ORIGINAL PAPER

Open Access

Simulating the effects of redesigned street-scale built environments on access/egress pedestrian flows to metro stations

Yanan Liu^{*} , Dujuan Yang, Harry J. P. Timmermans and Bauke de Vries

Abstract

In urban renewal processes, metro line systems are widely used to accommodate the massive traffic needs and stimulate the redevelopment of the local area. The route choice of pedestrians, emanating from or going to the metro stations, is influenced by the street-scale built environment. Many renewal processes involve the improvement of the street-level built environment and thus influence pedestrian flows. To assess the effects of urban design on pedestrian flows, this article presents the results of a simulation model of pedestrian route choice behavior around Yingkoudao metro station in the city center of Tianjin, China. Simulated pedestrian flows based on 4 scenarios of changes in street-scale built environment characteristics are compared. Results indicate that the main streets are disproportionally more affected than smaller streets. The promotion of an intensified land use mix does not lead to a high increase in the number of pedestrians who choose the involved route when traveling from/to the metro station, assuming fixed destination choice.

Keywords: Street-level built environment, Pedestrian flow, Route choice, Simulation

1 Introduction

To satisfy the increasing travel demand caused by fast urbanization, Chinese cities have resorted to new metro systems and large scale urban redevelopment in megacities. Metro stations generate large pedestrian flows in the streets surrounding the stations (Yang, Chen, Le, & Zhang, 2016; Zacharias, 2001). However, roads in metropolitan regions were designed mainly to serve vehicles rather than pedestrians. Therefore, the metro stations often led to a conflict between the original built environment and the large number of pedestrians generated by the new metro stations. The inconvenience for pedestrians is caused by larger block sizes, narrow sidewalks, lower land use mix (Clifton, Smith, & Rodriguez, 2007), fewer crossing facilities at intersections, larger width of vehicle roads associated with unsafe crossings (Li, Fisher, Brownson, & Bosworth, 2005; Lin, 2015). It has triggered

urban renewal processes in these areas aimed at improving the streetscape.

To improve the street-level built environment around metro stations for pedestrians who travel from/to the metro station, it is necessary to understand the relationship between street-level built environment characteristics and pedestrian route choice. Unfortunately, only a few studies examined the quantitative relationship between the street-scale built environment around metro stations and pedestrian route choice decisions (e.g., Guo & Loo, 2013; Liu, Yang, Timmermans, & de Vries, 2020; Lue & Miller, 2019; Nuworsoo, Cooper, Cushing, & Jud, 2013). They found a significant impact of built environment characteristics on pedestrian route choice. On average, pedestrians tend to prefer wide sidewalks, low FAR, more retail frontage, fence appearance, and signalized intersections. Moreover, there is evidence that density, number of walkway crossings, and road width influence shopping pedestrian route choice behavior (e.g., Borgers & Timmermans, 2014; Borgers &

^{*} Correspondence: y.liu.1@tue.nl

Department of the Built Environment, Eindhoven University of Technology, Vertigo 9.09, 5600 MB Eindhoven, the Netherlands

Timmermans, 2015; Tilahun & Li, 2015; Zhu & Timmermans, 2011). As the relationship between the built environment and pedestrian route choice varies by context (e.g., Moran, Rodriguez, & Corburn, 2018; Rodriguez et al., 2015; Tribby, Miller, Brown, Werner, & Smith, 2017), it is necessary to allow for heterogeneous, context-dependent preferences of pedestrians.

Different pedestrian route choice models have been applied. Several researchers have used macroscopic models to investigate the impact of different built environment design scenarios on the distribution of pedestrian flows. For example, Schelhorn, O'Sullivan, Haklay, and Thurstain-Goodwin (1999) applied an agent-based simulation model to predict pedestrian movements in urban centers considering road configuration and the spatial location of the attractions. Dijkstra, Jessurun, Timmermans, and de Vries (2011) simulated the movements of shopping pedestrians, considering a variety of shops, number of stores, road network characteristics, and pedestrianized streets. Rose, Ligtenberg, and van der Sperk (2014) included retail, drinking and dining facilities, and cultural attractions. Examples of microscopic pedestrian flow models include Daamen and Hoogendoorn (2003a), Daamen and Hoogendoorn (2003b), Ukkusuri, Miranda-Moreno, Ramadurai, and Isa-Tavarez (2012), and Chen (2017). The latter built a simulation model of pedestrian movement in a subway station.

