CHENGCHENG XU, Ph.D.[1, 2]
E-mail: xuchengcheng@seu.edu.cn
CHEN WANG, Ph.D.[1]
(Corresponding author)
E-mail: wkobec@hotmail.com
WEI WANG, Ph.D.[1, 2]
E-mail: wangwei@seu.edu.cn
JIE BAO, Ph.D. Candidate[1, 2]
E-mail: baojie@seu.edu.cn
MENGLIN YANG, Ph.D. Candidate[1, 2]
E-mail: 230149536@seu.edu.cn
[1] Jiangsu Key Laboratory of Urban ITS, Southeast University Si Pai Lou #2, Nanjing, 210096, China
[2] Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Si Pai Lou #2, Nanjing, 210096, China

Traffic and Space
Original Scientific Paper
Submitted: 26 Apr. 2016
Accepted: 8 Feb. 2017

INVESTIGATING SPATIAL INTERDEPENDENCE IN E-BIKE CHOICE USING SPATIALLY AUTOREGRESSIVE MODEL

ABSTRACT

Increased attention has been given to promoting e-bike usage in recent years. However, the research gap still exists in understanding the effects of spatial interdependence on e-bike choice. This study investigated how spatial interdependence affected the e-bike choice. The Moran's I statistic test showed that spatial interdependence exists in e-bike choice at aggregated level. Bayesian spatial autoregressive logistic analyses were then used to investigate the spatial interdependence at individual level. Separate models were developed for commuting and non-commuting trips. The factors affecting e-bike choice are different between commuting and non-commuting trips. Spatial interdependence exists at both origin and destination sides of commuting and non-commuting trips. Travellers are more likely to choose e-bikes if their neighbours at the trip origin and destination also travel by e-bikes. And the magnitude of this spatial interdependence is different across various traffic analysis zones. The results suggest that, without considering spatial interdependence, the traditional methods may have biased estimation results and make systematic forecasting errors.

KEY WORDS

e-bike; spatial autocorrelation; spatially autoregressive regression; random-parameter regression; survey data;

1. INTRODUCTION

World is facing a great challenge in environmental problems resulting from the increasing vehicle use and rapid urbanization. Vehicle emissions are considered the main source of urban air pollution. To reduce the climate impacts of urban vehicle use, increased attention has been given to promoting the usage of environmentally friendly and sustainable traffic modes, such as electric bikes (e-bikes) [1]. With the power assistance of electric motors and integrated batteries, e-bikes address the limits of trip distance and terrain associated with the traditional man-powered bicycles. It is expected to increase bicycle use in urban areas.

To develop effective policies and strategies for promoting e-bike usage, numerous studies have been conducted to investigate the characteristics of e-bike trips and mode choice behaviour [2-9]. E-bike choice models have been developed to link the likelihood of e-bike trips with various contributory factors, including socio-demographic characteristics, physically built environment, attitudinal factors, and trip features. Cherry and Cervero developed an e-bike choice model based on survey data collected in Kunming and Shanghai cities in China. The ownership of bikes, travel time, age and gender are the main contributing factors to the choice of e-bikes [2]. Sylvia et al. investigated the relationship between the e-bike choice and trip characteristics based on surveyed travel diary data for two weeks. The attitude towards environment, the trip characteristics, and the socio-demographic characteristics affects the choice of e-bike in a trip.

A number of studies have also been conducted to investigate the effects of e-bike use on the mode share of other travel modes [10-13]. Fyhri and Fearnley explored the potential effects of e-bike use in reducing motorized trips. Based on a case-control study, they found that e-bike use decreases the percentage of motorized trips by about 20%, and that this effect increases with an increase in the usage time of e-bikes [13]. In a study conducted by Montgomery, an origin–destination survey of e-bikes was conducted to explore the competition between e-bikes and bus rapid transit in China. The results showed that e-bike is a highly attractive alternative to bus rapid transit (BRT), and travellers may shift from BRT to e-bikes.

While numerous studies have been conducted to understand the factors that influence the e-bike choice, most studies did not consider the effects of spatial interdependence. Such spatial interdependence indicates that the probabilities of choosing e-bikes of nearby travellers are correlated. In other words, nearby travellers may have similar preferences in choosing e-bikes than more distant travellers. Such spatial interdependence is caused by the unobserved latent factors such as public transport service, road characteristics, traffic control conditions, etc. The spatial autoregressive structure in statistical regression models can be used to capture the spatial interdependence in e-bike choice. The existing e-bike mode choice models were generally developed by the discrete choice model without incorporating spatial autoregressive structure. The conventional modelling technique may lead to bias and invalid parameter estimates, if, in fact, the spatial interdependence affects the mode choice [14-15]. The policies and strategies based on the biased model may not effectively promote the e-bike use. This study fills the gaps in understanding the spatial interdependence in e-bike choice.

Recently, increased interest has been shown to investigate the spatial interdependence and its impacts on travel behaviour. This spatial interdependence has been increasingly found in various travel behaviour studies, such as public transit use [14], automobile ownership [15], and travel time loss [16]. Although considering spatial interdependence in travel behaviour studies is not completely new, relatively few studies investigated the spatial interdependence in e-bike choice. Moreover, previous studies generally assumed that the effects of spatial interdependence were the same for different persons. The spatially autoregressive mode choice model with varying spatial interdependence among different persons has not been implemented.

The aim of this study is to detect the spatial interdependence in e-bike choice and to evaluate its impact on e-bike choice. This study has the potential to contribute to the field of e-bike mode choice modelling by: (1) investigating whether the spatial dependence exists in both commuting and non-commuting e-bike trips, and how spatial dependence affects the traveller's choice of e-bikes; (2) providing a relatively new spatial autoregressive model to allow spatial interdependence to vary among different travellers. This new model is expected to better capture spatial interdependence and improve model estimation accuracy.

2. DATA SOURCES

Data were collected from an extensive household travel survey conducted in Shaoxing City in 2007. Shaoxing is a typical medium-sized city on the east coast of China. In 2007 the population was 908,500 and the total area covered 59.96 km2. The city area was divided into 25 traffic analysis zones (TAZs). The household survey was conducted by the local government on one typical weekday to make transportation system planning. Each member in the randomly selected household was asked to complete a questionnaire. A total of valid 7,320 questionnaires were used.

