Using geographic information systems and discrete choice modeling to estimate cycling trips volumes in an urban transport network

Álvaro Caviedes, Research Assistant, Universidad de los Andes, Cra 1E No 26A-40 of ML126
Phone: +57 1 339 4949 ext 1807, Email: aa.caviedes800@uniandes.edu.co

Daniel Paez, Dr, Assistant Professor, Universidad de los Andes, Cra 1E No 26A-40 of ML744
Phone: +57 1 339 4949 ext 3440, Email: dpaez@uniandes.edu.co*

*Corresponding author

Submitted for publication and presentation
Transportation Research Record

Committee number ANF20
(Bicycle transportation)

July 2014

Words: 5533 + 7 tables =7,283 (limit = 7,500)

Review Submission

Abstract

Bicycling is expected to increase as a mode of transportation for all trip purposes. City planners are constantly searching for technical methodologies that provide decision-support around bicycle infrastructure and changes attitudes from seeing cycling as a ‘recreational activity’ to a primary mode of transport. However, four steps transport models (generalization, distribution, mode split, and network assignment) have been designed for motorized transport and provide limited applicability to bicycling. Cycling trips are different to those by vehicle as users’ main concern is not road congestion caused by cars but other riding factors such as road safety and comfort. This paper presents a methodology that combines geographic information systems and discrete choice modeling to incorporate cycling in the fourth step of the transport model. Extensive geographically based surveys of routes were used to build a statistical model that considers both riders’ demographic and cycling infrastructure characteristics. Results from the case study in Bogota suggest that this methodology opens an opportunity to consider cycling as a main transport mode by allowing planners to assign cycling trips to an entire urban transport network, something not achieved before. However, the determination of the parameters of the methodology developed is highly dependent on the city of study and these appeared not to be transferable without proper fieldwork. Results from the investigation open research opportunities to develop the methodology for the assessment of different types of cycling infrastructure proposals and to apply it in different contexts both in developed and developing countries.

Keywords: Bicycle assignation model, Cyclist behavior, Path Size model

Introduction

Today, authorities pay particular attention to the performance of cities in terms of pollution and sustainable development. Transport projects for cities are key to sustainable development. These projects have to respond to mobility goals and new challenges such as social inequality and climate change. Additionally, it is expected today that transport systems should contribute to air quality and provide comfort and efficiency.

For every 10 minutes a person spends in a car, there is a 10% decrease in participation in community life (Putman, 2000) and research has shown that the greater the traffic volume in a city, the lower the social interactions (Appleyard, Gerson, & Lintell, 2009). In American cities, lack of physical activity due to excessive use of the car is leading to a decline in life expectancy in the new generation, increasing health care costs by 4% of the federal budget in the country (PeopleForBikes and Alliance for Biking & Walking).

Bycicling has been proven to be a sustainable transport mode option (Pucher & Buehler, 2006). It is considered by its users to be inexpensive and healthy (Heinen, Maat, & Van Wee, 2010).

The benefits for implementing politics and projects around the bicycle in a city are in many areas. Different studies conducted in the United States conclude that there is an increase in economic benefits due to proximity to stations for bicycles and bike paths. A survey conducted in Washington D.C concluded that the flow of people increased significantly due to the stations of public bicycles, generating a positive impact on sales and in the neighborhood (Buehler R. , 2014).

As for the public health benefits, a review study evidenced that in Copenhagen the population using the bike to commute (at least three hours per week) experienced 40% less risk of death compared with those who did not use it. In Shanghai, China, the behavior of more than 70,000 women were studied over six years and found that those using bicycles and walking as transportation were 40% less likely to die from heart disease or cancer compared with women who did not engage in physical activities, those with a more sedentary life (Jacoby & Pardo, 2010).

Based on the benefits of the bicycle, it is important to promote its use in a city. One method to do this is by planning, designing, and building cycling infrastructure. The bike paths have to connect at least the main zones in a city by direct routes, offer secure infrastructure and smooth pavement, protect the cyclist from motorized traffic, and attract users by integrating land use and transit (Kuijper & Braakman, 2009).

