,l-1}$) and lane-changing, because this coefficient (0.9647) is lower than 1 within significant level of 99%. This means that probability increases if speed of target cell is faster than speed of origin cell. Model 2 is the result analyzed with both the speed difference between origin and target cells and the speed difference between origin and the previous cell. In this model, the coefficient of the case between origin and the previous cells (1.0208) is bigger than 1, and this mean that the speed difference is positive implying which the probability of lane changing is greater. Model 3 is the result analyzed based on the combination of Model 2 and the dummy variables which have the density difference values. In Model 3, density difference’s dummy variables with previous cell have insignificant values. This means that drivers don’t significantly consider the density of previous cell when drivers change the lane. Except for these variables, Model 4 shows reasonable result, but density difference’s dummy variables with target cell have still insignificant values. It is because instances in which the density difference is 4 are rare. Model 5 has the most reasonable variables to explain the discretionary lane changing within significant level of 95%. 
Table 2 Result of logistic regression analysis (odd ratio)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{ij} - v_{ij+1}$</td>
<td>$-0.0477^{***}(.0019)$</td>
<td>$-0.0538^{***}(.0021)$</td>
<td>$-0.0441^{***}(.0030)$</td>
<td>$-0.0526^{***}(.0025)$</td>
<td>$-0.0531^{***}(.0028)$</td>
</tr>
<tr>
<td>$v_{ij} - v_{i+1,j}$</td>
<td>-</td>
<td>$1.0208^{**}.0048$</td>
<td>$1.0231^{**}.0057$</td>
<td>$1.0185^{**}.0029$</td>
<td>$1.0246^{**}.0051$</td>
</tr>
</tbody>
</table>

Dummy variables

<table>
<thead>
<tr>
<th>$d_{i,j} - d_{i,j+1}$</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coef. (std.Err.)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Reference value</td>
<td>Reference value</td>
<td>Reference value</td>
<td>Reference value</td>
<td>Reference value</td>
</tr>
<tr>
<td>$d_{i+1,j} - d_{i+1,j+1}$</td>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Coef. (std.Err.)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Reference value</td>
<td>Reference value</td>
<td>Reference value</td>
<td>Reference value</td>
<td>Reference value</td>
</tr>
</tbody>
</table>

*** <0.01, ** <0.05, *<0.1 (Statistical significance)

Figure 5 interpreted the odd ratios according to speed difference between origin (i, l) and target cells (i, l-1) which was found from table 2. We can calculate odd ratio,

$$P = \frac{e^{\beta x}}{1 + e^{\beta x}} = (e^{\beta})^{x} = 0.9931^x,$$

where $x$ is the speed difference between origin (i, l) and target cells (i, l-1). Compared with odd ratio of lane changing when speed difference = 0, odd ratio decreases as speed difference increases. It shows that the probability of lane changing occurrence increases as the speed difference between origin cell (i, l) and target cell (i, l-1) gets negatively greater.
Figure 6 interpreted the coefficient (odd ratio) of model 5 according to density difference between origin \((i, l)\) and target cells \((i, l-1)\). Compared with odd ratio of lane changing when density difference = 0 (reference value), odd ratio increases significantly as density difference increases. It means that the probability of lane changing occurrence increases as the density difference between origin cell \((i, l)\) and target cell \((i, l-1)\) gets positively greater.

**Figure 6 Change of Odd Ratio according to density difference between origin \((i, l)\) and target cells \((i, l-1)\)**

### 5. Conclusion

This paper evaluates microscopic traffic data to examine the likelihood of discretionary lane-changing as functions of travel conditions in the vicinity. To this end, traffic data regarding lane-changing maneuvers from right to left lanes were extracted as discretionary lane-changing. The extracted data were then aggregated into average speed and density *via* cell-based mesoscopic approach that divides freeway into short segments and lanes (i.e., cell), and aggregates traffic data within each cell. In this way, we can obtain localized macroscopic traffic features from available microscopic data. Descriptive statistics show that both speed difference and density difference significantly influence lane-changing likelihood. To examine how lane-changing probability varies with both speed difference and density difference, therefore, we constructed joint probability distribution of lane-changing as functions of those two variables, speed difference and density difference.

The constructed distribution indicates that drivers tend to change their lanes more frequently to travel faster and to have more space in the target lane. To quantify these relations, logistic regression was applied to the collected data. The outcomes of logistic regression show that the relations are statistically significant and non-linear.

In this paper, we assumed that drivers’ perception distance for lane-changing
decision, expressed as cell, is 40m (131.2ft), which may be the realistic distance under congested traffic conditions. However, it is also likely that drivers may scan different distance when they make lane-changing decisions. Therefore, this assumption needs to be verified in further studies. Although we attempted to extract data that well represent discretionary lane-changing, extracted discretionary lane-changing data include some non-discretionary lane-changing. This is because drivers often have specific, preferred traveled lanes and, therefore, change their lanes to reach those lanes. These should be confirmed in future research.

REFERENCE

14. Toledo, T., H.N. Koutsopoulos, and M.E. Ben-Akiva, *Modeling Integrated Lane-