The survey contained two parts: (1) individual and household characteristics; and (2) travel information of all trips in the whole day. The travel information included the geocodes and time for all trip origins and destinations, allowing the estimations of detailed travel time, origins and destinations of all trips. The purpose and transportation mode were also recorded for each trip. The travel purpose included work, shopping, education, home, entertainment, sports, medical care, and official business. Finally, the ODs between different origins and destinations, as well as the population density, trip production, and trip attraction data for each TAZ were obtained from the transportation system planning documents of the Shaoxing City.

The individual and household characteristics information included gender, occupation, number of household members, number of automobiles in the household, age, and annual household income. In the survey, age was divided into eight categories, including 0 to 14, 15 to 19, 20 to 24, 25 to 29, 30 to 39, 40 to 49, 50 to 59, and older than 60. The annual household income was divided into five categories, lower than RMB 10,000, 10,000 to 20,000, 20,000 to 50,000, 50,000 to 100,000, and higher than 100,000. The data about the intention to buy an automobile was also collected in the survey. The mode choice model for commuting trips was developed using the trips from home to work, and the model for non-commuting trips was developed using the first non-commuting trip in the surveyed day.

3. METHODOLOGY

The unobserved latent factors such as location amenities, road characteristics and traffic control conditions would lead to spatial interdependence of e-bike choice. Such spatial interdependence indicates that nearby travellers may have similar preference in choosing e-bikes than more distant travellers. To evaluate whether spatial interdependence plays a role in e-bike choice, the following statistical methods were used. The Moran's I statistic was used to test whether e-bike choice is spatially correlated at the aggregate level. The Bayesian spatial autoregressive logistic regression was used to detect whether e-bike choice is spatially correlated at disaggregate level, and to evaluate the impact of spatial interdependence on the e-bike choice for commuting and non-commuting trips.

3.1 Spatial correlations tests

The Moran's I is one of the most commonly used statistics to measure the spatial dependence. It was used to identify whether the proportions of e-bike trips in various TAZs are spatially correlated. The Moran's I statistic for an unstandardized spatial weight matrix C takes the following forms [17]:

$$Moran's\, I = \frac{N}{\sum_i \sum_j c_{ij}} \frac{\sum_i \sum_j c_{ij}(E_i - \overline{E})(E_j - \overline{E})}{\sum_i (E_i - \overline{E})} \tag{1}$$

where N is the number of spatial units (i.e. TAZ); E_i and E_j is the proportion of e-bike trips in the i-th and j-th TAZ; $\overline{E}$ represents the average proportion of e-bike trips across different TAZs; c_{ij} denotes an element of an unstandardized spatial weight matrix C, which measures the connection between areas i and j; c_{ij} equals 1 if TAZ i and j are neighbours, and 0 otherwise. The range of Moran's I statistic is between -1 and +1. Higher positive value indicates greater degrees of spatial interdependence while negative value indicates spatial dispersion. A value near zero indicates a spatially random pattern.

The statistical significance of the Moran's I index is usually tested by the Z-scores. The null hypothesis is that the proportion of e-bike trips is spatially independent in the study area. The z-score of Moran's I can be calculated by:

$$Z_I = \frac{I - E(I)}{SD(I)} \tag{2}$$

where $E(I)$ and $SD(I)$ are expectation and standard deviation of Moran's I. A positive Z_I score denotes that the neighbouring TAZs tend to have similar proportions of e-bike trips, whereas a negative Z score indicates that the proportions of e-bike trips tend to be more dissimilar among neighbouring TAZs.

3.2 Bayesian spatially autoregressive logistic regression

The Bayesian spatial autoregressive logistic regressions were used to detect the spatial interdependence in e-bike choice at individual level. In this model, a spatially autoregressive term was included to account for the spatial correlations of e-bike choices at disaggregate level. Assuming that $Y=\{y_m\}$ is a vector of binary observed e-bike choice indicators (m=1, 2, ..., M). If e-bike was chosen for a trip, $y_m=1$; otherwise, $y_m=0$. With regard to each y_m, $x_m=[x_{1,m}, x_{2,m}, ..., x_{K,m}]'$ is correspondingly a 1×K vector of explanatory variables with coefficient vector $\beta=[\beta_1, \beta_2, ..., \beta_K]$. The Bayesian spatial autoregressive logistic regression is written as:

$$z_m = \beta_0 + \beta x_m + \theta_m + \varepsilon_m, \varepsilon_m \overset{iid}{\sim} logistic(0,1)$$
$$y_m = \begin{cases} 0, \text{if } z_m \le 0 \\ 1, \text{if } z_m > 0 \end{cases} \tag{3}$$

where z_m is the latent variable for the logistic regression; ε_m denotes the error term and is assumed to be logistically distributed with the probability density function given by $f(\varepsilon)=e^{-\varepsilon}/(1+e^{-\varepsilon})^2$.

θ_m captures the correlations of e-bike choices among different travellers. It is assumed that travellers are more likely to choose e-bikes if their neighbours also travel by e-bikes. To define the neighbourhood for each traveller, an $M \times M$ neighbourhood matrix $C=(c_{ij})$ is defined as: $c_{ij}=1$ if individual i and j is located in the same TAZ and $c_{ij}=0$, otherwise θ_m takes the following form:

$$\theta_m = \rho \sum_{m=1}^{M} w_{ij} y_m, m = 1, ..., M \tag{4}$$

where w_{ij} is an element of spatial weight matrix W with $M \times M$ dimension; In addition, it is assumed that row sums $\Sigma_j w_{ij}$ are normalized to one as follows:

$$W = [w_{ij}] = \left[\frac{c_{ij}}{\sum_{j=1}^{M} c_{ij}} \right] \tag{5}$$

Equation 4 can be summarized in vector form as: $\theta=\rho WY$. The product term WY represents the spatially weighted average mode share of e-bike over the travellers neighbouring each individual m; parameter ρ is taken to describe the effect of spatial interdependence on e-bike choice. Although the model specified in *Equation 3* does not immediately present the spatially autoregressive structure, it can be observed when converting y_m into a function of the latent variable of z_m as $y_m=f(z_m)$. *Equation 3* is then given as follows:

$$Z = \beta_0 + \beta_x + \rho W f(Z) + \varepsilon$$
$$y_m = \begin{cases} 0, \text{if } z_m \le 0 \\ 1, \text{if } z_m > 0 \end{cases} \tag{6}$$

Following suggestions in previous studies [14-15], the spatial interdependence in e-bike choice is determined as an exogenous process. This can help to improve model estimation efficiency. In previous studies, spatial interdependence effect is assumed to be the same across different samples. However, spatial interdependence should differ across different samples [14-15]. To overcome these associations, parameter ρ in *Equation 6* was allowed to be different across various TAZs. Random parameter ρ is assumed to be normally distributed as $\rho \sim (\mu_\rho, \Sigma_\rho)$.