To evaluate the feasibility of a bike path project, normally the traffic planners start with the traditional four-step algorithm (trip generation, trip distribution, mode choice, and route assignment). The ultimate goal of the algorithm is to allow the allocation of traffic volumes to individual links so they can be designed (or evaluated) under different scenarios.

However, there is a gap in the four-step algorithm for bicycle trips because route assignment principles (fourth step) have been designed for motorized transport. Assigning a travel network refers to predicting and selecting the paths or routes used by travelers in order to map flows on the network. This step is modeled on a trade off principle and it is assumed the user will choose the route with less congestion and/or distance. Both bicycling users do not always follow this assumption and other factors such as road safety, personal safety, and comfort are important.

Additionally, traditional route assignment needs to have capacity constraints in every link of the network so when shortest routes are congested, users take alternatives. Increases in bicycling users on a link would rarely cause additional traffic, making the fourth step inapplicable for bicycling.

This study proposes an alternative to assigning bicycle trips to the network. It combines advanced geographic information systems and discrete choice modeling to incorporate cycling in the fourth step of the transport model. Results from the case study in Bogota suggest that this methodology opens an opportunity to consider cycling as a main transport mode by allowing planners to assign cycling trips to an entire urban transport network.

The next section presents the research background and how findings from previous research studies were used to support the methodology proposed.

1. RESEARCH BACKGROUND

It is recognized that how cyclists choose a route differs from the drivers of motorized vehicles. While drivers seek to reduce travel time and vehicle operation costs, cyclists when choosing a route in addition to valuing travel time, they consider its suitability (safety, traffic volume, traffic speed, dedicated lanes, type of terrain, etc.).

Each cyclist chooses the route based on its preferences and the factors that affect its decision. Ehrgott et al. (2012) proposed a model to determine a set of possible routes taking into account two specific objectives, travel time and suitability of the route. Papinski and Scott (2011) developed a model to create bicycling routes that minimize travel time and the travel distance between two zones. These options were compared with observed routes to find their differences. This conclusion was that there are other factors that affect route selection as each person has a pattern of travel depending on personal preferences. Therefore, knowing the trip information of each user helps determine transportation habits (Papinski & Scott, 2011).

Factors that affect the route selection can be divided in two groups: infrastructure characteristics and user characteristics. In relation to infrastructure characteristics, the distance of a route affects negatively its selection. As a result, users are more likely to choose routes that included not segregated infrastructure for short distances (Kang & Fricker, 2013). The value of travel time savings for cyclists is greater than that of other transport modes (Borjesson & Eliasson, 2012). For this reason, cyclists often value highly on-road cycling infrastructure improvements (such as bike lanes). Today, there is evidence that favors assessing bicycling improvements based on speed and comfort as they are cost effective compared with other types of investment (Borjesson & Eliasson, 2012). In this respect, bicycling should be seen as a competitive mode as the embodiment of travel and not simply as a means to better health or reduction of car traffic (Borjesson & Eliasson, 2012).

Many studies have covered the area of user characteristics. Roach, Gliebe, and Dill (2009) followed 164 cyclists with GPS units in order to understand better the cyclist route preferences. The results were coded to a very detailed bicycle network. The authors used 1449 commute trips to estimate a bike path choice model. This model was then incorporated into a generation algorithm according to the travel choice set. The main method was to calculate multiple permutations of path attributes and, then, to take into account the overlap of alternative routes. Findings showed that cyclists are sensitive to the effects of distance, frequency, slope, signals (e.g. the presence or absence of signs), and traffic volumes. The provision of bicycle lanes tries to compensate for the negative effects of adjacent traffic, but this is not to the same magnitude as having a street with low traffic volume. Finally, the route preferences of cyclists differ between daily travel and other utilitarian trips (Roach, Gliebe, & Dill, 2009).

Therefore, it could be concluded that cyclists are more sensitive to distance and less sensitive to other characteristic infrastructure of daily trips. This research by Roach, Gliebe, and Dill also developed a modified method for route assignment called labeling network. In this method alternate routes were chosen by maximizing individual criteria, thus allowing it to justify a larger portion of observed routes than other methods such as shortest path. Also, for the route selection model, they highlighted the drawbacks of Multinomial Logit for route selection as it appears to give more utility to routes that share a link. A solution to this might be the use of Path Size Logit (Roach, Gliebe, & Dill, 2009).

