LI LI, Ph.D. Candidate[1]
E-mail: limao2401040236@126.com
QING-CHANG LU, Ph.D.[2]
(Corresponding Author)
E-mail: qclu@sjtu.edu.cn
YUN-TAO CHANG, Ph.D.[1]
E-mail: changyuntao@tongji.edu.cn
ZHONG-REN PENG, Ph.D.[2]
E-mail: zpeng@dcp.ufl.edu
[1] Key Laboratory of Road and Traffic Engineering
    of the Ministry of Education
    College of Transportation Engineering, Tongji University,
    No. 4800 Cao'An Highway, Jia-Ding District,
    Shanghai 201804, China
[2] State Key Laboratory of Ocean Engineering, School of
    Naval Architecture, Oceanic & Civil Engineering,
    Shanghai Jiao Tong University, No. 800 Dongchuan Rd.,
    Min-Hang District, Shanghai 200240, China

A COMPOSITIONAL ANALYSIS OF UNBALANCED USAGES OF MULTIPLE LEFT-TURN LANES

ABSTRACT

Lane usage measures distribution of a specific traffic movement across multiple available lanes in a given time. Unbalanced lane usages decrease the capacity of subject segment. This paper took multiple left-turn lanes at signalized intersections as case study, and explored the influences of some factors on the lane usage balance. Lane usages were calculated from field collected lane volumes and the constant-sum constraint among them was explicitly considered in the statistical analysis. Classical and compositional analysis of variance was respectively conducted to identify significant influential factors. By comparing the results of compositional analysis and those of the classical one, the former ones have better interpretability. It was found that left-turn lane usages could be affected by parameter variance of geometric design or traffic control, such as length of turning curve, length of upstream segment, length of signal phase or cycle. These factors could make the lane usages achieve relative balance at different factor levels.

KEY WORDS

lane usage; drivers' lane preferences; double and triple left-turn lanes; classical analysis of variance; compositional analysis of variance;

1. INTRODUCTION

Multiple lanes are often designed on the segments where there is large demand of specific traffic movements, but the capacity is not fully utilized in most cases. An important reason of the capacity loss is unbalanced traffic distribution on the multiple lanes. This phenomenon did not get deserved attention in previous traffic studies, which may reach biased research conclusions or incorrect guidance for engineering practice. For instance, when designing traffic signal timing plan, it is usually assumed in an implicit way that multiple lanes have equal queue lengths in the red phase, and the queues are discharged at identical rate during the green phase. If the assumed condition does not occur which is often seen in real traffic, unbalanced traffic distribution is presented as one lane with no queued vehicles but other lanes still having some. It results in the waste of green time and reduction of intersection capacity. The unbalance degree of traffic distribution on multiple lanes can be measured by lane usage [1-2], at individual level, which indicates the ratio of specific lane being chosen by a group of drivers in a given period. If more drivers prefer one lane, its lane usage is definitely higher than the others. Previous studies had found that many external factors could affect the drivers' lane choices, such as traffic control or geometric design [3-5]. At aggregated level, these factors could also affect lane usages, though this analyzing view has never been taken before, at least not to the authors' knowledge.

The drivers' lane choices are either-or choices, i.e. one driver can only choose one lane at one time and give up the others. If taking the drivers in a given period as a whole, their choices directly result in strict negative correlation existing among the usages of multiple lanes. Unfortunately, this implicit constraint has not been taken into account in previous studies on unbalanced lane usages [1-2]. Yousif et al. made polynomial regression analysis between lane usages and total volume of multiple lanes [1]. In their study, strict negative correlation of lane usages was only held

in the regression model of one lane usage while ignored in other models. Most studies that worked on identifying the influential factors of lane usages had not even noticed this constraint. Instead of individual lane usages, lane utilization, which is defined as ratio of average multilane volume with respect to the highest one [6], is more common to be taken as dependent variable in the statistical analyses of unbalanced multilane traffic distribution [7-8]. Sando and Moses figured out the influences of intersection geometrics on the utilization of triple left-turn lanes [7], but the obtained results could not interpret the trade-off of three individual lane usages. In the authors' opinion, the reason that strict negative correlation of lane usages is hardly considered in previous studies could be attributed to the fact that no proper method had been found to hold this constraint in statistical analysis or result interpretation. This paper intends to fill this gap.

In this study, multiple left-turn lanes (MLTLs) at signalized intersections are taken as case study. Each left-turn lane (LTL) volume is considered as aggregated lane choice results of a group of left-turn drivers, and it is used to calculate LTL usage to measure the probability of each LTL being chosen by drivers. For the arriving vehicles in the red phase and discharging vehicles in the green phase, their speed deviation is limited, so it is assumed that only external factors affect LTL usages. Two regression analyses, i.e. classical and compositional analysis of variances, were conducted to identify the significant influential factors. In order to use common methods to solve compositional regression problem, an approach called "staying in the simplex" [9] is introduced in this study. The remainder is organized as follows. Section 2 introduces the methodology used in this study. Section 3 describes the selected study sites and the work on data collection. Section 4 compares the results of classical and compositional analysis of variance, and provides result interpretations and discussions. Section 5 concludes the main findings of this paper.

2. METHODOLOGY

2.1 Case study and data structure

Two common MLTLs configurations, double left-turn lanes (DLTLs) and triple left-turn lanes (TLTLs), are taken as case studies in this paper. Their typical layouts include: 1) two unshadowed LTLs; 2) one shadowed LTL and one unshadowed LTL; 3) three unshadowed LTLs; 4) one shadowed LTL and two unshadowed LTLs. Lane usage of the i-th LTL, f_{usa-i}, can be calculated as:

$$f_{usa-i} = \frac{V_i}{V} \tag{1}$$

where V_i is volume of the i-th LTL, and V is total volume of the MLTLs. LTL usages are named as f_{usa-in} or $f_{usa-out}$ for inside or outside lane of DLTLs, and f_{usa-in},

$f_{usa-med}$, or $f_{usa-out}$ for inside, median or outside lane of TLTLs. The name of each LTL is defined according to its location to road central line. The closest one is the inside LTL, so the further and farthest ones are the median and outside lanes, respectively.

Besides individual LTL usages, another data structure, i.e. MLTL volume composition, is applied to represent the lane usages in an integrated way. The compositional data, or composition, is a set of data elements, and each element is called a component, whose value reflects the importance of the component within the whole. The components of MLTL volume composition are LTL volumes, and the ratio-scaled components can hold the information of LTL usages and the constraint they are subject to. DLTL and TLTL volume composition, Y_{DLTLs} and Y_{TLTLs}, can be respectively organized as 2-part or 3-part ratio-scaled compositional data [10]:

$$\begin{cases} Y_{DLTLs} = k[VIL, VOL] \\ Y_{TLTLs} = k[VIL, VML, VOL] \end{cases} \tag{2}$$

where k is the total volume of MLTLs; VIL, VML and VOL are the volume components of the inside, median and outside LTLs, and their summation always equals 1, which is called the constant-sum constraint [10]. LTL usages are also subject to the constraint. Under such constraint, if one factor can increase the value of one LTL volume component, it definitely decreases the sum of other components. Hence, it is reasonable to infer that this factor could affect the drivers' lane preferences, owing to the intrinsic connection between lane usage and individual lane choices.