This paper concentrates on the simulation of pedestrian flows from/to a metro station. It aims to provide feedback to urban planners about the consequences of different renewal scenarios on the distribution of pedestrian flows. Simulated flows can be used to elaborate on particular scenarios or combined different scenarios in light of a set of different and possibly conflicting objectives. It contributes to the knowledge of pedestrian-friendly street-scale (micro) built environment design surrounding metro stations.

The paper is organized as follows. First, we will discuss the data collected for this study. It is followed by an outline of the simulation approach. Next, we discuss the results of the simulations, which include the results of uncertainty analysis. The paper is concluded with a discussion of the main results.

2 Study area and data

This study is conducted in the area surrounding Yingkoudao (YKD) metro station, located in the city center of Tianjin, China, with a size of 2000*2000 m squared. YKD station has a large passenger flow and is the transit station of metro lines No.1 and No.3. The area is the biggest commercial and business center in Tianjin. It includes various land uses (e.g., shopping malls, business, residential, primary and middle schools, restaurants, entertainment, and open spaces). The case study area

includes not only streets with wide sidewalks, wide roads, very high-rise buildings, and well-maintained crossing facilities, but also streets with narrow sidewalks, low buildings, and few crossing facilities.

The core of the simulation is a pedestrian route choice model. It requires data on observed pedestrian flows and street-level characteristics of the built environment. Data on pedestrian volume at the metro station were collected in March 2016 by five trained Bachelor students from Tianjin University. They collected data on the number of pedestrians who entered respectively exited the Yingkoudao metro station during three different time periods: morning (8:00–9:00), noon (12:00–13:00), and afternoon (17:00–18:00). Phone cameras were used to capture pedestrian volumes at every exit/entrance of the station. Observations were made on two randomly selected weekdays and two randomly selected weekend days (excluding national holidays). Observations were made for 2 min in each time period that includes arrivals/departures of subway trains.

The number of pedestrians who entered/exited the metro station in each video was counted by the investigators. To avoid significant error, each video was processed twice by two different investigators. If the difference in the number of counted pedestrians was bigger than 2, the video was processed again; otherwise, the average was used as the final count. Based on the counted number of pedestrians in the 2-min videos, the total number of pedestrians for each exit/entrance for each time period was extrapolated. The peak count was observed between 17:00 and 18:00. The lowest number of passengers was observed for the 8:00–9:00 time period. Estimated counts suggest that 5040 pedestrians/hour enter/exit the station in the morning (8:00–9:00), 6810 at noon (12:00–13:00), and 11,220 in the afternoon (17:00–18:00).

The second data collection, which was a paper-based questionnaire was conducted in September 2018. The data collection targeted pedestrians who used the YKD metro station to enter/leave the study area. The questionnaire asked details of the egress or assess trips such as origin and destination (OD), start time, end time, the exact route involved, and trip purpose. The trip purposes include work/school, leaving/returning home, shopping, visiting others, recreation (restaurants, entertainment, parks, public facilities, etc.), services (banks, government, post office, etc.), and other. The routes were drawn on a map. In addition to trip information, respondents were invited to report their age, gender, whether they work close to the station and for how many years, and whether they live close to the station and for how many living years. The data involved information about 515 trips of 302 randomly selected respondents.

The third data collection, conducted in November 2018, concerned the street-level built environment and road network. The street-level built environment data were mainly collected from topographic maps from the Tianjin Government, the 2018 Baidu satellite map, and the 2018 Baidu street view map. The data included building area, land use, length of building lot coverage (BLC), block area, sidewalk length, sidewalk width, number of lamps, road width, and the number of vehicle lanes on each road. Fence presence (physical separation between pedestrians and vehicles/cyclists) and traffic lights at intersections were collected from the 2018 Baidu street view map. Street greenery areas were collected from the 2018 Baidu satellite map. The selection of these variables is based on the literature and the main considered street-level elements in the practice of urban planning and design.

3 Simulation approach

To explore the impact of the street-scale built environment on the distribution of pedestrian flows, a Netlogo simulation was developed and applied. The simulation concentrated on pedestrian access/egress trips to the metro station. The route choice of each pedestrian was simulated using a latent class route choice model.