Therefore, the literature reviewed shows that there is an approximation to understand how the cyclists choose their route. However, there is a problem assuming that the bicycle users have the same behavior choosing the travel route. The problem assigning the bicycle OD matrix to the network has not been completely resolved. In the next section an explanation of the methodology used to address this problem is presented.

2. METHODOLOGY

The aim of this study is to develop a methodology to assign bicycle trips (based on surveys as part of a four-step transport model) to the network. We argue that there are different cyclist profiles, with different socio-demographic characteristics, that affect their decision to choose a route between a particular origin and destination. For example, it could be possible that young people are more disposed to share the route with traffic compared with old people.

A way to route utility by each type of user is using a discrete choice model such as Path Size Logit or PSL (Ben-Akiva & Bierlaire, 2003). The purpose of a PSL model applied to bicycling route utility is to determine the probability of using a specific route against the other options based on the characteristics of the person and the route.

For example, we have two zones of study, one origin and one destination. There are also only three possible routes to go from the origin to the destination: Route A, Route B, and Route C. Each route option has different characteristics of slope, distance, availability of bike paths, etc. Furthermore, based on the socio demographic characteristic of cyclists in the origin zone, we can divide that population into three cyclist profiles: Profile I, Profile II, and Profile III. With the discrete choice model, constructed from the information of the attributes of the routes and the profiles, it is possible to determine the proportion of trips of each cyclist profile in each route. We can note, for example, that of the Profile I cyclists, 50% will go by Route A, 20% by Route B, and 30% by Route C.

2.1. Route set generation sub-model

The methodology proposed, based on PSL, is different to that used by the four-step model, as we consider only a limited number of routes. Therefore, it is necessary to create a sub-model that generates all the possible and feasible routes that a cyclist has from a particular origin to a destination. To this end, we modified, using geographic information systems (GIS), the label generation model (Broach, Gliebe, & Dill, 2009) for route set generation. In this model a “label” refers to the attribute or characteristic that is going to be evaluated in the bicycle network

This sub-model for route set generation has the following steps:

1. A city network is used and all links for cycling (such as highways, main streets, secondary streets, and bike paths) are connected. The links are then converted from vector to raster or grip with a size of 3 x 3 meters.
2. With this raster a short distance route between transport zones are calculated. Short routes will be those with the lowest sum of the grid values. This shortest route will be used as the basis for other alternatives

3. Based on the literature, the following factors are used to determine different route options:
   - Bike paths: this label will maximize the bike paths on the route.
   - Highways: this label will maximize the highways on the route. (This label was chosen to determine whether there is a cyclist profile that prefers to use this route rather than the other options).
   - First order streets: this label will maximize the first order streets in the route. First Order Street refers to streets that follow in priority after the highway. A high traffic volume normally characterizes them.
   - Second order streets: this label will maximize the second order streets on the route. Second Order Street refers to the street that follows in priority after the First order streets. These streets are mainly the neighborhood streets, where traffic speed and volume are low compared with highways or first order streets.
   - Slope: this label will minimize the slope on the route.
   - Intersections: this label will minimize the number of dangerous intersections on the route. For the study, we consider a dangerous intersection one where the cyclist has to cross a highway or a first order street.
   - Motorized traffic volume: these labels will minimize the motorized traffic volume on the route.
4. To calculate the routes based on the labels, a reference grid, with a lower utility value, is constructed. For example, if we want to determine the route with the bike path label, the grid network will have a value of 2, except for the grids that represent the bike path; they will have a value of 1.
5. Definition of the label cost function:

$$Cost\ network = Weight * (Short\ route\ network) +\ (1 - Weight) * (Label\ network)$$

The weight is a value between 0 and 1.