Based on the above specifications, full data likelihood of Bayesian spatially autoregressive logistic regression model is given as:

$$
\begin{aligned}
f(\mathrm{Y}|\Theta) &= \prod_{m=1}^{M} f(y_m|\beta_0,\beta,\rho) = \\
&= \prod_{m=1}^{M}\left[\frac{e^{\beta_0+\beta x_m+\rho_r \mathrm{W}_m \mathrm{Y}}}{1+e^{\beta_0+\beta x_m+\rho_r \mathrm{W}_m \mathrm{Y}}}\right]^{y_m}\cdot \\
&\cdot\left[\frac{1}{1+e^{\beta_0+\beta x_m+\rho_r \mathrm{W}_m \mathrm{Y}}}\right]^{1-y_m}
\end{aligned}
\tag{7}
$$

where Θ represents the vector of all parameters, including constant β_0, variables coefficient β, random parameter vector ρ, variance Σ_ρ for ρ and mean μ_ρ for ρ. Accordingly, $\Theta= [\beta_0, \beta, \rho, \mu_\rho, \Sigma_\rho]$. A Markov Chain Monte Carlo (MCMC) simulation-based Bayesian approach is used to estimate the posterior distribution of model parameters Θ. The non-informative prior distributions for all parameters Θ is specified as:

$$
\begin{aligned}
&\beta_0 \sim Normal(\overline{\mu}_0,\overline{\Sigma}_0),\beta \sim Normal(\overline{\mu},\overline{\Sigma}),\\
&\rho_r \sim Normal(\mu_\rho,\Sigma_\rho), r=1,2,...,R\\
&\mu_\rho \sim Normal(\overline{a}_\rho,\overline{b}_\rho),\\
&\Sigma_\rho \sim Inverse\ gamma(\overline{c}_\rho,\overline{d}_\rho)
\end{aligned}
\tag{8}
$$

where all the priors of fixed parameters vector, and the mean of random parameters ρ follow normal distributions. The variance of random parameter ρ was assumed to follow an inverse gamma distribution [18-19]. The non-informative prior distribution was used for each parameter. Accordingly, the hyper-parameters with over lines in *Equation 8* were set as:

$$
\begin{aligned}
&\overline{\mu}_0=0,\overline{\Sigma}_0=10^6,\overline{\mu}=0_k,\overline{\Sigma}=10^6 \mathrm{I}_k,\\
&\overline{a}_\rho=0,\overline{b}_\rho=10^6,\overline{c}_\rho=0.001,\overline{d}_\rho=0.001
\end{aligned}
\tag{9}
$$

To assess the effects of variables on e-bike choice preference, elasticity is computed as:

$$
E_k=\frac{\partial Y_k}{\partial x_k}\cdot\frac{x_k}{Y_k}=[1-P(i)]\beta_k x_k \tag{10}
$$

where β_k represents the estimated parameter associated with the k-th variable x_k. Elasticity for a continuous independent variable represents the percentage change in the dependent variable resulting from a 1% change in an independent variable [20]. Each sample in the dataset has an elasticity that depends on the value of x_i and the estimated probability of choosing e-bike; it is usually to report the average elasticity in the sample. Note that *Equation 10* cannot be used to calculate the elasticity of an indicator variable. The elasticity of an indicator variable x_i is computed as a pseudo-elasticity (see *Equation 11*) [20]. It can be explained as the percent change in the probability of choosing e-bike when the indicator variable is changed from 0 to 1.

$$
E_i=\left[\frac{EXP[\Delta(\chi'\beta)][1+EXP(\chi_i\beta_i)]}{EXP[\Delta(\chi'\beta)][1+EXP(\chi_i\beta_i)]+1}-1\right]\cdot 100 \tag{11}
$$

3.3 Bayes factors analysis

Bayesian comparison of two competing models M_1 and M_2 can be performed by the ratio of the models' posterior probabilities which is defined as Bayes factors [19]. Assuming that we have equal preferences of these two models, the prior probability of model m_1 is equal to model m_2. In this case, the ratio of the models' posterior probabilities can be expressed as:

$$
\begin{aligned}
PO_{21}&=\frac{f(M_2|Y)}{f(M_1|Y)}=\frac{f(M_2|Y)/f(Y)}{f(M_1|Y)/f(Y)}=\\
&=\frac{f(Y|M_2)\pi(M_2)}{f(Y|M_1)\pi(M_1)}=\frac{f(Y|M_2)}{f(Y|M_1)}=B_{21}
\end{aligned}
\tag{12}
$$

where PO_{21} is termed the posterior model odds of model M_2 versus model M_1; $f(M|Y)$ is the posterior probability for model M; $f(Y|M)$ is the marginal likelihood of the data under model M; B_{21} is the Bayes factors of model M_2 versus model M_1. The previous study in Bayesian statistics suggested that significant difference exists between two models if the log of Bayes factor for these two models is larger than 3 [19].

4. DATA ANALYSIS AND RESULTS

4.1 Results of spatial correlations tests

The Moran's I statistic was calculated to investigate whether the e-bike choice is spatially correlated at an aggregated TAZ level. *Table 1* gives the Moran's I statistics and test results. The e-bike choice percentages at both origin and destination for commuting and non-commuting trips are all positively spatially correlated with a p-value lower than 5%. The test results indicate the presence of considerable spatial interdependencies in e-bike choice at both trip origin and destination. Accordingly, spatial interdependencies should be incorporated in the e-bike choice models for both commuting and non-commuting trips.