Since the drivers' lane choices could be affected by the design of traffic control and road geometrics at the intersection equipped with MLTLs, some factors are selected to reflect the features of these designs. Since the attributions of the factors are fixed at each site, they should be taken as categorical variables in the statistical analysis. To identify the significant factors, the analysis of variance (ANOVA) is applied in this study to check which factor could affect the mean of MLTL volume composition.

2.2 Classical analysis of variance model

Since the vector of lane usages is naturally multivariate, it seems to be suitable as the dependent variable of a multivariate regression model, which could be used to figure out the reasons of unbalanced lane usages. However, this analyzing method proved to be infeasible in theory as well as in practice, because the vector is a compositional data in essence. Some academics explained why the compositional regression problem cannot be solved by existing methods developed for classical multivariate analysis if the compositional data appears as the independent variable or dependent variable or both [11]. The reasons include:

– Independent components mix together and closed exhibit negative correlations exist among them.
– Covariance between two components depends on other components in the dataset.
– Variance matrices are always singular due to the constant sum constraint.
– Due to bounded range of value, the component cannot be normally distributed in real space.

To verify the infeasibility of classical multivariate analysis in the case of compositional regression analysis, the authors have tried to make multivariate analysis of variance (MANOVA) to LTL usages in virtue of the statistical package STATA [12]. The model estimation has failed and the error message on exact co-linearity of dependent variables was reported. Probably due to the incapability of multivariate regression way to deal with the lane usages, some previous studies formulated the alternative models as follows [1-2]:

$$\begin{cases} f_{usa-i} = \theta_{mi} x_m + \varepsilon_i \, (i = 1, ..., n-1) \\ f_{usa-n} = 1 - \sum_{i=1}^{n-1} f_{usa-i} \end{cases} \tag{3}$$

where x_m is the m-th influential factor of f_{usa-i}; θ_{mi} is its coefficient; n is the lane number of MLTLs; ε_i is the error term. Such model is also formulated in this study, and its estimates will be interpreted and compared with those of the compositional regression model developed in the next section.

2.3 Compositional analysis of variance model

As mentioned in Section 2.1, under the constant-sum constraint, MLTL volume composition can reflect LTL usages in an accurate and integrated way. But the constraint also makes the statistical analysis of the composition hard to conduct until the "stay in the simplex" approach was proposed by some academics. To release the constraint, several types of log-ratio transformations were developed by Aitchison [9]. Then Mateu-Figueras and her colleagues have found that one of the transformations, the isometric logratio transformation (ILR), is not only able to release the constraint, but also to translate the compositional data without changing its angles and distances and obtain its Cartesian coordinates with respect to an orthonormal basis [13]. The translated compositional data can be analysed in the classical way, and similarly, the translated statistical model can be estimated by the classical method as well. This is the "stay in the simplex" approach. Its application in this study can be divided into the following steps:

1) Formulate compositional analysis of variance (CANOVA) model.
2) Define an orthonormal basis and translate the CANOVA model to the version without the constant-sum constraint by conducting ILR transformation.

3) Estimate the transformed model by classical method on defined orthonormal basis.
4) Translate the obtained results not subject to the constant-sum constraint to the ones subject to it by inverse ILR transformation, and interpret them.

Firstly, CANOVA model can be formulated as a classical ANOVA model:

$$Y_i = a \oplus b_{xi} \oplus \varepsilon \tag{4}$$

where Y_i is MLTL volume composition; a is the compositional constant; b_{xi} is the unknown coefficient of the influential factor xi, subject to a compositional constant set $\{b_{x1}, ..., b_{xo}, ..., b_{xp}\}$; p is level number of xi; ε is assumed to be subject to normal distribution with null compositional expectation and a constant variance; $\oplus$ and $\odot$ are Aitchison operations to the compositional data [10], whose functions are similar to summation and multiplication to real vector. At some level of xi, Y_i has a conditional expected value μ_{io}:

$$\mu_{io} = E[Y_i \mid b_{xi} = b_{xo}] = a \oplus b_{xo} \tag{5}$$

But μ_{io} is not unique, because b_{xo} can always be replaced by $b_{xq} = b_{xo} \odot (c \odot a)$ to get the same model, where $\odot$ is also an Aitchison operation that are similar with vector subtraction. Hence, the estimated coefficients are non-identifiable:

$$c \oplus b_{xq} = a \oplus (c \odot a) \oplus b_{xo} \odot (c \odot a) = a \oplus b_{xo} \tag{6}$$

To avoid this problem, *Equation 5* should be reformulated into an equivalent model with identifiable coefficients [11]. It is assumed that xi has no effect on Y_i at its first level. The variance of Y_i can be interpreted as its average response to the upgraded level of xi from its first level. The obtained results will be interpreted in this way. Since similar condition happens in classical ANOVA, this treatment is also applied in its model estimation and result interpretation, as shown in Section 5.1.

Secondly, an orthogonal basis with physical meaning should be defined before the ILR transformation to CANOVA model. Since LTL volume components have physical meanings, they are naturally taken to build the basis by assigning them in a hierarchical way referring sequential binary partition (SBP) method [14]. On the defined basis, the CANOVA model can be transformed as:

$$ILR(Y_i) = ILR(a) + ILR(b_{xi}) + ILR(\varepsilon) \tag{7}$$

After the transformation, CANOVA model becomes a MANOVA model not holding the constant-sum constraint of Y_i, i.e. LTL usages, any more, which is ready to be estimated by classical methods. Finally, since the constant-sum constraint is released in current stage, the estimated coefficients do not correspond to the LTL volume components one by one. They need to be translated by inverse ILR transformation to let the coefficients subjected to the constraint. The

inverse-transformed coefficients can correspond exactly to each component, which will make the result interpretations much easier.