The simulation involves five steps: 1) setting up the road network and associated built environment attributes in the study area; 2) generating the simulated pedestrians and OD pairs for each pedestrian; 3) defining scenarios to explore their effect on pedestrian flows; 4) predicting the route choice of each pedestrian based on the latent class route choice model; 5) conducting uncertainty analysis. Figure 1 shows these steps.

3.1 Setting up the road network and built environment attributes

The pedestrian road network was drawn manually in an ArcGIS shapefile using a sidewalk and vehicle road layer. Two-sided sidewalks on roads with more than two vehicle lanes were drawn as two separate links, while two-sided sidewalks on roads with two or less than two vehicle lanes were drawn as a single link. For these roads, we assume that pedestrians are influenced by the built environment on both sides, as pedestrians can easily cross the roads.

The links were spatially attached to four shapefile layers (building layer, block layer, sidewalk layer, and vehicle road layer). For the two-sided sidewalks presented by a single link, the properties of the building and block layer on both sides of the link were summed as properties of the link. The properties number of lamps respectively street greenery area in the sidewalk layer were also summed as single properties of the link. The properties sidewalk length, sidewalk width, fence presence, and traffic lights at intersections in the sidewalk layer and road width in the vehicle road layer on both sides of the sidewalks were averaged. For roads with more than two vehicle lanes, represented by two links, all properties in the building, block, and sidewalk layers on each side of the road were spatially attached to the link on the corresponding side.

The built environment attributes of each link are shown in Table 1. The link-level data were converted to route-level data for pedestrian road choice modeling. The road-level built environment attributes are listed in Table 2. Land use mix is defined to include the vertical and horizontal dimensions and was calculated as defined in Eq. (1) (Cerin, Macfarlane, Ko, & Chan, 2007; Lau, Giridharan, & Ganesan, 2005; Song, Merlin, & Rodriguez, 2013).

Table 1 Built environment attributes at the link level

Attributes	Description	Levels and units
Link length	The length of a link.	A continuous number in meters.
Presence of fence	A fence along both sides of the link is physically splitting the sidewalk and vehicle lanes/ cycling lanes.	Binary variable: 1, the link has a fence. 0, does not have a fence.
Width of sidewalk	The average width of the sidewalks on the two sides of a link.	A continuous number in meters.
Gross floorage of each type of land use	The gross floorage of the buildings with the same land use along two sides of a route.	A continuous number in m ² .
Building front length	The gross length of the front cover by buildings along two sides of a link.	A continuous number in meters.
Total building area	The entire gross floorage of buildings on the two sides of a link.	A continuous number in m ² .
Block area	Total block area along a link.	A continuous number in m ² .
Street greenery area	The gross area of trees, grasslands, etc. on the two sides of a link.	A continuous number in m ² .
Traffic lights presence	The presence of traffic signal lights at both endpoints (intersections) of a link.	Categorical variable: 2, both endpoints have traffic signal lights; 1, one endpoint has traffic signal lights; 0, both endpoints have no traffic signal lights.
Number of lamps	The total number of street lamps on the two sides of a link.	A positive integer
Width of the road	The width between building restriction lines, including sidewalk width, vehicle lane width, and related separation facility width on a link.	A continuous number in meters.

$$\text{Land use mix} = -1 * \left[\sum_{h=1}^H \left(\frac{P_h}{P} \right) * \ln \left(\frac{P_h}{P} \right) \right] / \ln(H) \quad h = 1, 2, 3, \dots, 10 \quad (1)$$

where h is a land use category, P_h is the gross floorage of land use h in m², P is the gross floorage of all studied land uses in m².