   If the weight is 1, the cost network will be the same as the short route network; consequently, the route that minimizes the cost will be the short route. If the weight is 0, then the cost network will be the label network; and the route that minimizes the cost for this network will be the one that maximizes the label. Figure 1 shows how the cost network is calculated. The sum of each grid of the short route network and each grid of the label network will give the cost network.
6. If the weight is 1, the cost network will maximize the label. However, it is possible that the route calculated will have a long distance compared with the short route and it will not be a feasible option for the cyclist. For this reason, it is necessary to create an iterative model that will maximize the label network in each iteration. In the first iteration the weight is 0, which will increase by 0,1 in each iteration until the weight is 1 or the admissible route distance is exceeded.
7. The admissible route distance is the distance that is accepted as feasible because it does not have a significant difference to the short route distance; so the cyclist will have this route as an option for a trip. The admissible route distance is calculated from field information, and it is possible to know how much (in %) the cyclists ride compared with the short route.
8. The process is repeated with another attribute.

2.2. Selection of alternative

After the calculation of each feasible option between an origin-destination pair, it is possible to determine the probability of a cyclist profile to choose a route. For this purpose, PSL was used as a discrete choice model that could calculate the probability of using a route over the other existing options, based on the utility that it generates for the user.

The PSL has the next form:

$$\Pr(i\text{l}\mathrm{C_n}) = \frac{exp^{V_{in}+In\,(PS_{in})}}{\sum_{j\in C_n} exp^{V_{jn}+In\,(PS_{jn})}}$$

Where V_{in} is the utility of the route i and PS is the path size coefficient.

The PS coefficient is a correction term for overlapping links. It avoids the error for correlated routes and prevents an overestimation in the probability for routes that are sharing links with other routes. This factor only affects routes with overlapping links; routes with unique links will not need adjustment (Ramming, 2002).

$$PS_{in} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_n} \left(\frac{L_i}{L_j}\right)^{\gamma} \delta_{aj}}$$

Where:

- j are the routes of a set of alternatives (Cn) for an origin destination pair
- Γ_i are the links of the route i
- l_a is the distance of the link a on the route i
- L_i is the distance of the route i
- δ_{aj} is a parameter equal to 1 if the route j includes the link a, otherwise it is 0. The minimum value of δ_{aj} is 0, because at least one route has to have the link a, otherwise the link belongs to the other set of routes Cn.
- γ is a parameter to penalize routes with long distances compared with the other options in the set Cn. Because we are taking into account this fact in the set generation model, we assign γ as 0

$$PS_{in} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{aj}}$$

The PS coefficient is always between 0 and 1. In the PSL, this coefficient is inside Ln (Ln (PS)), resulting in a negative value that will decrease the utility (consequently, the probability) of a route if it has overlapping links with other routes (Ramming, 2002).

The other value in PSL is the utility. For our purpose, we define the utility as:

$$V_{in} = \beta_0 + \sum_{1}^{n=route\ attribute} \beta_n * x_n + \sum_{1}^{k=profile\ cyclist\ attribute} \beta_k * x_k$$

Where:

- V_{in} is the dependent variable, the utility that a route generates to a cyclist based on the characteristics of the route and the user
- β_0 is the constant of the model
- β_n are the parameters for the route attributes
- β_k are the parameters for the cyclist profile attributes

*The attributes defined for the route are:

- Percentage of bike path on the route
- Percentage of highway on the route
- Percentage of First Order Street on the route
- Percentage of Second Order Street on the route
- Average slope
- Public traffic volume on the route
- Private traffic volume on the route

*These attributes are calculated for each of the routes determined in the set model generation.

The attributes defined for the profile cyclist are:

- Gender
- Age
- Stratum
- Income
- Education
- Skills using the bicycle (1 to 5, where 5 means high skills)
- Experience using the bicycle as a transport mode for daily trips

3. CASE OF STUDY

Both the route set generation sub-model and route selection model were applied in Bogota, Colombia. Bogota has more than 360 km of bike paths, and generates approximately 611500 trips by bicycle per day, which corresponds to 3% of the total trips of the city (Preparado por la Unión Temporal Steer Davies & Gleave Limited y Centro Nacional de Consultoría, 2011). There are three peak hours for the bicycle trips: 6 am to 8 am, 12 pm to 2 pm, and 5 pm to 6 pm. In these peak hours, the percentage of trips made by bicycle increases more than 3%. Also, 72% of these trips are mandatory trips (work, study, job searching, and work issues). Low-income people make the majority of cycling trips in Bogota. Work commuting is the most common purpose for these trips. For people with high-income, the main reason for using a bicycle is recreational. The average travel time made by bicycle is 26 minutes, and the distance traveled average is 6,6 km. Additionally, low-income people in Bogota have to expend more time and travel long distances compared with the averages (Preparado por la Unión Temporal Steer Davies & Gleave Limited y Centro Nacional de Consultoría, 2011).