3. DATA COLLECTION AND REDUCTION

Six signalized intersections in downtown Shanghai, China equipped with eight MLTLs, four DLTLs and four TLTLs, were selected as the study sites. Their locations are listed in *Table 1* and shown in *Figure 1a*. The geometrics of two intersections with two MLTLs, (D2, D3) and (D1, T1), are illustrated in *Figures 2a* and *2b*. Upper speed limit of these intersections is 60 km/h, and LTL width is 3.5 metres. The intersections are controlled by fixed signal timing with exclusive left-turn phases. LTL volume was extracted from the traffic videos captured in the field. As illustrated in *Figure 1b*, video camera was deployed on the pedestrian overpass and directed to MLTLs. Since the drivers' lane choices have been found to be affected by signal indicator [4-5], LTL volumes were counted from traffic videos phase by phase, and they were summed up as cycle volume. A heavy vehicle was equivalent to two passenger cars in volume counting. The volume was counted in an untraditional way in this study. One vehicle was counted

in when it arrived to the queue back, rather than when it passed the stop line. If no queue existed, it was still counted at the stop line. This method was derived from a priori knowledge: since the lane volume is the aggregated outcome of the drivers' lane choices, the choices definitely have been made when the subject vehicles enter the queue back. The LTL volumes were organized as 424 DLTL red/green phase compositions (or 212 DLTL cycle compositions) and 336 TLTL red/ green phase compositions (or 168 TLTL cycle compositions). Six study scenarios, SD1/ST1, SD2/ST2, and SD3/ST3, were designed to analyze the variance of LTL usage or MLTL volume composition in the red phase, green phase and the cycle. External influential factors were selected from three aspects: (1) the length of signal phase/cycle; (2) the length of turning curve; (3) drivers' lane changing opportunities on intersection approach and upstream segment. The factor measures are listed in *Table 2* and their values at each site are listed in *Table 1*. The measures of some factors and their values are illustrated in *Figure 2*. Except F5, all factors are ordered categorical variables in the statistical analysis, and their levels are ordered as 1, 2, 3 and 4 in *Table 1*. The division criteria of the factor levels are: (1) The interval between two levels

Table 1 – Summary of study sites

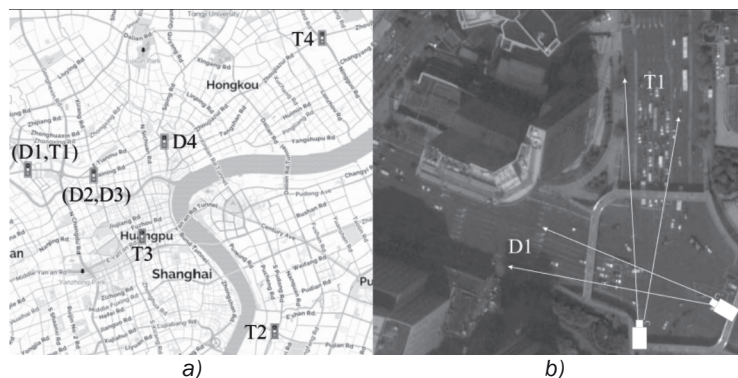| Layout	DLTLs				TLTLs			
Study site location	D1: EB Tianmu Western Rd. at Hengfeng Rd.	D2: WB Haining Rd. at Xizang Northern Rd.	D3: NB Xizang Northern Rd. at Haining Rd.	D4: WB Haining Rd. at Wusong Rd.	T1: SB Hengfeng Rd. at Tianmu Western Rd.	T2: EB Pujian Rd. at Pudong Sourthern Rd.	T3: EB Yan'an Rd. at Shimen First Rd.	T4: WB Zhoujia Zui Rd. at Ningguo Rd.
F1	220s(3)	210s(2)	200s(1)	260s(4)	219s(2)	104s(1)	242s(4)	230s(3)
F2r	190s(2)	193s(2)	164s(1)	205s(3)	183s(2)	76s(1)	194s(3)	184s(2)
F2g	30s(2)	17s(1)	36s(2)	55s(3)	36s(1)	28s(1)	48s(2)	46s(2)
F3	6(2)	4(1)	6(2)	4(1)	6(3)	28s(1)	48s(2)	46s(2)
F4	4(2)	4(2)	3(1)	4(2)	4(2)	3(1)	4(2)	4(2)
F5	Yes	Yes	No	No	Yes	No	No	Yes
F6	158m(2)	107m(1)	123m(1)	289m(3)	142m(1)	172m(2)	195m(3)	300m(4)
F7	95m(2)	77m(1)	108m(3)	200m(4)	94m(1)	157m(2)	180m(3)	200m(3)
F8	4(3)	4(3)	2(1)	3(2)	3(2)	2(1)	2(1)	4(3)

Figure 1 – a) Location of the study sites; b) Video capture method

Table 2 – Influence factors for LTL usages and MLTL volume composition

| Category	No.	Name	Measures
Signal phase /cycle length	F1	Cycle length	Counted in seconds per cycle.
	F2	Phase length	Counted in seconds at red phase (F2r) and green phase (F2g).
Length of turning curve	F3	Length of longitudinal movement	Measured in lane number.
	F4	Length of lateral movement	Measured in lane number.
Lane changing opportunity	F5	Shadowed LTL	Equipped with shadowed LTL or not; 0=Unequipped, 1=Equipped.
	F6	Length of upstream segment	Measured from end of white solid line to upstream segment of MLTLs.
	F7	Distance to the first left-turn sign	Measured from end of white solid line to the first left-turn sign at upstream.
	F8	Number of other lanes	The number of through and right-turn lanes in the same approach of MLTLs.
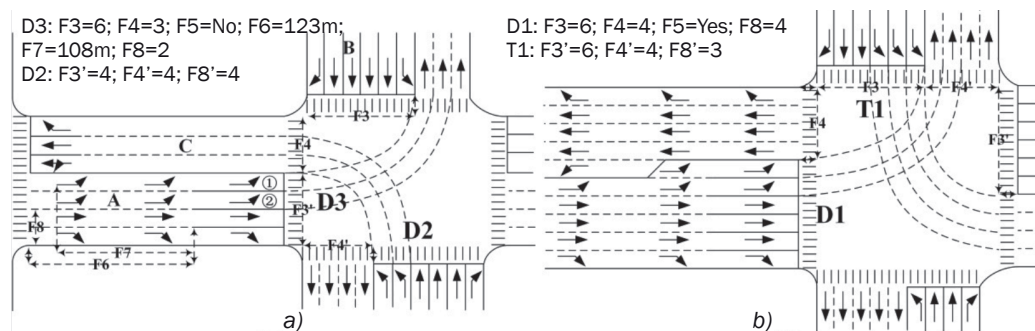
Figure 2 – a) D2 and D3; b) D1 and T1

of F1, F2r and F2g is more than 10 seconds; (2) the levels of F3, F4 and F8 are divided by the lane number; (3) the gap between two levels of F6 and F7 is more than 20 metres.

4. DESCRIPTIVE ANALYSIS

The collected samples are descriptively explored in a compositional view. Sample scatterplots of DLTL and TLTL volume compositions at each scenario are illustrated in the diagrams of *Figures 3* and *4*. Two coordinate axes of the diagrams in *Figure 3* scale inside and outside LTL usages (*VIL* and *VOL* defined in *Equation 2*). All samples scatter along the sloping dot lines, as they cover all possible combinations of *VIL* and *VOL*. All combinations of the components of TLTL volume composition fall into the equilateral triangles (*VIL, VML, VOL*) in *Figure 4*. If one sample is closer to the point *VIL* than *VML* and *VOL*, the usage of inside LTL in this sample is larger than the usages of median and outside LTLs. In other words, more drivers choose the inside LTL. The crossing of dot lines in the figures denotes the balanced status of the MLTL volume composition, i.e. [0.500, 0.500] for DLTL volume composition and [0.333, 0.333, 0.333] for the TLTL one. Two descriptive statistics, central tendency (CT) and metric standard deviation (MSD) [11], were computed. CT measures the mean of each LTL volume component, while MSD measures the average spread of all components to CTy.