The Path Size Correction factor (PSC) is a specific factor in the Path Size Correction (PSC) model that

represents the influence of overlapping routes on route choice probabilities (Prato, 2009). In this study, it is calculated as defined in Eq. (2).

$$PSC_l = - \sum_{k \in l} \left[\frac{L_k}{L_l} \cdot \ln \sum_{a \in C} \delta_{ka} \right] \quad (2)$$

where link k belongs to route l , L_k is the length of link k , L_l is the length of route l , δ_{ka} equals 1 if route a includes

Table 2 Built environment attributes at the route level

Estimated variables	Input data from Table 1	Units at route-level
Length of the route	Sum of the variable across all links of the route.	A continuous number in meters.
Street greenery area/1000		A continuous number in m ² .
Number of lamps		A continuous number.
Presence of fence	Sum of the variable, weighted by the length of each link.	A continuous number between 0 and 1.
Width of sidewalk		A continuous number in meters.
Number of traffic lights		A continuous number from 0 to 2.
Width of the road		A continuous number in meters.
Land use mix	Eq. (1)	A continuous number between 0 and 1.
BLC	Total "building front length" of each link divided by route length	A continuous number between 0 and 2.
The floorage ratio (FAR) on each route	Total "building area" of each link, divided by total "Block area" of each link.	A continuous number.
PSC	Eq. (2)	A continuous number.

link k and is equal to zero otherwise, C is the choice set of the alternative routes.

3.2 Generating simulated pedestrians

The number of simulated pedestrians is based on the video data collected in 2016. Thus, the route choice of 5040 pedestrians was simulated in the morning (8:00–9:00), 6810 pedestrians were simulated for the noon period (12:00–13:00), while the simulation involved 11,220 pedestrians for the afternoon period (17:00–18:00). Each simulated pedestrian needs to be characterized in terms of the selected socio-demographic characteristics, while in addition trip purpose, and OD pair need to be imputed from the travel survey data collected in 2018. Age, gender, working years close to the station, living years close to the station, OD pair, and trip purpose of the simulated pedestrians were generated using Monte Carlo draws from the observed profiles of the sampled pedestrians. This approach guarantees that any correlations between socio-demographics, OD pairs, and travel purposes are retained in the simulations.

Table 3 shows the distribution of age, gender, if they work nearby the station and how many working years, and if they live nearby the station and how many living years. Seven trip purposes were distinguished in the data collection, but to simplify the simulation only three were used: work/school, going home, and going shopping (which includes the other categories). The revealed data showed that different trip purposes are distributed quite differently at different times of the day. There were more work/school trips in the morning than in the afternoon, and more shopping trips at noon compared to the morning. The distributions are presented in Table 4.

3.3 Defining the scenarios

In 2019, the Tianjin government planned a series of projects to improve the built environment of the study area, especially for the roads highlighted in Fig. 2 (Tianjin Bureau of Commerce, 2019). Based on these projects, we developed 4 new design scenarios by changing the street-level built environment of 6 links. Three links are named Chifeng Road-1 (C-1), Chifeng Road-2 (C-2), and Chifeng Road-3 (C-3). Three links named Harbin Road-1 (H-1), Harbin Road-2 (H-2), and Harbin Road-3 (H-3) are on Harbin Road. Chifeng Road which is directly connected with the metro station is the main road. Harbin Road does not directly connect with the metro station. The 4 scenarios are shown in Table 5. A baseline scenario without any changes to the built environment serves as a benchmark.

The detailed new designs are shown in Table 6. The 4 new design scenarios add fences between sidewalks and vehicle lanes, crossing facilities, and increase sidewalk width in line with the main goals of the Tianjin government's projects. Besides these changes, links C-1, C-2, and C-3 in the Chifeng Road with land use mix change scenario additionally increase land use mix by about 10% with more restaurants and entertainment shops reflecting Tianjin government's intentions. As shown in Table 6, all 6 links will have a fence between sidewalks and vehicle lanes. All 6 links will have traffic lights at both endpoints. The "baseline" column in Table 6 is the current situation of each link in the study area. The "New design" column in Table 6 represents the new value of each variable on the link.

Table 3 Distribution of pedestrian characteristics

	Category	Percentage
Age	15 to 19 years	11.6%
	20 to 29 years	50.0%
	30 to 39 years	19.2%
	40 to 73 years	19.2%
Gender	Female	60.3%
	Male	39.7%
If they work close to the station and for how many working years	Do not work close	71.2%
	0 to 1 year	4.3%
	More than 1 year	24.5%
If they live close to the station and for many living years	Do not live close	72.5%
	0 to 1 year	7.0%
	More than 1 year	20.5%
Trip purpose	Going to work/school	20.8%
	Going home	32.0%
	Going shopping and others	47.2%