To find the parameters of the model PSL, extensive geographically based surveys that included socio demographic characteristics were conducted. To determine the place to make the survey, we chose the origin-destination pairs with the highest number of bicycle trips according to mobility surveys. Then, we identified which were the main streets, the bike paths, and secondary roads that connect those OD pairs. With these options, we selected 12 points across the city (covering the entire city and the entire strata) to make the survey. In total, 605 surveys were conducted.

Cyclists were stopped at intersections so they could answer a set of questions that included socio demographic characteristics (described in the previous section), bicycle skills, frequency of trips, and its geographically-based travel route. For the route, interviewed people had to draw on a map the route, describing the links on the transport network used during the bicycle journey.

4. RESULTS

Socio demographic variables that affect the route selection based on the survey were identified for Bogota. For this purpose, we conducted standard linear regressions. In the regression, the dependent variable was % of the route – the bike path – and the independent variable was the level of study. The main results from the regression analysis are shown in table 1.

TABLE 1 Linear Regression Use of Bike Paths Based on Education Level

VARIABLES	(Reg1) %Bike Path	(Reg2) % Bike Path
Bikepath_MS	0.0657**	
	(0.0282)	
Bikepath_B		0.129***
		(0.0300)
Constant	78.59***	78.06***
	(1.058)	(1.011)
Observations	605	605
R-squared	0.009	0.030

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

In Table 1, MS refers to the cyclists whose maximum level of study is middle school, and B refers to the cyclists whose maximum level of study is high school. Comparing the regression 1 with the regression 2, we found that, on average, the highest level of education means more use of bike paths.

TABLE 2 Linear Regression Use of Highways Based on Education Level

VARIABLES	(Reg1) Highway	(Reg2) Highway	(Reg3) Highway
Highway_Stratum1and2	0.944***		
	(0.0211)		
Highway_MS		0.849***	
		(0.0597)	
Highway_B			0.696***
			(0.164)
Constant	2.115***	5.586***	6.934***
	(0.303)	(0.518)	(0.581)
Observations	605	605	605
R-squared	0.768	0.251	0.029

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In this case, shown in Table 2, the dependent variable was the percentage of highway on the route selected. The results from this regression show that, on average, people with a low income find it more significant to use highways than people with high income. Similarly, for education, people with a low level of study, on average, prefer to use the highway compared with people with a high school degree. It is important to note that for al the cyclists, the bike path on the route is significant, but the significance has a variance depending on the characteristics of the cyclist.

By using the same regression methodology, it was possible to conclude other relationships between route characteristics and user demographics. For example, those users who consider themselves competent or have more than three years of experience would also favor routes with a highway component.

On the other hand, people with less experience (less than one year) prefer segregated cycling infrastructure such as bike paths.

In relation to income, the results from our analysis suggest that residents of low-income areas, on average, have a significant portion of routes on highways. However, it is important to note that although Bogota has more than 360 km of bike paths, there are some low-income districts without segregated bicycle paths. Therefore, it cannot be concluded that a low-income person prefers on-road cycling as today limited off-road alternatives are offered.

In contrast, based on the data collected, low-income users are not educated but highly experienced riding bicycles.

It was also interesting to find that in the survey, the women use more bike paths compared with the men. And the men found it more attractive to use the highway than the women. It is possible that the survey did not represent well the behavior of the female gender, the variance is not significant, and the percentage of women in the survey (13%) is less than the percentage of women in the mobility survey (26%).