As shown in the diagrams, the samples have higher concentration to the balanced status in SD1/ST1 or SD3/ST3 than in SD2/ST2, and the highest concentration seems to be achieved in SD3/ST3. This trend can be also found in *Table 3*. The MSD in SD2/ST2 (0.434/0.487) is larger than that in SD1/ST1 (0.175/0.255) or SD3/ST3 (0.163/0.249), which indicates the green phase compositions spread wider than the red phase or cycle ones. In addition, the distribution of red phase compositions (SD1/ST1) looks similar to that of cycle ones (SD3/ST3). If checking the results of compositional analysis in *Tables 5* and *6*, it can be found that significant factors in SD1/ST1 are similar to the ones in SD3/ST3, but some of them are not significant in SD2/ST2, especially in SD2. This could be attributed to the much shorter green phase length than the red phase length, which makes red phase volume higher than green phase volume. So cycle volumes, as their summation, could have similar amount and composition with the red phase volume. Moreover, the CT of *VIL* in SD1/SD3 (0.479/0.485) is smaller than the CT of *VOL* (0.521/0.515). It indicates

that the inside LTL attracts more drivers than the outside LTL in SD1/SD3, but it is less attractive in SD2. In TLTL scenarios, the median LTL always attracts the most left-turn drivers (0.356/0.377/0.362), and the attractions of the inside and outside LTLs are similar to those in DLTL scenarios.

Table 3 – Descriptive statistics

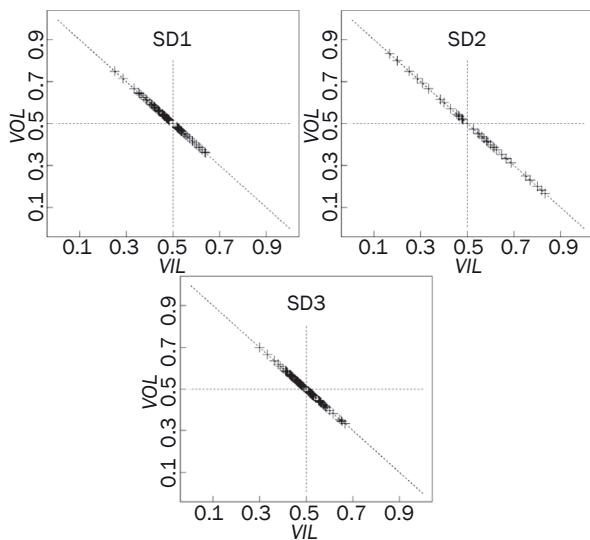| Scenario	CT			MSD
	VIL	*VML*	*VOL*	
SD1	0.479	--	0.521	0.175
SD2	0.516	--	0.484	0.434
SD3	0.485	--	0.515	0.163
ST1	0.331	0.356	0.313	0.255
ST2	0.306	0.377	0.317	0.487
ST3	0.326	0.362	0.312	0.249

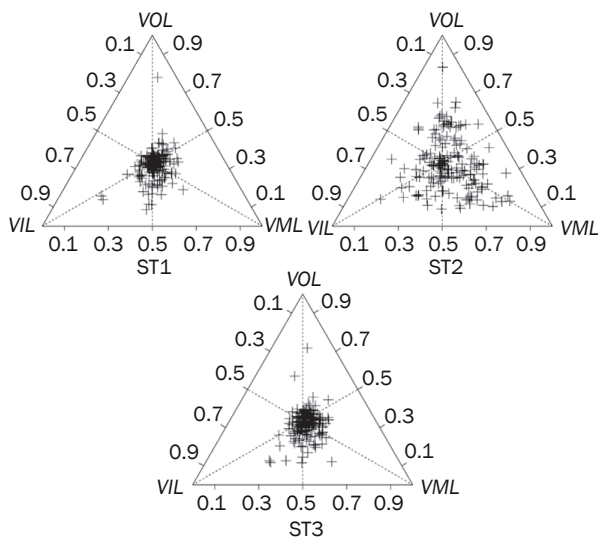*Figure 3 – Sample distributions in SD1/SD2/SD3*

Figure 4 – Sample distributions in ST1/ST2/ST3

5. RESULTS AND DISCUSSION

The classical ANOVA model and CANOVA model developed in Section 2 are to inspect if the mean of LTL usages or MLTL volume composition could be significantly affected by the upgraded influence of some factor. Null hypothesis is the mean of the usage in *Equation 3* or the mean of the composition in *Equation 4* has the same expectation at different factor levels. The hypotheses are tested at 95% confidence level by STATA [12] and R package "compositions" [11].

The significant factors for LTL usages and their estimated coefficients are listed in *Table 4*. The first level of a factor is taken as the basis when estimating its coefficients at the higher level, so the number of each factor coefficient is one less than its levels (bracketed next to the factor code). Here, take F2r for example. This factor has three levels (3L) in both DLTL and TLTL scenarios, but it is only significant in the latter scenario, as shown in *Table 4*. So its coefficients can only be found in $f_{usa-med}$ column of ST1. According to *Equation 3*, it indicates that the median and outside LTL usages are significantly affected by F2r in ST1, but the inside LTL usage is not. The upgrading F2r from its first level to the second level could increase the median LTL usage (0.017) while decreasing the usage of the outside LTL (-0.017). In other words, some drivers originally preferring the outside LTL could choose the median LTL if the red phase length increases to the second level.

Unlike the factor coefficients listed in *Table 4*, *Tables 5* and *6* report the component values of MLTL volume composition at each level of significant factors. The component value is calculated according to the linear relationship defined by the CANOVA model, i.e. adding the component coefficient estimated at an upgraded level of a factor to the component value at the first level of the factor, which is assumed to have no influence on the component. Again, take DLTL volume composition (bolded in *Table 5*) under the influence of F2r, i.e. the length of the red phase, in SD1 for example. Inside LTL volume component (*VIL* column in SD1 of the table) is 0.493 at the first level of F2r. If upgrading F2r to the second or third level, *VIL* decreases to 0.479 or 0.464. It indicates that with the prolonged red phase, the ratio of inside LTL volume with respect to total DLTL volume declines. The decrease of *VIL* is added to *VOL* owing to the sum-constant constraint, which means that more drivers would choose the outside LTL rather than the inside LTL if they have to spend longer time waiting behind the stop line. To observe the component variance clearly, the varying trend of each component is presented by symbol "↗" or "↘" in *Tables 5* and *6*. The symbol "↗" denotes the increase of LTL volume component owing to the upgraded factor, and *vice versa*. In this way, the trade-off among the LTL volume components can be easily observed. The unbalance level of each component is also known by comparing its value at each factor level with the balanced component value.