Table 4 Distribution of trip purpose by time of day

	Trip purpose: go to work/school	Trip purpose: go home	Trip purpose: go shopping and other
Morning	73%	10%	17%
Noon	11%	20%	69%
Afternoon	6%	54%	40%

3.4 Simulating route choice

The core of the simulation is the route choice model. In this study, a latent class route choice model that accounts for overlapping routes was used. The PSC latent class logit model is a combination of the path size correction logit model and latent class discrete choice model and considers the effect of overlapping routes and pedestrian preference heterogeneity. The probability P_{ns1} of pedestrian n belonging to class 1 is shown in Eq. (3), while the probability P_{ns2} for the same pedestrian belonging to class 2 is shown in Eq. (4).

$$P_{ns1} = \frac{\exp(V_{n|s1})}{\exp(V_{n|s1}) + \exp(V_{n|s2})} \quad n = 1, \dots, N \quad (3)$$

$$P_{ns2} = \frac{\exp(V_{n|s2})}{\exp(V_{n|s1}) + \exp(V_{n|s2})} \quad n = 1, \dots, N \quad (4)$$

$$V_{n|s1} = \beta_{gender|s1}X_{gender} + \beta_{age|s1}X_{age} + \beta_{work|s1}X_{work} + \beta_{live|s1}X_{live} + \beta_{purpose1|s1}X_{purpose1} + \beta_{purpose2|s1}X_{purpose2} + constant_{s1} \quad (5)$$

Fig. 2 Scenarios in the study area

Table 5 Scenarios

Scenario	Harbin Road	Chifeng Road
Baseline scenario	Baseline	Baseline
Harbin Road scenario	New design	Baseline
Chifeng Road without land use mix change scenario	Baseline	New design
Chifeng Road with land use mix change scenario	Baseline	New design
Harbin Road & Chifeng Road without land use mix change scenario	New design	New design

$$V_{n|s2} = \beta_{gender|s2}X_{gender} + \beta_{age|s2}X_{age} + \beta_{work|s2}X_{work} + \beta_{live|s2}X_{live} + \beta_{purpose1|s2}X_{purpose1} + \beta_{purpose2|s2}X_{purpose2} + constant_{s2} \quad (6)$$

where $V_{n|s1}$ is the utility of pedestrian n belonging to class 1, $V_{n|s2}$ is the utility of pedestrian n belonging to class 2. $\beta_{gender|s1}$, $\beta_{age|s1}$, $\beta_{work|s1}$, and $\beta_{live|s1}$ are the parameters of gender, age, working in the study area or not, and living in the study area or not for class 1. $\beta_{purpose1|s1}$ is the parameter for trip purpose going to work/school for class 1, $\beta_{purpose2|s1}$ is the parameter for trip purpose going home for class 1. For class 2, $\beta_{gender|s2}$, $\beta_{age|s2}$, $\beta_{work|s2}$, and $\beta_{live|s2}$ represents the effects of gender, age, working in the study area or not, and living in the study area or not. $\beta_{purpose1|s2}$ is the parameter for trip purpose going to work/school, while $\beta_{purpose2|s2}$ is the parameter for trip purpose going home. Because the categorical variables were effect-coded, the effect of shopping and other can be derived by summing the parameters for the other two trip purposes.

The probability of pedestrian n of class 1 choosing route k from choice set D is $P_{nk|s1}$, while $P_{nk|s2}$ is the probability of agent n of class 2 choosing route k from choice set D . These probabilities can be expressed as

$$P_{nk|s1} = \frac{\exp(V_{nk|s1} + \beta_{PSC} \cdot PSC_k)}{\sum_{l \in D} \exp(V_{nl|s1} + \beta_{PSC} \cdot PSC_l)} \quad n = 1, \dots, N \quad (7)$$

$$P_{nk|s2} = \frac{\exp(V_{nk|s1} + \beta_{PSC} \cdot PSC_k)}{\sum_{l \in D} \exp(V_{nl|s2} + \beta_{PSC} \cdot PSC_l)} \quad n = 1, \dots, N \quad (8)$$

$$V_{nk|s1} = \beta_{1|s1}X_1 + \beta_{2|s1}X_2 + \beta_{3|s1}X_3 + \beta_{4|s1}X_4 + \beta_{5|s1}X_5 + \beta_{6|s1}X_6 + \beta_{7|s1}X_7 + \beta_{8|s1}X_8 + \beta_{9|s1}X_9 + \beta_{10|s1}X_{10} \quad (9)$$