Finally, considering the geographically based surveys of routes conducted, longer routes would have more segments in main roads and less in secondary roads. This is due to the fact that routes using highways in most cases represent the shortest route. Therefore, users of long distance routes are more likely to prefer the shortest path even though it might not provide proper cycling infrastructure.

Before a PSL model could be developed for all types of users, a route set generation sub-model was developed for the area of influence of the surveys conducted. Seven options (one for each label) and the short route were created for 545 surveys. The remaining surveys (10% of the total) were reserved for validation purposes of the route set sub-model and network assignation model. The results for the eight options analyzed are presented in the following table.

TABLE 3 Label Characteristics

Option	Label	Characteristics						
		% of bike path	% of highway	% of First Order Street	% of Second Order Street	Public traffic volume	Private traffic volume	Slope %
1	**Bike path**	79,55%	4,71%	4,45%	11,19%	34,21	163,73	1,95

2	**Highway**	2,80%	65,26%	14,18%	17,87%	362,76	1183,6	1,99
3	**First Order Street**	11,50%	16,67%	49,33%	22,51%	107,31	530,03	2,08
4	**Second Order Street**	5,21%	5,74%	5,42%	83,62%	148,59	420,09	2,13
5	**Slope**	17,29%	27,53%	18,41%	36,77%	182,1	672,38	2,06
6	**Intersection**	31,39%	29,37%	9,82%	29,41%	188,41	897,36	1,98
7	**Traffic Volume**	65,67%	4,84%	2,68%	26,80%	56,49	351,31	1,93
8	**Short route**	14,13%	30,16%	19,96%	27,85%	364,32	675,77	2,01

With this route set, a discrete choice model based on PSL was developed. This PSL uses statistical descriptions and the regressions deducted from the surveys. As a result, four combinations of model results appear to have statistical significance. They are described next.

Model 1

Model 1 takes into account the % of bike path on each route. This model allows determining how much the % of bike path on a route increases the utility for cyclists. The results from model 1 are shown in table 4. The following equation expresses the model in a mathematical way:

$$Prob(option\ i|route\ set\ j) = \frac{e^{\beta_0+\beta_1\times\%route\ path_i+\ln PS}}{\sum_i^j e^{\beta_0+\beta_1\times\%route\ path_i+\ln PS}}$$

TABLE 4 Paramaters Model 1

Option	Parameter	Value	Z
All the options	β1 (%Bike path)	0.07	12.37
1	β0	-1.16	-2.97
2	β0	0(base)	.
3	β0	-0.24	-0.42
4	β0	-0.08	-0.22
5	β0	0.30	0.77
6	β0	-2.90	-3.11
7	β0	-1.96	-4.92
8	β0	-0.52	-0.94

Model 2

In this model, we included the gender factor. This variable was investigated in relation to the % of bike path or segregated infrastructure on each route. The first was because the gender was a dummy variable where 1 is male and 0 female. The second was to evaluate the PSL based on the interaction between gender and bike path; this interaction reveals whether there is more utility for gender over the other. The following equation expresses the model in a mathematical way:

$$Prob(option\ i|route\ set\ j) = \frac{e^{\beta_0+\beta_1\times\%route\ path_i+\beta_2\times\%route\ path_i\ \times dummy\ gender\ +\ \ln PS}}{\sum_i^j e^{\beta_0+\beta_1\times\%route\ path_i+\beta_2\times\%route\ path_i\ \times dummy\ gender\ +\ \ln PS}}$$

TABLE 5 Paramaters Model 2

Option	Parameter	Value	Z
All the options	β1 (%Bike path)	0.12	5.08

All the options	β2 (dummy gender * % bike path)	0.07	-2.13
1	β0	-1.20	-2.99
2	β0	0(base)	
3	β0	-0.22	-0.37
4	β0	-0.07	-0.18
5	β0	0.31	0.78
6	β0	-2.87	-3.06
7	β0	-1.99	-4.86
8	β0	-0.50	-0.90

Model 3

In model 3 the utility was only calculated based on the percentage of highway. As expected, the sign of the constant is negative, which could be interpreted as people do not prefer to use the highway if they have another option.