5.1 Results of classical ANOVA

Some trends are observed from the results of classical ANOVA in *Table 4* if comparing them with that of CANOVA in *Tables 5* and *6*. Some factors impose similar effects on the mean of LTL usages and LTL volume components. For example, the upgraded F4 could result in the decrease of inside LTL usage (-0.019 in f_{usa-in} row of DLTLs in *Table 4*) and VIL (0.474-0.493=-0.019 in *VIL* column of SD1 in *Table 5*). A similar condition also happens to F6 in SD1 (-0.018/0.020 in f_{usa-in} row of DLTLs in *Table 4*; 0.479-0.460=0.019, 0.479-0.499=0.020 in *VIL* column of SD1 in *Table 5*). Under the influence of other factors, such as F4 in SD2 or F8 in SD2, the variance of the usage and volume component of specific LTL are also close to each other.

However, the contradictions between two ANOVA results are much greater than their similarity. An obvious contradiction is that the number of significant factors presents different patterns in the red phase, green phase and cycle. From the results in *Table 4*, it is found that more factors could affect LTL usages in the green phase (SD2/ST2 row) than in the red phase (SD1/ST1 row) or cycle (SD3/ST3 row), while the condition is inversed for MLTL volume composition, as shown in *Tables 5* and *6*. Since the same dataset is used in the two ANOVAs, this contradiction indicates that the two methods have methodological difference.

Moreover, many factors only have significant effects on the median and outside LTL usages, but the inside LTL usage is totally unaffected, as shown in *Table 4*. The most extreme example is that no factor significantly affects the inside LTL usage in ST1 (f_{usa-in} row in *Table 4*). This result is not convincible for the authors from the lane choosing drivers' view, because it indicates that although many factors could affect the median and outside LTL usages, the proportion of the drivers choosing the inside LTL rather than others is quite steady in a given number of drivers, in other words, no driver leaves or enters the lane. This is unimaginable in real traffic. Such inconvincible result do not appear in *Tables 5* and *6*. Upgraded factor influence could be imposed on all LTL volume components, but the impact degree on each component is different. It means that the variance of one factor could induce some vehicles to switch to other LTLs in different probabilities, rather than all of them switching to one lane. Obviously, these results are more reasonable than those of the classical ANOVA. The detailed interpretations of the CANOVA results can be referred to Section 5.2.

In the authors' opinion, the aforementioned contradictions derive from different roles of the constant-sum constraint in the two ANOVAs. As shown in *Equation 3*, the constraint is only applied in the classical ANOVA model of the last LTL usage after the rest n-1 models are developed. Such formulation implies that the

Table 4 – ANOVA results for the factor influences on LTL usage

			F1 (4L/4L)	F2r (3L/3L)	F2g (3L/2L)	F3 (2L/3L)	F4 (2L/2L)	F5 (2L/2L)	F6 (3L/4L)	F7 (4L/3L)	F8 (3L/3L)
SD1/ST1	DLTLs	f_{usa-in}	-	-	-	-	-0.019	-	-0.018/ 0.020	-	-
	TLTLs	f_{usa-in}	-	-	-	-	-	-	-	-	-
		$f_{usa-med}$	-	0.017/ -0.006	-	0.019/ 0.019	0.011	0.021	-0.013/ -0.019/ 0.011	-	-
SD2/ST2	DLTLs	f_{usa-in}	-	-	-0.028/ 0.007	-0.031	0.048	0.020	0.018/ 0.031	-0.006/ -0.048/ 0.006	-
	TLTLs	f_{usa-in}	-	-	-0.028	-	-0.027	-	-	0.011/ -0.024	-
		$f_{usa-med}$	-	-	0.027	0.031/ -0.004	0.006	0.016	0.012/ 0.004/ 0.057	0.012/ 0.032	-0.008/ 0.049
SD3/ST3	DLTLs	f_{usa-in}	-	-	-	-0.014	-0.004	-	-	-	-
	TLTLs	f_{usa-in}	-	-	-	-0.008/ -0.034	-	-	-	-	-
		$f_{usa-med}$	0.009/ 0.030/ -0.007	-	-	0.023/ 0.017	0.011	0.021	0.009/ -0.017/ 0.021	-0.009/ 0.003	0.013/ 0.033

variances of the n-1 LTL usages are independent, yet they do not comply with the fact. But the constraint is explicitly considered in the formulation of the CANOVA model, and it can be held in significant test and coefficient estimation. This characteristic of the CANOVA makes the trade-off among LTL volume components and can be attributed to the upgraded factor influences. Hence, it is reasonable to believe that the results of CANOVA are more convincible than those of classical ANOVA. Further in the text, the CANOVA results will be mainly interpreted and discussed.

5.2 Results of compositional ANOVA

5.2.1 Signal phase/cycle length

The influence of signal indicator on MLTL volume composition has been considered in the scenario design. As shown in *Tables 5* and *6*, the significant factors in SD1/ST1 are different than those in SD2/ST2. At individual level, it indicates that the drivers' lane preferences could vary with the switch of signal

phases. Some factors are significant in both red and green phases, but their influence is presented in an opposite way in the two periods, which will be detailed later. Besides the influence of signal light switch, the influences of cycle length and red phase length (F1 and F2r) on the MLTL volume composition are reported in the tables. These two signal timing factors determine the amount of LTL volume as well as affect the LTL usages, i.e. the probability of each LTL being chosen by the drivers. The longer phase/cycle gives the drivers more time to make their lane choices. Since the green phase length (F2g) does not show significant influence, so it is omitted from the tables.

Cycle length (F1)

F1 has four levels in both DLTL and TLTL scenarios. In SD3, the increase of the cycle length to the second or third level could reduce the ratio of inside LTL volume to total DLTL volume (0.486>0.475>0.468 in *VIL* column of SD3 in *Table 5*). But this trend could be reversed at the fourth level of F1 (0.468<0.512, ↘↘↗ in *VIL* column of SD3) when the composition achieves