Table 1 – Test of Moran's I for spatial correlations

Characteristics	Moran's I	Z Score	p-value
The percentage of e-bike at commuting trip origin	0.197	2.382	0.017
The percentage of e-bike at commuting trip destination	0.402	4.135	<0.001
The percentage of e-bike at non-commuting trip origin	0.851	9.321	<0.001
The percentage of e-bike at non-commuting trip destination	0.224	2.478	0.013

Table 2 – Summary of candidate variables

Variables		Min	Max	Mean	SD [a]
Trip duration (minutes)		2	90	19.21	12.674
Gender	Male	-	-	0.52	- [b]
	Female	-	-	0.48	-
Occupation	Student	-	-	0.16	-
	Worker	-	-	0.11	-
	Public officials	-	-	0.29	-
	Other occupations	-	-	0.44	-
Traveller's age	Traveller's age lower than 20	-	-	0.14	-
	Traveller's age between 20 and 40	-	-	0.34	-
	Traveller's age between 40 and 50	-	-	0.37	-
	Traveller's age above 50	-	-	0.15	-
Annual household income	Higher than RMB 20,000	-	-	0.78	-
	Lower than RMB 20,000	-	-	0.22	-
Intention of buying an automobile	Buying automobile in the next five years	-	-	0.087	-
	Others	-	-	0.913	-
Population density of origin	Higher than 0.015 persons/m^2	-	-	0.522	-
	Lower than 0.015 persons/m^2	-	-	0.478	-
Residents at destination	Higher than 0.015 persons/m^2	-	-	0.523	-
	Lower than 0.015 persons/m^2	-	-	0.477	-
Origin–destination volumes (×1,000/day)		0.123	33.150	5.075	6.896

Note: [a] standard deviation; [b] not applicable

4.2 Estimation results of Bayesian spatial autoregressive logistic regressions

To investigate whether spatial interdependence exists in e-bike choice at the disaggregated level, the Bayesian spatial autoregressive logistic regressions were developed for commuting and non-commuting trips separately. The probability of e-bike choice was linked with travel time, gender, occupation, age, household characteristics, and traffic volume. A spatial autoregressive term was included in each model to capture the potential spatial interdependence in e-bike choice (see *Equation 4*). The descriptive statistics of the initially considered explanatory variables for model developments are given in *Table 2*.

Three different models were developed for both commuting and non-commuting trips. More specifically, the first model was developed by the traditional logistic regression approach, in which spatial autoregressive term was not included. This model represents the conventional technique of e-bike choice modelling. The spatial autoregressive term was included in the following two models. In the e-bike choice model with fixed spatial interdependence, the parameter of the spatial autoregressive term was fixed across different observations in the dataset. While in the model with random spatial interdependence, the parameter of the spatial autoregressive term was allowed to be different across various TAZs.

Spatially Autoregressive Model for Commuting Trips

The results of the e-bike choice models for commuting trips are given in *Table 3*. The Bayes factor was calculated using *Equation 12* to compare the fitness of the model without spatial interdependence and the model with fixed spatial interdependence for commuting trips. The logarithms of the marginal likelihoods of these two models are -6,651.336 and -6,625.851, respectively. The logarithm of the Bayes factor of these two models is 25.485, indicating that inclusion of the spatial interdependence significantly increases the fitness of the e-bike choice model for commuting trips. Similarly, the Bayes factor was also used to compare the fitness of the model with fixed spatial interdependence and the model with random spatial interdependence. The marginal likelihood of the random spatial interdependence model is -6,614.740. Accordingly, their Bayes factor is 10.111, indicating that the model with random spatial interdependence has better fitness than the model with fixed spatial interdependence.

As indicated by the negative regression parameter (see *Table 3*), females are less likely to choose e-bikes for commuting travels. According to the estimated pseudo-elasticity presented in *Table 5*, the likelihood of choosing e-bikes by females is about 7.315% lower than the likelihood by male. The parameter of public officials is positive, indicating that the public officials

Table 3 – E-bike choice models with and without spatial interdependencies for commuting trips

Variables		Without spatial interdependence		Without fixed spatial interdependence		With random spatial interdependence	
		Parameter	95% CI [a]	Parameter	95% CI	Parameter	95% CI
Constant		-3.707	(-4.062, -3.314)	-5.950	(-6.504, -5.502)	-5.777	(-6.173, -5.308)
Trip duration (minutes)		-0.006	(-0.010,-0.002)	-0.007	(-0.010, -0.003)	-0.007	(-0.010,-0.003)
Gender	Female	-0.132	(-0.214,-0.048)	-0.143	(-0.227, -0.061)	-0.142	(-0.228,-0.059)
Occupation	Student	-	-	- [b]	-	-	-
	Worker	-	-	-	-	-	-
	Public official	0.300	(0.214,0.386)	0.306	(0.221,0.396)	0.313	(0.225,0.405)
Traveller's age	Lower than 20	0.764	(0.421,1.084)	0.689	(0.392,1.002	0.733	(0.407,1.047)
	Between 20 and 40	2.492	(2.206, 2.759)	2.403	(2.174,2.662)	2.483	(2.210, 2.740)
	Between 40 and 50	1.950	(1.669, 2.215)	1.863	(1.637,2.123)	1.928	(1.656,2.183)
Annual household income	Higher than RMB 20,000	0.206	(0.150,0.263)	0.190	(0.138,0.239)	0.201	(0.137,0.256)
Intention of buying an automobile	Buying in the next 5 years	0.108	(0.076,0.139)	0.108	(0.076,0.140)	0.113	(0.076,0.145)
Origin–destination volumes		-0.033	(-0.040,-0.0267)	-0.039	(-0.046,-0.032)	-0.039	(-0.046,-0.032)
Spatial interdependence at origin				3.794	(2.836, 4.831)	3.486	(3.148, 3.752)
S.D. of spatial interdependence at origin						0.694	(0.429, 1.178)
Spatial interdependence at destination				4.080	(3.012,5.295)	3.350	(2.740,3.848)
S.D. of spatial interdependence at destination						0.886	(0.622,1.171)
Logarithms of the marginal likelihood		-6651.336		-6625.851		-6614.740	

Note: [a] 95% confidence interval of parameter estimates, [b] this variable was not significant in the model

are more likely to travel by e-bikes, compared with travellers of other occupations. The average pseudo-elasticity of 14.576 suggests that the likelihood of choosing e-bikes by public officials is about 14.576% higher than the likelihood by travellers of other occupations. The parameter estimates of the traveller's age suggest that 20- to 50-year-old travellers are more likely to use e-bikes for commuting trips than travellers older than 50. The travellers of age between 20 and 40 have the largest probability of choosing e-bikes.