$$V_{nk|s2} = \beta_{1|s2}X_1 + \beta_{2|s2}X_2 + \beta_{3|s2}X_3 + \beta_{4|s2}X_4 + \beta_{5|s2}X_5 + \beta_{6|s2}X_6 + \beta_{7|s2}X_7 + \beta_{8|s2}X_8 + \beta_{9|s2}X_9 + \beta_{10|s2}X_{10} \quad (10)$$

where $V_{nk|s1}$ is the total utility of route k in the latent class 1, $V_{nl|s1}$ is the total utility of route l in latent class 1, l is a route in choice set D , β_{PSC} is the estimated coefficient for the PSC factor, PSC_k is the PSC value for route k , PSC_l is the PSC value of route l in set D , $x_1 \dots x_{10}$ are the 10 built environment variables listed in Table 7. $\beta_{1|s1} \dots \beta_{10|s1}$ are the corresponding parameters of the ten built environment attributes in class 1, $\beta_{1|s2} \dots \beta_{10|s2}$ are the corresponding parameters of the ten built environment attributes in class 2.

The actual simulation involved the following steps: for each simulated pedestrian and each trip.

- (1) The membership probability was calculated as a function of gender, age, work location, residential location, and trip purpose.

Table 6 Built environment changes on each link

Road	Links	Fence		Sidewalk width (m)		Traffic lights		Land use mix	
		Base line	New design	Base line	New design	Base line	New design	Base line	New design
Harbin Road	H-1	0	1	1.5	1.8	0	2	0.70	0.70
	H-2	0	1	3	3.6	1	2	0.66	0.66
	H-3	0	1	1.5	1.8	1	2	0.65	0.65
Chifeng Road without land use mix change	C-1	0	1	2	2.4	1	2	0.69	0.69
	C-2	0	1	2	2.4	1	2	0.71	0.71
	C-3	0	1	2	2.4	2	2	0.74	0.74
Chifeng Road with land use mix change	C-1	0	1	2	2.4	1	2	0.69	0.76
	C-2	0	1	2	2.4	1	2	0.71	0.78
	C-3	0	1	2	2.4	2	2	0.74	0.81

Table 7 Parameters of the route choice model

Estimated variables	Class 1	Class 2
	Parameter	Parameter
Total route length	−0.065***	0.008**
Street greenery area/1000	−0.947	−0.116
Number of lamps	0.073	0.050
Fence presence	12.105**	5.805**
Sidewalk width	−1.277***	0.315
Number of traffic lights	−7.762***	1.480
Road width	−0.251***	0.139***
Land use mix	5.470	−0.093
Building lot coverage (BLC)	5.614	12.927***
FAR	−4.860***	−3.010**
PSC	1.555***	1.555***
Class membership		
Constant	0.549**	0
Female	0.069	0
Male	−0.069	0
Age lower than 23 years old	0.038	0
Age higher than 23 years old	−0.038	0
Work nearby the station	0.868***	0
Not work nearby the station	−0.868	0
Live nearby the station	0.060	0
Not nearby the station	−0.060	0
Trip purpose: work/school	0.131	0
Trip purpose: home	−0.030	0
Trip purpose: shopping and other	−0.101	0

Source: Liu et al., 2020. Note: *** indicates significant values at the 1% level. ** indicates significant values at the 5% level

- (2) A Monte Carlo drawn was made from these membership probabilities to simulate the latent class of that pedestrian.
- (3) Given the simulated latent class, the route choice probabilities were calculated using the route choice model for each route in the pedestrian's choice set.
- (4) A Monte Carlo drawn was made to simulate the chosen route.