Table 5 – DLTL volume composition under the influence of significant factors

Factor	VIL	VOL
SD1		
F1(4L)	-	-
F2r(3L)	0.493/0.479/0.464‖↘↘	0.507/0.521/0.535‖↗↗
F3(2L)	-	-
F4(2L)	0.493/0.474‖↘	0.507/0.526‖↗
F5	0.496/0.462‖↘	0.504/0.538‖↗
F6(3L)	0.479/0.460/0.499‖↘↗	0.521/0.560/0.501‖↘↗
F7(4L)	0.464/0.460/0.493/0.499‖↘↗↗	0.536/0.540/0.507/0.501‖↘↗↗
F8(3L)	0.493/0.499/0.462‖↘↗	0.507/0.501/0.538‖↘↗
SD2		
F1(4L)	-	-
F2r(3L)	-	-
F3(2L)	-	-
F4(2L)	0.478/0.531‖↗	0.522/0.469‖↘
F5	-	-
F6(3L)	-	-
F7(4L)	-	-
F8(3L)	0.478/0.538/0.527‖↗↘	0.522/0.462/0.473‖↘↗
SD3		
F1(4L)	0.486/0.475/0.468/0.512‖↘↘↗	0.514/0.525/0.532/0.488‖↗↗↘
F2r(3L)	-	-
F3(2L)	0.492/0.478‖↘	0.508/0.522‖↗
F4(2L)	-	-
F5	0.498/0.471‖↘	0.502/0.529‖↗
F6(3L)	0.481/0.468/0.512‖↘↗	0.519/0.532/0.488‖↗↘
F7(4L)	0.475/0.468/0.487/0.512‖↘↘↗	0.525/0.532/0.513/0.488‖↗↘↘
F8(3L)	0.487/0.512/0.471‖↗↘	0.513/0.488/0.529‖↘↗

relative balance ([0.512, 0.488] at the fourth level of F1). This reversal could be attributed to the highest green ratio along with the longest cycle length (F2g/F1=55s/260s at D4 in *Table 1*) among all study sites. Although the green phase length (F2g) does not affect the DLTL volume composition, more vehicles arriving to D4 during the longest green time (F2g=55s at D4) could release the unbalance of the red phase composition in SD1. This unbalance improving function of the green time ratio can also be found in ST3. The highest green time ratio along with the longest cycle (F2g/F1=48s/242s at T3 in *Table 1*) could also bring TLTL volume composition close to balance ([0.341, 0.343, 0.316] at the fourth level of F1 in ST3 in *Table 6*). So proper adjustment of the green time ratio in signal timing could improve unbalanced traffic distribution on MLTLs.

Red phase length (F2r)

F2r has three levels in SD1. The increase of the red phase length could gradually reduce the ratio of the inside LTL volume with respect to the total DLTL volume (0.5>0.493>0.479>0.464 in *VIL* column in SD1 of *Table 5*). Correspondingly, the outside LTL could attract more left-turn drivers (0.5<0.507<0.521<0.535 in *VOL* column of SD1). In addition, F2r has three levels in ST1. Opposite to SD1, the upgraded F2r plays

the balanced role for TLTL volume composition, as the composition could be closer to the balanced status at the fourth level of F2r ([0.334, 0.338, 0.327] in ST1). In other words, the increased red phase length could make the drivers choose three LTLs of similar probabilities. Hence, F2r could have the opposite effect on drivers' lane preferences in DLTLs and TLTLs.

5.2.2 Length of vehicle turning curve

The turning curve of the left-turn vehicles at intersection is presented as a partly edge of the circle or ellipse. Its length is hard to estimate accurately for the left-turn drivers in real traffic, but it can be estimated in separate steps. The left-turning movement can be divided into longitudinal and lateral movements. The former one is to pass the entry approach of crossing road, the approach "B" in *Figure 2a*, while the latter one is to pass the exit approach in opposite direction of the same segment, approach "C" in the figure. In this paper, the two moving distances are measured by the lane number, as this measurement is more reasonable for the on-task drivers. Take the drivers on approach "A" in *Figure 2a*, for example. The length of longitudinal movement (F3) is the width of "B" as 2 lanes, which can be estimated when the drivers get enough sight of "B". The length of lateral movement (F4) is the width

Table 6 – TLTL volume composition under the influence of significant factors

	VIL	VML	VOL
ST1			
F1(4L)	--	--	--
F2r(3L)	0.409/0.302/0.334 ‖↘↗	0.346/0.363/0.338 ‖↗↘	0.245/0.335/0.327 ‖↗↘
F3(3L)	0.334/0.347/0.308 ‖↗↘	0.338/0.362/0.357 ‖↗↘	0.327/0.291/0.335 ‖↘↗
F4(2L)	0.409/0.311 ‖↘	0.346/0.356 ‖↗	0.245/0.333 ‖↗
F5	0.371/0.302 ‖↘	0.344/0.363 ‖↗	0.285/0.335 ‖↗
F6(4L)	0.308/0.409/0.335/0.295 ‖↗↘↘	0.357/0.346/0.338/0.371 ‖↘↘↗	0.335/0.245/0.327/0.334 ‖↘↗↗
F7(3L)	0.308/0.409/0.313 ‖↗↘	0.357/0.346/0.356 ‖↘↗	0.335/0.245/0.331
F8(3L)	0.371/0.308/0.295 ‖↘↘	0.344/0.357/0.371 ‖↗↗	0.285/0.335/0.334 ‖↗↘
ST2			
F1(4L)	--	--	--
F2r(3L)	--	--	--
F3(3L)	0.349/0.278/0.317 ‖↘↗	0.359/0.400/0.357 ‖↗↘	0.292/0.322/0.326 ‖↗↘
F4(2L)	--	--	--
F5	0.339/0.282 ‖↘	0.356/0.385 ‖↗	0.305/0.333 ‖↗
F6(4L)	0.317/0.329/0.349/0.237 ‖↗↗↘	0.357/0.372/0.359/0.423 ‖↗↘↗	0.326/0.299/0.292/0.341 ‖↘↘↗
F7(3L)	--	--	--
F8(3L)	0.339/0.317/0.237 ‖↘↘	0.366/0.357/0.423 ‖↘↗	0.295/0.326/0.340 ‖↗↗
ST3			
F1(4L)	0.398/0.306/0.279/0.341 ‖↘↘↗	0.354/0.360/0.382/0.343 ‖↗↗↘	0.248/0.334/0.339/0.316 ‖↗↗↘
F2r(3L)	--	--	--
F3(3L)	0.331/0.333/0.306 ‖↗↘	0.343/0.372/0.360 ‖↗↘	0.307/0.295/0.334 ‖↘↗
F4(2L)	0.398/0.307 ‖↘	0.354/0.362 ‖↗	0.248/0.331 ‖↗
F5	0.369/0.305 ‖↘	0.350/0.359 ‖↗	0.281/0.336 ‖↗
F6(4L)	0.306/0.398/0.341/0.279 ‖↗↘↘	0.360/0.354/0.343/0.382 ‖↘↘↗	0.334/0.248/0.317/0.340 ‖↘↗↗
F7(3L)	0.306/0.398/0.307 ‖↗↘	0.360/0.354/0.364 ‖↘↗	0.334/0.248/0.329 ‖↘↗
F8(3L)	0.369/0.306/0.279 ‖↘↘	0.350/0.360/0.382 ‖↗↗	0.281/0.334/0.339 ‖↗↗

of "C" as 2 lanes, which is easily observed when the drivers approach the intersection.