With regard to trip characteristics, travel time and origin–destination volumes significantly affect the probability of choosing e-bikes. More specifically, the estimated parameter of trip duration is negative, indicating that travellers are less likely to choose e-bikes for a commuting trip with longer travel time. Travellers would like to use motorized trip mode, such as the public transit and passenger car, for longer trips. The average elasticity of -0.095 suggests that one-percent increase in travel time is associated with -0.095% decrease in the probability of choosing e-bikes. The parameter of origin–destination traffic volume is negative, indicating that the probability of choosing e-bike decreases as traffic volume increases. Possible explanation is that traffic congestion may reduce travellers' preference to travel by e-bikes. Increasing bicycle volume decreases cyclists' perceptions of comfort. Moreover, heavy vehicle traffic volume is also negatively related to cyclists' comfort for on-street bicycle facilities because of the increased risks of collision [22-23]. This result is consistent with the finding from previous studies [22-23].

Interestingly, the parameter of intention of buying an automobile is positive, indicating that travellers with intention to buy an automobile are currently more likely to travel by e-bikes. This may suggest that e-bike is just intermediate mode for these travellers before buying an automobile, and they are likely to shift from e-bikes to passenger cars [4]. Therefore, appropriate policy initiatives should be implemented to prevent the transition from e-bike to passenger car. As expected, the annual household income is positively related to the probability of taking trips by e-bikes. This explains the positive parameter of the annual household income.

Both spatial interdependence at trip origin and destination produce statistically significant random parameters (see *Table 3*). The spatial interdependence at trip origin results in a random parameter that is normally distributed, with a mean of 3.486 and a variance of 0.694. Given these distributional parameters, the spatial interdependence at the trip origin is always positive with varying magnitude across different TAZs. This indicates that travellers are more likely to choose e-bikes if their neighbours at the same trip origin also travel by e-bikes. The probability of choosing e-bikes increases with an increase in the spatially weighted average mode share of e-bikes at trip origin. And the magnitude of such spatial interdependence is different across various trip origins.

The parameter of the spatial interdependence at trip destination is also normally distributed, with a mean of 3.350 and a variance of 0.886. It indicates that the positive spatial interdependence also exists on the destination side of the trips. Travellers are also more likely to choose e-bikes if their neighbours at the same trip destination also travel by e-bikes. The possible explanation for the spatial interdependence at trip destination is that travellers choose trip mode based on both origin characteristics and destination characteristics [14]. Like the spatial interdependence at trip origin, the magnitude of such spatial interdependence is also different across various trip destinations.

Spatially Autoregressive Model for Non-commuting Trips

Table 4 gives the estimation results of the Bayesian spatial autoregressive logistic regression models for non-commuting trips. The contributing factors to

Table 4 – E-bike choice models with and without spatial interdependencies for non-commuting trips

Variables		Without spatial interdependence		Without fixed spatial interdependence		With random spatial interdependence	
		Parameter	95% CI [a]	Parameter	95% CI	Parameter	95% CI
Constant		-3.162	(-3.536, -2.778)	-4.656	(-5.173,-4.142)	-4.853	(-5.464,-4.308)
Shopping		-0.449	(-0.653,-0.261)	-0.456	(-0.653,-0.261)	-0.494	(-0.740,-0.273)
Trip duration (minutes)		-0.019	(-0.026,-0.012)	-0.019	(-0.027,-0.012)	-0.022	(-0.031,-0.013)
Occupation	Student	-0.485	(-0.897,-0.052)	*-0.473*	(-0.895,-0.063)	-0.487	(-0.980,-0.014)
	Worker	-	-	- [b]	-	-	-
	Public official	-	-	-	-	-	-
Traveller's age	Lower than 20	1.176	(0.641,1.687)	1.031	(0.519,1.544)	1.080	(0.511,1.688)
	Between 20 and 40	2.347	(2.047, 2.346)	2.293	(2.018,2.583)	2.528	(2.142, 3.001)
	Between 40 and 50	1.499	(1.208,1.795)	1.411	(1.146,1.702)	1.543	(1.211, 1.912)
Annual household income	Higher than RMB 20,000	0.241	(0.151,0.331)	0.237	(0.141,0.333)	0.264	(0.164,0.376)
Origin–destination volumes		-0.026	(-0.039,0.014)	-0.031	(-0.044,-0.019)	-0.036	(-0.053,-0.022)
Spatial interdependence at origin				5.746	(4.208, 7.256)	6.060	(4.632, 7.373)
S.D. of spatial interdependence at origin						0.224	(0.082, 0.464)
Spatial interdependence at destination				4.590	(2.664, 6.547)	3.620	(2.611, 4.988)
S.D. of spatial interdependence at destination						0.296	(0.067, 0.636)
Logarithms of the marginal likelihood		-1888.912		-1865.869		-1862.307	

Note: [a] 95% confidence interval of parameter estimates, [b] this variable was not significant in the model e-bike choice are different between commuting and non-commuting trips. Some variables are significant in the model for commuting trips but are not significant in the model for non-commuting trips, and their impacts are also quite different. The Bayes factor analyses also indicate that including spatial interdependence significantly increases the model fitness for non-commuting trips, and that the model with random spatial interdependence has better fitness than the model with fixed spatial interdependence.

As shown in *Table 4*, the parameter of student is negative, with an average pseudo- elasticity of -24.956 (see *Table 5*), indicating that the likelihood of choosing e-bikes by students for non-commuting trips is 24.956% lower than the likelihood of travellers of other occupations. Regarding traveller's age, the model estimation results suggest that travellers older than 50 are less likely to travel by e-bikes for non-commuting trips, and travellers of the age between 20 and 40 have the largest probability of choosing e-bikes for non-commuting trips.

As expected, an increase in annual household income was found to increase the probability of choosing e-bikes for non-commuting trips. With regard to trip characteristics, shopping, travel time and origin–destination volumes are the main contributing factors to the choice of e-bikes for non-commuting trips. The results are similar to those in the model for commuting trips. Specifically, the increasing travel time and traffic volume reduce the probability of choosing e-bikes. The negative parameter of shopping indicates that travellers are less likely to travel by e-bikes for shopping. The pseudo-elasticity of -25.484 suggests that the likelihood of choosing e-bikes for shopping is 25.484% lower than the likelihood for other travel purposes.

Finally, the spatial interdependence at trip origin and destination also significantly affect the probability of choosing e-bikes for non-commuting trips, and result in statistically significant random parameters. The parameter for the spatial interdependence at trip origin follows a normal distribution with a mean of 6.060 and a variance of 0.224. The spatial interdependence at trip destination results in a random parameter that is normally distributed, with a mean of 3.620 and a variance of 0.296. Therefore, the spatial interdependence also exists in the e-bike choice for non-commuting trips. The average elasticities of the spatial interdependence at trip origin and destination are 0.754 and 0.456, indicating that one-percent increase in spatial interdependence is associated with the 0.754%, and 0.456% increases in the likelihood of choosing e-bikes for non-commuting trips, respectively.