3.5 Uncertainty analysis

Because the route choice model is a probabilistic model, every set of Monte Carlo draws may lead to a different simulated route. Thus, the results of the simulation are prone to simulation error. In order to quantify the resulting uncertainty, in this study, an uncertainty analysis was conducted at the link level by time of day. Coefficients of variation (CV) of the pedestrian flow on each link and time of day, based on n runs were calculated as shown in Eq. (11). The CV was calculated after respectively 50, 100, 150, 200, 250, and 300 runs.

$$CV_{kt|n} = \sigma_{kt|n} / \bar{\mu}_{kt|n} \quad n = 50, 100, 150, 200, 250, 300 \quad (11)$$

where $CV_{kt|n}$ is the coefficient of variation based on n runs of t link k at time of day t , $\sigma_{kt|n}$ is the standard deviation of the pedestrian flows based on n runs of link k at time of day t , $\mu_{kt|n}$ is mean value pedestrian flow based on n runs of link k at time of day t ,

In addition, based on Miller's (1992), 95% confidence levels were calculated as

$$CI_{kt|n} = \frac{s}{\bar{x}} \pm z_{\alpha/2} \sqrt{(n-1)^{-1} \left(\frac{s}{\bar{x}} \right)^2 \left(0.5 + \left(\frac{s}{\bar{x}} \right)^2 \right)} \quad n = 50, 100, 150, 200, 250, 300 \quad (12)$$

where $CI_{kt|n}$ is the confidence interval on n runs of link k at time of day t , $\frac{s}{\bar{x}}$ is equal to $\sigma_{kt|n} / \mu_{kt|n}$ when sample x is an asymptotic normal distribution with mean μ and variance $\sigma_{kt|n}^2$, s is the sample variance, $z_{\alpha/2}$ is the 95% percentile of the standard normal distribution, n is the number of runs, and α is 0.05 for the common 95% probability level.

4 Scenario analysis

4.1 Simulation results

4.1.1 Results of the baseline scenario

First, the pedestrian flows for the Baseline scenario were simulated. The simulation was run separately for the three different time periods (morning, noon, and afternoon). For each time period, the simulation was repeated 200 times. The simulated pedestrian flows are partly shown in Fig. 3 (top part) and the absolute number of pedestrians on the 6 links are shown in Table 8. The width of the red links represents the magnitude of the simulated pedestrian flows. The thicker the red link, the more pedestrians were simulated to pass that link. The distribution of pedestrian flows across the different time periods shows the higher accumulation of pedestrians near the metro station and less pedestrians further away from the metro station. The number of pedestrian flows is calculated based on the average number of 200 runs. Table 8 only shows the pedestrian flows on the 6 links. Links H-2 and H-3 have more pedestrians compared to links C-1, C-2, C-3, and H-1. The results for the other links in Fig. 3 (bottom part) in the whole study area are not shown. The destinations/origins of each trip that have the station either as origin or destination are shown in Fig. 3 as green circles. Bigger circles reflect a higher number of pedestrians who have this link as their destination/origin. As expected, the biggest circle is observed for the metro station.

Fig. 3 Pedestrian flow on each link by time period and the distribution of destinations/origins

4.1.2 Results of the 4 scenarios

The simulated pedestrian flows for each time period (morning, noon, and afternoon) under the 4 hypothetical scenarios are shown in respectively. Figures 4, 5, 6, and 7. The horizontal axis is the link. The vertical axis is the absolute number of simulated pedestrians. The numbers on the bars represent the difference between the simulated

pedestrian flow on each link under the considered scenario compared and the baseline scenario.

Figure 4 indicates that for the Harbin Road Scenario in the morning (green color) the differences in pedestrian flows with the baseline are very small (from -0.2 to $+4$) for all 6 links. It indicates that at this time of day the pedestrian flows on the 6 links are hardly affected by

Table 8 Absolute number of pedestrians on the 6 links by time of day

	Morning	Noon	Afternoon
C-1	33.8	58.0	120.7
C-2	31.0	55.4	116.0
C-3	29.0	78.6	122.7
H-1	43.2	122.8	190.7
H-2	113.2	333.4	542.3
H-3	170.0	497.4	776.3

the Harbin Road scenario. The same applies to the noon period. In the afternoon (grey color), all reported links attract more pedestrians (from + 21.3 to + 276) compared to the baseline scenario. The increase in the number of pedestrians is much higher for links H-1, H-2, and H-3 than for C-1, C-2, and C-3. Thus, the simulated impact of this scenario is higher for Harbin Road than for Chifeng Road.