Length of longitudinal movement (F3)

F3 has two levels in the DLTL scenarios, and it is significant in SD3. The increased length of longitudinal movement could induce DLTL volume composition out of balance (0.5>0.492>0.478 in *VIL* column, 0.5<0.508<0.522, in *VOL* column of SD3 of *Table 5*). In addition, F3 has three levels in the TLTL scenarios. Its negative influence on the balance of TLTL volume composition is found in ST1 and ST3. The composition achieves relative balance at the first level of F3 ([0.334, 0.338, 0.327] in ST1, [0.331, 0.343, 0.307] in ST3 of *Table 6*), yet it is away from the balance at the higher levels. On the contrary, the relative balance of the composition is achieved at the highest level of F3 in the green phase ([0.317, 0.357, 0.326] at the second level of F3 in ST2). So, the shorter length of longitudinal movement could bring fluctuation to TLTL volume composition in the green phase but lead to balance in the red phase.

Length of lateral movement (F4)

F4 has two levels in both DLTL and TLTL scenarios. Increased lateral movement length could aggravate the unbalance of DLTL volume composition in both red phase (0.5>0.493>0.474 in *VIL* column, 0.5<0.507<0.526 in *VOL* column in SD1 of *Table 5*) and in green phase (0.478<0.5<0.531 in *VIL* column, 0.522>0.5>0.469 in *VOL* column in SD2). Opposite to the influence of F3, F4 could make TLTL volume composition close to relative balance at its second level in the red phase or cycle period ([0.311, 0.356, 0.333] in ST1, [0.307, 0.362, 0.331] in ST3 of *Table 6*). With the influences of F3 and F4 in mind, providing symmetrical distances of lateral and longitudinal movements to left-turn drivers (4 lanes at the first level of F3, 4 lanes at the second level of F4 in the TLTL design of *Table 1*) could be helpful to maintain the balance of TLTL volume composition.

The opposite influence of F4 in SD1/SD2 (or F3 in ST1/ST2) is noteworthy. Increased length of lateral movement in the DLTL scenarios could induce more drivers to choose the outside LTL in the red phase (0.526>0.474 at the second level of F4 in SD1 of *Table 5*). But in the green phase their preferences switch to the inside LTL (0.531>0.469 at the second level of F4 in SD2). This could be owing to the larger turning radius of the outside LTL, which could make the drivers obtain a faster acceleration when signal light turns from red (SD1) to green (SD2). If the drivers arrive at the intersection in the green phase (SD2), they could want to pass through as soon as possible, so the smaller turning radius of the inside LTL becomes more favourable. In addition, another example of phase-dependent factor influence is that of F3 in the TLTL scenarios. In the red phase, *VIL* increases

if F3 upgrades from the first level to the second one, then it decreases at the third level (↗↘ in *VIL* column in ST1 of *Table 6*). But in the green phase, this trend is reversed (↘↗ in *VIL* column in ST2). Under its influence, the fluctuation of *VOL* follows the opposite way to that of *VIL* (↘↗ in *VOL* column in ST1; ↗↘ i in *VOL* column in ST2). The opposite influences of F3 and F4 in different phases could be attributed to their roles changed with signal phase switch. In the red phase (SD1/ST1), the turning movement is infeasible for the drivers. They have to wait behind the stop line until the red phase ends. But in the green phase (SD2/ST2), the turning movement can be finished in a short time, so the drivers' attitudes about the turning curve would be changed. In the red phase F3 and F4 are desired distances for the drivers, while the distances become implementable for the drivers in the green phase.

5.2.3 Drivers' lane changing opportunities

The drivers can choose their target lanes with lane changing or not. The appearance of lane changing behaviour depends on its necessity as well as whether there is enough space to conduct it. These influences on the drivers' lane choices are captured by the factors F5-F8. Shadowed LTL (F5) can provide the drivers additional lane changing opportunities near the stop line. Only the inside LTL is designed as shadowed lane if necessary in the study sites. Besides, the longer upstream segment (F6) would provide the drivers with more opportunities to pre-adjust to their target LTLs. Similarly, the first left-turn sign (F7) set at the upstream segment of MLTL could inform the drivers of its existence ahead, so they can make the lane choices in advance. The number of through and right-turn lanes in the same approach of MLTLs could be also related to the lane changing opportunities of the left-turn drivers. With the fewer through and right-turn lanes, the drivers could enter the LTLs with fewer lane changings or even directly enter the LTLs when entering the approach.

Shadowed LTL (F5)

In *Tables 5* and *6*, the values of LTL volume components at the first level of F5 refer to the ones without shadowed LTL influence. F5 is significant in nearly all scenarios except SD2. Relative balance of DLTL volume composition could be achieved in SD1 and SD3 when the inside LTL is unshadowed ([0.496, 0.504] in SD1, [0.498, 0.502] in SD3 at the first level of F5 in *Table 5*). Shadowed LTL could result in unbalanced lane traffic on the DLTLs (0.462<0.538 in SD1, 0.471<0.529 in SD3 at the second level of F5). At individual level, it demonstrates that there are more drivers who do not want to use the inside shadowed lane of the DLTLs at the cost of a lane changing from outside lane. Such repulsion of the shadowed inside lane is found in all TLTLs scenarios while leads to different consequences.

In the green phase, the shadowed LTL could impair the balance of TLTLs volume composition ([0.339, 0.356, 0.305] at the first level of F5 in ST2 of *Table 6*). However, the unshadowed median and outside LTLs could be more attractive for drivers than the shadowed inside one in ST1 and ST3, and relative balance of the composition could be achieved at the second level of F5 ([0.302, 0.363, 0.335] in ST1, [0.305, 0.359, 0.336] in ST3). Finally, it is interesting to find that the median LTL of TLTLs benefits the most from the existence of the shadowed LTL. There is more outflow of the inside LTL that switches to the median LTL than the outside LTL. It indicates that when the drivers are unwilling to change to the inside LTL, they do not like the outside LTL either, which makes the median LTL naturally the best choice.

Length of upstream segment (F6) and Distance to the first left-turn sign (F7)

Both F6 and F7 have significant influence on MLTL volume composition. F6 could make DLTL volume composition achieve relative balance at its highest level ([0.499, 0.501] in SD1, [0.512, 0.488] in SD3 at the third level of F6 in *Table 5*). As result, the composition balance could benefit from the longer upstream segment. This trend is also found in the TLTL scenarios. Relative balance of TLTL volume composition is achieved at the third level of F6 ([0.335, 0.338, 0.327] in ST1, [0.341, 0.343, 0.317] in ST3 of *Table 6*). But if the length of upstream segment is overlong, it could impair the composition balance ([0.295, 0.371, 0.334] in ST1, [0.279, 0.382, 0.340] at the fourth level of F6 in ST3). A similar condition is also found in ST2, when relative balance of the composition is achieved at the first level of F6 ([0.317, 0.357, 0.326] in ST2). The overlong upstream segment could weaken the drivers' confidence to pass the intersection ahead and induce them to choose the lanes accommodating their other demands. F7 has similar impact on MLTL volume composition as F6 does. The compositions could achieve relative balance at the highest level of F7 ([0.499, 0.501] in SD1, [0.512, 0.488] at the fourth level of F7 in SD3 of *Table 5*; [0.313, 0.356, 0.331] in ST1, [0.307, 0.364, 0.329] at the third level of F7 in ST3 of *Table 6*). In practice, if it is impossible to extend the upstream segment, setting the left-turn signs closer upstream of the intersection of MLTL could be a good idea to balance the composition.