Effects of spatial interdependence on E-bike choice prediction

To evaluate the effects of spatial interdependence on predicted e-bike choice probability, the estimation results of the model without spatial interdependence and the model with random spatial interdependence were compared. The parameters of explanatory variables are quite similar between these two models. However, the constant of the model without spatial interdependence is much greater than the constant of the model with spatial interdependence. This result indicates that the model without spatial interdependence may provide biased forecasting of e-bike choice probability.

Table 5 – Elasticity analyses for different variables

Variables		Commuting trip		Non-commuting trip	
		Mean	S.D.[a]	Mean	S.D.
Trip duration (minutes)		-0.095	0.071	-0.363	0.288
Shopping		-[b]	-	-25.484	1.930
Gender	Female	-7.315	0.232	-	-
Occupation	Student	-	-	-24.956	1.806
	Worker	-	-	-	-
	Public official	14.576	1.295	-	-
Traveller's age	Lower than 20	33.695	4.330	42.384	12.411
	Between 20 and 40	53.823	37.711	70.374	30.597
	Between 40 and 50	49.491	30.615	51.161	21.861
Annual household income	Higher than RMB 20,000	11.294	0.721	11.721	0.848
	Lower than RMB 20,000 [c]				
Intention of buying an automobile	Buying in the next 5 years	35.893	14.489	-	-
Origin–destination volumes		-0.143	0.223	-0.174	0.223
Mean of spatial interdependence at origin		0.735	0.171	0.754	0.241
Mean of spatial interdependence at destination		0.708	0.163	0.456	0.124

Note: [a] Standard deviation of elasticity; [b] This variable was not significant in the model, [c] Reference level

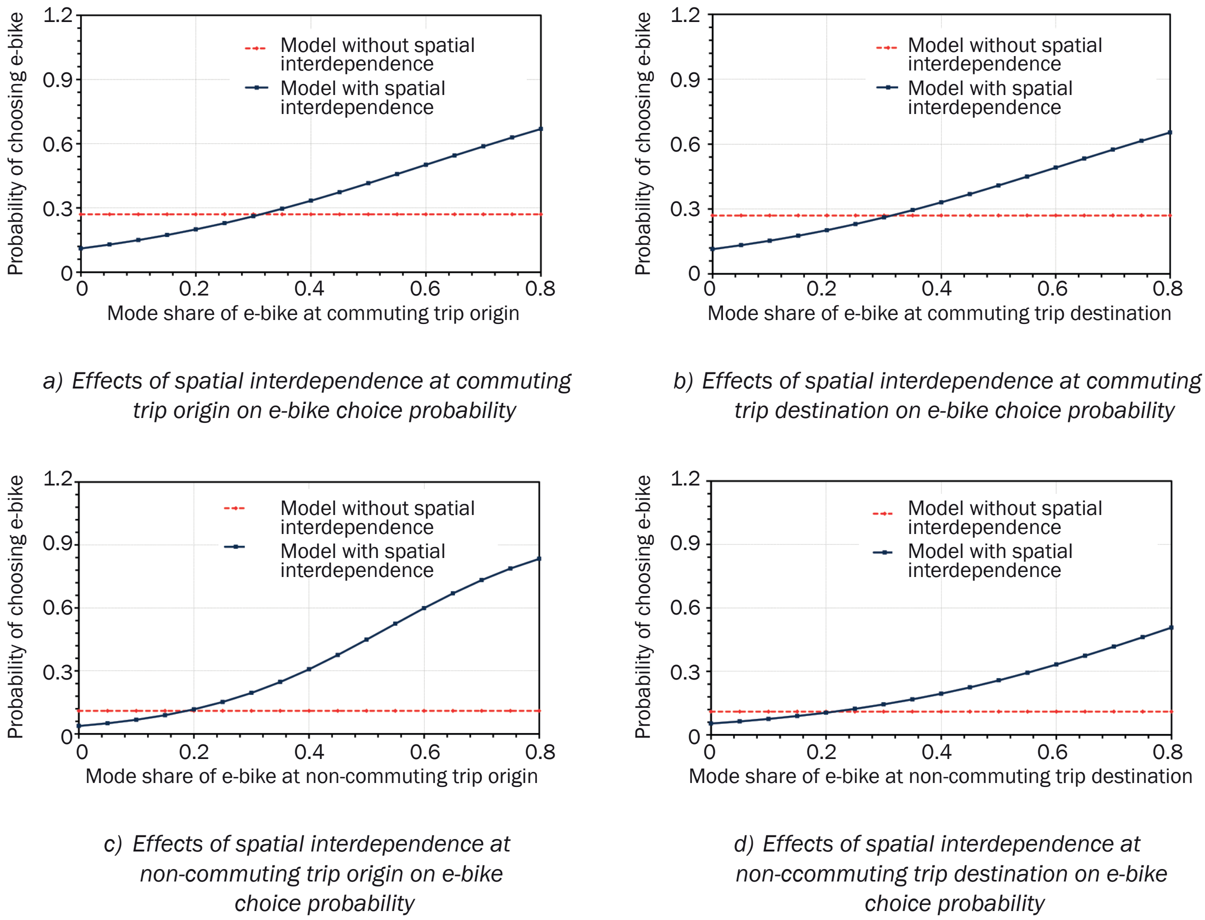
Figure 1a and *1b* were used to illustrate more clearly this point. *Figure 1a* compares the predicted e-bike choice probabilities with different mode shares of e-bikes at trip origin. Note that the sample means of the explanatory variables were used to calculate the probability in *Figure 1a* and *1b*. As shown in *Figure 1a*, the model without spatial interdependence provides an overestimated e-bike choice probability when the mode share of e-bikes at commuting trip origin is low; while it provides an underestimated e-bike choice probability when the mode share of e-bikes at commuting trip origin is high. This systematic prediction error also exists in *Figures 1b*, *1c* and *1d*. The unobserved factors such as public transport service, road characteristics and traffic conditions lead to spatial interdependence that the e-bike choice probabilities of nearby travellers are correlated. The conventional e-bike choice model generally did not account for such spatial interdependence. Accordingly, the model without spatial interdependence will have biased parameter estimates and systematic forecasting errors.