$$Prob(option\ i|route\ set\ j) = \frac{e^{\beta_0+\beta_1\times\%highway_i+\ln PS}}{\sum_i^j e^{\beta_0+\beta_1\times\%highway_i+\ln PS}}$$

TABLE 6 Paramaters Model 3

Option	Parameter	Value	Z
All the options	β1 (%Highway)	-0.019	-3.63
1	β0	2.35	9.85
2	β0	0(base)	
3	β0	-0.55	-1.01
4	β0	0.99	3.51
5	β0	0.33	1.01
6	β0	-2.73	-3.68
7	β0	0.95	3.58
8	β0	-0.71	-1.41

Model 4

Finally, in model 4, the following independent variables were included:

- % of highway on the route and
- level of study
- if the cyclist only finished middle school or earned a bachelor's degree:

$$Prob(option\ i|route\ set\ j) = \frac{e^{\beta_0+\beta_1\times\%highway_i+\beta_2\times\%highway_i\times dummy\ middle\ school+\beta_3\times\%highway_i\times dummy\ bachelor+\ln PS}}{\sum_i^j e^{\beta_0+\beta_1\times\%highway_i+\beta_2\times\%highway_i\times dummy\ middle\ school+\beta_3\times\%highway_i\times dummy\ bachelor+\ln PS}}$$

TABLE 7 Paramaters Model 4

Option	Parameter	Value	Z

All the options	β1 (%Highway)	-0.014	-2.61
All the options	β2 (dummy Middle school x %highway)	0.004	3.06
All the options	β3 (dummy bachelor x %highway)	-0.053	-2.06
1	β0	2.11	8.68
2	β0	0(base)	
3	β0	-0.63	-1.16
4	β0	0.95	3.38
5	β0	0.30	0.91
6	β0	-2.70	-3.64
7	β0	0.80	3.00
8	β0	-0.72	-1.43

As in model 2, on average, people with low levels of education prefer more to use highways than people with high levels of education. It is important to note that the existence of highways on the route option generates, on average, a disutility (B1 = -0.014) for all users. However, people with high levels of education have a higher disutility compared with other users.

However, validation of each model, using the 10% of surveys not used, showed a different result for each model. The coefficient of determination (r2) is 0.46 for model 1, 0.09 for model 2, 0.38 for model 3 and 0.60 for model 4. Statistics tests for model 4 appear to be more significant than the others and selected as the most appropriate for the case of Bogota.

Finally, having chosen model 4 as the best model to assign the trips for this case of study, we made a comparison with all or nothing model, which was normally used in Bogota. Hence, for the remaining surveys, we simulated the selected route based only on the distance, the short distance. Once we identified the shortest route, we calculated the % of bike path on the route and compared it with the value of the observed value. Then, we determined the coefficient of determination to evaluate how much all or nothing model explained the reality. The all or nothing model only explains 4 of 100 scenarios simulated.

5. CONCLUSIONS

Based on the sample and the model, we found that, on average, having a bike path on the route increases the utility of the cyclists, and having a highway decreases it. However, this utility or disutility is different for different kinds of cyclists. For example, for females it is more significant to have bike paths on their route than for men. Also, people with high levels of educations prefer, on average, to use the bike path or segregated infrastructure.

Based on results for Bogota, it could be argued that people who have less disutility using a highway, probably have low-incomes and live in a zone with a general low-income standard. However, the research cannot determine whether this preference for highways is based on their income level or related to the fact that low-income areas have poor segregated bicycle paths.

The set generation sub-model appears as an appropriate methodology to calculate the different options that cyclists have to get to their destination for two main reasons. In the first, routes are avoided that have a long distance and will not be feasible for the cyclists. The average distance of the options will be realistic because it was calculated from an observation of the reality. Second, when the options are calculated, there is an option that perfectly fits with the option that the cyclist of the sample chose. In this regard, it is possible to calculate the Path Size Logit because we have information about the options that the cyclists have, and that which they choose.

The PSL, with the characteristics of the route and the cyclist in the utility function, was found to be a better option to assign bicycle trips to the network than using an all or nothing approach based on the shortest path. Survey results demonstrated in this study that cyclists have different behaviors and value differently the characteristics of the routes. Although some of them prefer to go on the shortest route, there are other cyclists who prefer to use the bike paths regardless of whether these are a longer distance.