Number of other lanes (F8)

F8 has three levels in both DLTL and TLTL scenarios, and it is significant in almost all scenarios except in SD2. Relative balance of MLTL volume composition is achieved at the second level of F8 ([0.499, 0.501] in SD1, [0.512, 0.488] in SD3 of *Table 5*; [0.308, 0.357, 0.335] in ST1, [0.317, 0.357, 0.326] in ST2, [0.306, 0.360, 0.339] in ST3 of *Table 6*). It seems that keeping the number of through and right-turn lanes at a moderate

level could make the composition easy to achieve relative balance.

The last but not the least important finding is that giving the drivers more lane choosing freedoms could change insignificant factors to be significant for the composition balance. F5-F7 are insignificant in SD2 but turn to be significant in ST2. This could be attributed to the one more LTL of TLTLs than DLTLs, which enlarges the drivers' lane choice set. Another example is that the number of significant factors in SD1 is much larger than that in SD2. Since drivers have to stop in the red phase, the external influential factors are more likely to be considered by the drivers. In the green phase most drivers focus on passing the intersections as soon as possible, so they may care less about external influences. Maybe this leads to the insignificance of many factors in the green phase. The two examples indicate that if the drivers have more alternative lanes or time to make lane choice, their lane preferences could be easy to present.

6. CONCLUSION

The presented work intends to figure out the reasons of unbalanced lane usages on MLTLs. LTL volumes are collected at some DLTLs and TLTLs. Two statistical methods, i.e. classical ANOVA and compositional ANOVA, have been introduced to identify significant influential factors of LTL usages. By comparing the results of two analyses, it is found that ignorance of the implicit constant-sum constraint among the LTL usages in classical ANOVA could lead to biased results. The CANOVA is a more convincible way to analyse unbalanced multilane usages. Based on this finding, the results of CANOVA have been interpreted and discussed in detail. Some suggestions could be proposed to improve unbalanced traffic distribution on the MLTLs in practice:

- Prolong cycle length along with increased green time ratio;
- Provide relative symmetrical lateral and longitudinal moving lengths for drivers;
- Rebuild shadowed LTL as unshadowed LTL if there is enough space in intersection;
- Extend the length of upstream segment of MLTLs;
- Set traffic sign or pavement marking of MLTLs close to upstream intersection;
- Deploy the proper number of through and right-turn lanes in the same approach of MLTLs.

It is noteworthy that the influences of some factors could be opposite in different signal phases or MLTL layout. Customized countermeasures should be carefully designed to improve unbalanced LTL traffic. In addition, since unbalanced lane usages are often seen at other multilane roadway segments, it is believed that the compositional analysing methods

introduced in this paper could also be applied to other applications.

ACKNOWLEDGMENTS

李立，博士研究生[1]
电子邮件：limao2401040236@126.com
路庆昌，博士（通讯作者）[2]
电子邮件：qclu@sjtu.edu.cn
常云涛，博士[1]
电子邮件：changyuntao@tongji.edu.cn
彭仲仁，博士[2]
电子邮件：zrpeng@sjtu.edu.cn
[1] 同济大学交通运输工程学院道路与交通工程教育部
　重点实验室
　中国上海市嘉定区曹安公路4800号（201804）
[2] 上海交通大学船舶海洋与建筑工程学院海洋工程国
　家重点实验室
　中国上海市闵行区东川路800号（200240）

多左转车道非平衡利用率的成分分析研究

摘要

车道利用率反映了一定时间内某一特定方向交通流在多条可用车道上的分布状况。非均衡的多车道利用率会降低多车道路段的通行能力。本文以信号控制交叉口多左转车道为研究对象，确定了对左转车道利用率平衡性有影响的因素。基于现场采集的车道流量数据计算出各车道利用率，并在统计分析中考虑了多车道利用率所满足的定和约束。经典方差分析和成分方差分析分别用于确定对车道利用率具有显著性影响的因素。通过比较分析结果发现，成分方差分析所得结论更加可信。研究发现，左转车道利用率可能受到几何设计或交通控制参数变化的影响，如转弯半径长度、上游路段长度和信号周期及相位时长。这些因素可使多左转车道利用率在其不同影响等级下达到相对平衡的状态。

关键词

车道利用率；驾驶员车道偏好；双左转及三左转车道；经典方差分析；成分方差分析

REFERENCES

[1] Yousif S, Al-Obaedi J, Henson R. Drivers' lane utilization for United Kingdom motorways. Journal of Transportation Engineering. 2013;139:441-447.

[2] Golias J, Tsamboulas R. Macrolevel estimation of highway lane usage. Journal of Transportation Engineering. 1995;121(1):40-49.

[3] Dubedi A, Chakroborty P, Kundu D, Reddy KH. Modeling automobile drivers' toll-lane choice behavior at a toll plaza. Journal of Transportation Engineering. 2012;138(11):1350-1357.

[4] Bugg Z, Rouphail NM, Schroeder B. Guidance for simulation-based modeling of auxiliary through lanes. Transportation Research Record: Journal of the Transportation Research Board. 2012;2311:51-58.

[5] Bugg Z, Rouphail NM, Schroeder B. Lane choice model for signalized intersections with an auxiliary through lane. Journal of Transportation Engineering. 2013;139(4):371-378.

[6] Transportation Research Board. Highway capacity manual. Washington, D.C.: Transportation Research Board; 2010.

[7] Sando T, Moses R. Influence of intersection geometrics on the operation of triple left-turn lanes. Journal of Transportation Engineering. 2009;135(5):253-259.

[8] Sando T, Mussa R. Site characteristics affecting operation of triple left-turn lanes. Transportation Research Record: Journal of the Transportation Research Board. 2003;1852:55-62.

[9] Glahn VP. Statistical modelling on coordinates. Paper presented at: Compositional Data Analysis Workshop—CoDaWork'03; 2003 Oct 15-19; Catalonia, Spain.

[10] Aitchison J. The statistical analysis of compositional data. London: Chapman & Hall; 1986. (Reprinted in 2003 with additional material by The Blackburn Press).

[11] Boogaart KGVD, Tolosana-Delgado R. Analyzing compositional data with R. Berlin Heidelberg: Springer-Verlag; 2013.

[12] StataCorp. Stata: release 14. College Station: StataCorp; 2015.

[13] Mateu-Figueras G, Pawlowsky-Glahn V, Egozcue J. The principle of working on coordinates. Chichester, UK: John Wiley & Sons, Ltd.; 2011.

[14] Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R. Modeling and analysis of compositional data. Chichester, UK: John Wiley & Sons, Ltd.; 2015.