Attention should be paid to such systematic prediction errors. The traditional e-bike mode choice model generally did not consider the spatial interdependence. It is usually used to predict the mode share of e-bikes in travel demand forecasting in traffic planning. However, the above analyses clearly illustrate that without considering spatial interdependence, the traditional e-bike choice model will produce biased mode share for travel demand forecasting.

4.3 Predictive performance

The receiver operating characteristic (ROC) curves were further used to test the predictive performance of the developed spatial autoregressive models for commuting and non-commuting trips [21]. The ROC curves illustrate the relationship between sensitivity and 1 - specificity. The sensitivity is usually called the true positive rate, which measures the proportion of the cases of choosing e-bikes that are correctly identified. And 1- specificity is usually called the false alarm rate, which represents the proportion of the cases of not choosing e-bikes that are mistakenly identified as the cases of choosing e-bikes.

The ROC curves of the models with random spatial interdependence for commuting and non-commuting trips are given in *Figure 2*. The areas under the ROC curve (AUCs) for these two models are 0.771 and 0.794, respectively, indicating that the e-bike choice



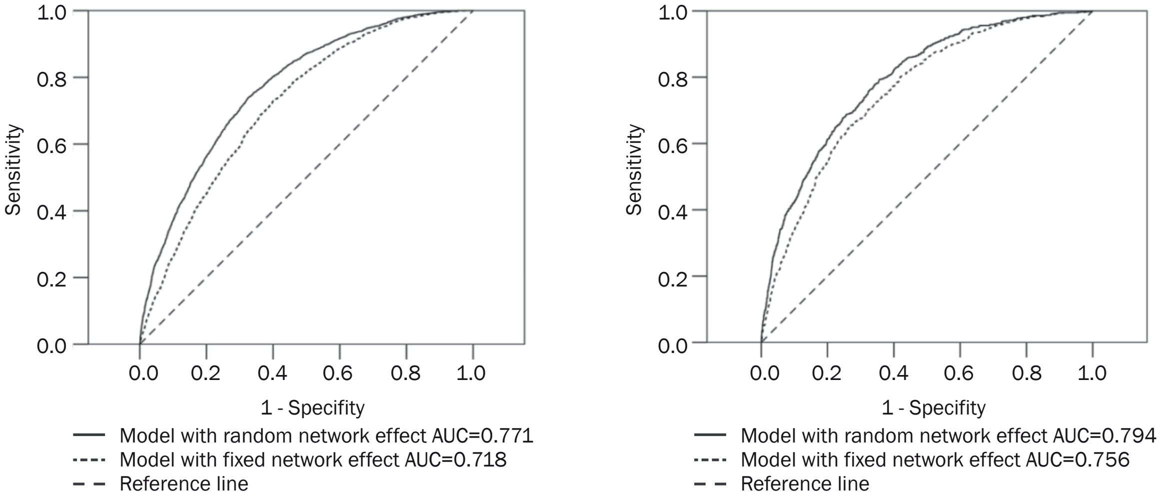
a) Effects of spatial interdependence at commuting trip origin on e-bike choice probability

b) Effects of spatial interdependence at commuting trip destination on e-bike choice probability

c) Effects of spatial interdependence at non-commuting trip origin on e-bike choice probability

d) Effects of spatial interdependence at non-ccommuting trip destination on e-bike choice probability

Figure 1 – Comparisons of E-bike choice probability between models with and without spatial interdependence



a) Models for commuting trips *b) Models for non-commuting trips*

Figure 2 – ROC Curves of models with fixed and random spatial interdependence

models with random spatial interdependence can provide good predictive performance for both commuting and non-commuting trips. For comparison, we have also developed the ROC curves of the e-bike choice models with fixed spatial interdependence for commuting and non-commuting trips. As shown in *Figure 2*, the ROC curves for the models with random spatial interdependence are always to the above of the models with fixed spatial interdependence, indicating that allowing the spatial interdependence to be different for various TAZs can further improve the predictive performance of the spatial autoregressive logistic regression models.

5. CONCLUSION

To promote the e-bike usage, numerous studies have been conducted to develop the e-bike choice model. However, most studies did not consider the effects of spatial interdependence on the choice of e-bikes. Without considering the spatial interdependence, the conventional e-bike choice model may produce biased parameter estimates and forecast errors. The aim of this study is to detect whether the spatial interdependence exists in the e-bike choice, and to investigate how spatial interdependence affects the mode choice of e-bikes. The Moran's I statistics was first used to investigate whether the e-bike choice is spatially correlated at an aggregated TAZ level. The results showed that considerable spatial interdependencies exist in e-bike choice at both trip origin and destination. The Bayesian spatial autoregressive logistic regressions were developed for commuting and non-commuting trips separately to explore the effects of the spatial interdependence on e-bike choice at individual level. The contributing factors to e-bike choice decision-making are different between commuting and non-commuting trips.

The estimation results of the e-bike choice models for both commuting and non-commuting trips indicated that the spatial interdependence exists not only at the origin side of e-bike trips, but also at the destination side of e-bike trips. Travellers are more likely to choose e-bikes if their neighbours at the same trip origin and destination also travel by e-bikes. And the magnitude of this spatial interdependence in e-bike choice decision-making is different across various TAZs. The Bayes factor analyses showed that the inclusion of spatial interdependence significantly increases the fitness of the e-bike choice models for both commuting and non-commuting trips.

One of the important findings is that without considering spatial interdependence, the traditional e-bike choice model may have biased estimation results and make systematic forecasting errors. More specifically, the traditional model without considering the spatial interdependence provides an overestimated e-bike choice probability when the mode share of e-bikes at the trip origin is low; while providing an underestimated e-bike choice probability when the mode share of e-bikes at the trip origin is high.

The ROC curves suggest that the predictive performance of e-bike choice models for commuting and non-commuting trips are satisfactory. The e-bike choice model with random spatial interdependence provides better prediction accuracy than the model with fixed spatial interdependence. The results can help transportation agencies to better understand how spatial interdependence affects the e-bike choice decision-making, and to develop policy initiatives and engineering measures to promote e-bike usage. Moreover, the developed models have the potential to improve the prediction accuracy in the mode share of e-bikes for travel demand forecasting.

ACKNOWLEDGEMENT

This research was sponsored by the National Natural Science Foundation of China (Grant No. 51508093 and 51561135003).