In conclusion, the aim of this study, to develop an assignment method for bicycle trips to a transport network, was achieved. Based on these findings, with the structured model it is possible to assign the bicycle trips to the network if an OD matrix is available for the city. Also, this methodology is applicable to any city, however, the parameters have to be calculated again (with a survey) based on the determination of the socio demographic characteristics that affect the study dependent variable.

For future work, it is of great value to determine which are more justifiable from statistical manner reasons that justify the perception of people of low levels of education and low incomes in relation to highways or main roads.

6. References

Jacoby, E., & Pardo, C. *Ciudades del automovil, obesidad y cambio climático:se alinean las crisis y también las soluciones.* Bogotá: Slow Research, 2010.

Putman, R. *Bowling Alone: The Colla´se and Rivival of American Community*. Simon & Schuster, 2000.

Appleyard, D., Gerson, S., & Lintell, M. *Livable Streets*. University of California Press, 2009..

PeopleForBikes and Alliance for Biking & Walking. *Protected bike lanes mean business.*

Pucher, J., & Buehler, R. Cycling for everyone: Lessons from Europe. *Transportation Research Record:Journal of the Transportation Research Board*, 2006, pp. 58-65.

Heinen, E., Maat, K., & Van Wee, B. (2010). The Role of Attitudes Toward Characteristics of Bicycle Commuting on the Choice to Cycle to Work over Various Distances. *Transport Research* , Vol. 106.

Martens, K. The Bicycle as a Feeding Mode: Experiences from Three European Countries. *Transportation Research*, 2004, pp. 281-294.

Buehler, T., & Handy, S. Fifty years of bicycle policy in Davis, California. *Transportation Research Record: Journal of the Transportation Research Board*, 2008, pp. 52-57.

Tolley, R. *The Greening of Urban Transport: Planning for Walking and Cycling in Western Cities*, 1990.

Clifton, K., Morrissey, S., & Ritter, C. Mode choice and consumer spending: examination of grocery store trips. *Proceedings of the 91st TRB annual meeting*, 2012.

Buehler, R. Business appears to be picking up near capital bikeshare stations. http://mobilitylab.org/2014/03/05/business-appears-to-be-picking-up-near-capital-bikeshare-stations. Accessed 03, 05, 2014.

Kuijper, D., & Braakman, B. Five main requirements for cycling-inclusive infrastructure. In G. T. Services, *Cycling-Inclusive Policy Development: A Handbook* (pp. 57-65). Utrecht, 2009.

Ortúzar, J., & Willumsen, L. *Modelos de transporte.* 2008: Universidad de Cantabria, 2008.

Broach, J., Gliebe, J., & Dill, J. Development of a multi-class bicyclist route choice model using revealed preference data. *Presented at the 12th International Conference on Travel Behavior Research, India*, 2009.

Ehrgott, M., Wang, J., Raith, A., & Houtte, C. A bi-objective cyclist route choice model. *Transportation Research Part A: Policy and Practice*, 2012, pp. 652-663.

Papinski, D., & Scott, D. A GIS-based toolkit for route choice analysis. *Journal of Transport Geography*, 2011, 434-442.

Kang, L., & Fricker, J. Bicyclist commuters' choice of on-street versus off-street route segment. *Springer Transportation*, 2013, 887-902.

Borjesson, M., & Eliasson, J. The value of time and external benefits in bicycle. *Transportation Research Part A: Policy and Practice*, 2012, 673-683.

Ben-Akiva, M., & Bierlaire, M. Discrete choice methods and their applications to short term travel decisions. *Handbook of Transportation Science*, 2003.

Ramming, M. *Network Knowledge and Route Choice.* Massachusetts Institute of Technology, 2002.

Preparado por la Unión Temporal Steer Davies & Gleave Limited y Centro Nacional de Consultoría. (2011). *Informe de indicadores Encuesta de Movilidad de Bogotá 2011*. Bogotá: SECRETARÍA DISTRITAL DE MOVILIDAD.