徐铖铖，王晨，王炜，包杰，杨梦琳

电动自行车选择行为的空间相关性研究

摘要：论文研究空间相关性对电动自行车选择行为的影响。莫兰统计量检验表明电动自行车选择行为在交通小区宏观层面上呈现显著的空间相关性，贝叶斯空间自回归模型被进一步用来研究电动自行车选择行为在微观层面上的空间相关性，针对通勤和非通勤出行分别构建模型，结果表明在两种出行的起点和终点都存在空间相关性，即某个出行者所在出行起点和终点的邻居选择电动自行车出行，则该出行者将更容易选择电动自行车出行，并且这种空间相关性的强度在不同交通小区中存在差异。论文研究结果表明不考虑空间相关性的传统模型对电动自行车分担率的预测存在系统性误差。

关键词：电动自行车；空间自相关；空间自回归模型；随机参数回归

REFERENCES

[1] Cherry C, Weinert J, Yang XM. Comparative Environmental Impacts of Electric Bikes in China. Transportation Research Part D: Transportation and Environment. 2009;14(5):281-290. Available from: http://www.sciencedirect.com/science/article/pii/S1361920908001387

[2] Cherry C, Cervero R. Use characteristics and mode choice behavior of electric bike users in China. Transport Policy. 2007;14:247-257. Available from: http://www.sciencedirect.com/science/article/pii/S0967070X07000169

[3] Chiu Y, Tzeng G. The market acceptance of electric motorcycles in Taiwan experience through a stated preference analysis. Transportation Research Part D: Transportation and Environment. 1999;4:127-146. Available from: http://www.sciencedirect.com/science/journal/13619209

[4] Cherry C, Yang H, Jones L, He M. Dynamics of Electric Bike Ownership and Use in Kunming China. Transport Policy. 2016;45:127-135. Available from: http://www.sciencedirect.com/science/article/pii/S0967070X15300524

[5] Jones LR, Cherry C, Vu TA, Nguyen QN. The effect of incentives and technology on the adoption of electric motorcycles: A stated choice experiment in Vietnam. Transportation Research Part A: Policy and Practice. 2013;57:1-11. Available from: http://www.sciencedirect.com/science/article/pii/S0965856413001675

[6] Zhang Y, Li Y, Yang X, Liu Q, Li C. Built Environment and Household Electric Bike Ownership: Insights from Zhongshan Metropolitan Area, China. Transportation Research Record. 2013;2387:102–111. Available from: http://trrjournalonline.trb.org/doi/10.3141/2387-12

[7] Popovich N, Gordon E, Shao Z, Xing Y, Yang Y, Handy S. Experiences of electric bicycle users in the Sacramento, California area. Travel Behaviour and Society. 2014;1:37-44. Available from: http://www.sciencedirect.com/science/article/pii/S2214367X13000185

[8] Heyvaert S, Vanhaverbeke L, Knapen L, Declercq K, Coosemans T, Joeri VM. Choosing an Electric Vehicle as a Travel Mode: Travel Diary Case Study in a Belgian Living Lab Context. Presented at the Transportation Research Board 94th Annual Meeting; 2015.

[9] Lee A, Molin E, Maat K, Sierzchula W. Electric Bicycle Use and Mode Choice in the Netherlands. Presented at the Transportation Research Board 94th Annual Meeting; 2015.

[10] Montgomery BN. Cycling Trends and Fate in the Face of Bus Rapid Transit: Case Study of Jinan, Shandong Province, China. Transportation Research Record. 2000;2193:28-36. Available from: http://trrjournalonline.trb.org/doi/ref/10.3141/2193-04

[11] Dill J. Rose G. Electric bikes and transportation policy insights from early adopters. Transportation Research Record. 2012;2314:1-6. Available from: http://trrjournalonline.trb.org/doi/10.3141/2314-01

[12] Fyhri A, Sundfør HB. Ebikes-who wants to buy them, and what effect do they have? TØI Report 1325/2014. Institute of Transport Economics, Oslo; 2014. Available from: https://www.toi.no/getfile.php/Publikasjoner/T%C3%98I%20rapporter/2014/1325-2014/sum-1325-2014.pdf

[13] Fyhri A, Fearnley A. Effects of e-bikes on bicycle use and mode share. Transportation Research Part D: Transportation and Environment. 2015;36:45-52. Available from: http://www.sciencedirect.com/science/article/pii/S1361920915000140

[14] Goetzke F. Network effects in public transit use: evidence from a spatially autoregressive mode choice model for New York. Urban Studies. 2008;45:407-417. Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.364.2358

[15] Adjemian MC, Williams LJ. Estimating spatial interdependence in automobile type choice with survey data. Transportation Research Part A. 2010;44:661-675. Available from: http://www.sciencedirect.com/science/article/pii/S0965856410000911

[16] Li Q, Lam WHK, Tam M L. Vehicle Travel Time Prediction in Spatio-Temporal Space. Applied Mechanics and Materials. 2013;253:1662-1665. Available from: http://www.scientific.net/AMM.253-255.1645

[17] Lawson A. Bayesian Disease Mapping Hierarchical Modeling in Spatial Epidemiology. CRC; 2009.

[18] Gelman A, Carlin J, Stern H, Rubin D. Bayesian Data Analysis. 2nd ed. London: Chapman and Hall; 2004.

[19] Ntzoufras I. Bayesian Modeling Using WinBUGS. Wiley; 2009.

[20] Washington SP, Karlaftis MG, Mannering FL. Statistical and Econometric Methods for Transportation Data Analysis. Boca Raton, FL: Chapman & Hall/CRC; 2003.

[21] Weiss GM, Provost F. Learning when training data are costly: The effect of class distribution on tree induction. Journal of Artificial Intelligence Research. 2003;19:315-354. Available from:

http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.60.8100

[22] Li Z, Wang W, Liu P, Ragland D. Physical environments influencing bicyclists' perception of comfort on separated and on-street bicycle facilities. Transportation Research Part D: Transportation and Environment. 2012;17:256-261. Available from: http://www.sciencedirect.com/science/article/pii/S1361920911001556

[23] Li Z, Wang W, Shan XF, Jin J, Lu J, Yang C. Analysis of bicycle passing events for LOS evaluation on physically separated bicycle roadways in China. Presented at 89th Annual Meeting of the Transportation Research Board. Washington, DC; 2010. Available from: https://trid.trb.org/view/2010/C/910355