In the Chifeng Road without land use mix change Scenario (Fig. 5), the number of pedestrians is simulated to increase for links C-1, C-2, and C-3, but to decrease for links H-1, H-2, and H-3 compared to the baseline scenario. The increase is slightly higher on C-3 than C-1 and C-2. Thus, the improvements on the built environment of Chifeng Road negatively affect the pedestrian flows on the Harbin Road especially for H-1 and H-2 links. The increase in the number of pedestrians on Chifeng Road exceeds the decrease in the number of pedestrians on Harbin Road. It suggests that pedestrians from other links would be attracted to Chifeng Road. The impact has a similar pattern for all times of day, but changes are bigger in the afternoon than in the morning and noon.

In the Chifeng Road with land use mix change Scenario (Fig. 6), the changes in pedestrian flows are very similar to the simulated changes under the Chifeng Road without land use mix change Scenario (Fig. 5). Upon reflection, this negligible difference can be explained by the very small parameter estimates for land use mix, while the latent class membership probabilities are the same for the two Chifeng scenarios.

In the Harbin Road & Chifeng Road without land use mix change scenario (Fig. 7), the number of pedestrians is simulated to increase for links C-1, C-2, C-3 in the three time periods. The increase for links H-1, H-2, H-3 only happens in the afternoon period. The number of pedestrians decreases on links H-1, H-2, in the morning and H-1, H-2, H-3 at noon compared to the baseline scenario. The increase is much higher on links C-1, C-2, and C-3 than on links H-1, H-2, and H-3 in each time period. The results indicate that the improvement of the built environment of both Chifeng Road and Harbin Road is simulated to positively affect the pedestrian flows on Chifeng Road all day and the pedestrian flows on Harbin Road mainly in the afternoon. The improvement has negative effects on pedestrian volumes of Harbin Road in the morning and noon. Based on the changing number of pedestrians, the simulated impact is higher for Chifeng Road than for Harbin Road.

4.2 Uncertainty analysis

To quantify the uncertainty in the simulated pedestrian flows due to simulation error, coefficients of variation (CV) were calculated for each link and time of day after respectively 50, 100, 150, 200, 250, and 300 runs. Table 9 and Fig. 8 presents the results for the morning time

Fig. 4 Results of the Harbin Road scenario

Fig. 5 Results of the Chifeng Road without land use mix change Scenario

period for each link under the Baseline scenario. Results for other time periods and the other scenarios are in the same order of magnitude. As seen in Table 9, the CVs for each link do not change much beyond 150 runs and become stable after 200 runs. The CVs on all links are smaller than 9.3%. Link C-3 has higher CVs than the other links. At the maximum number of 300 runs, the confidence intervals are around 3%. Thus, overall, uncertainty due to simulation error is small.

4.3 Interpretation

The simulated results suggest that pedestrian flows on the main roads near the metro station are affected by the simulated changes in the built environment. The Harbin Road & Chifeng Road without land use mix change scenario suggests that the simulated impact is higher for the main road which is directly connected with the metro station. The simulated

Harbin Road scenario has a very limited impact on the distribution of pedestrian flows in both the morning and noon time period. This can be explained by differences in trip purpose and socio-demographics of the pedestrians for the different time periods. In the morning, the largest share of pedestrians concern working/school pedestrians. They have a higher probability to belong to class 1, for which the increased utility of the presence of a fence is canceled out by the decreased utility of increasing sidewalk width and adding more traffic lights. At noon, there were more pedestrians for the purpose of shopping and other, who have a higher probability to belong to class 2. This class more likely shops at the large shopping malls on the west side of Harbin Road. The shortest path to reach the west side of Harbin road is from the main road next to the metro station or via Chifeng Road to the north part of the shopping street.

Fig. 6 Results of the Chifeng Road with land use mix change Scenario

Fig. 7 Results of the Harbin Road & Chifeng Road without land use mix change scenario

In the afternoon, the Harbin Road scenario has a positive influence on pedestrian flows on all 3 links of Harbin Road and the 3 links of Chifeng Road. Thus, the improvement of the streetscape Harbin Road has synergistic effects on Chifeng Road. The increased attraction of Harbin Road seems to lead to more pedestrians traveling through the links of Chifeng Road. Especially, there are more pedestrians going home (54%) and “shopping and other” pedestrians (40%) in the afternoon. These pedestrians have a higher probability of belonging to class 2.

The Chifeng Road with land use mix change scenario and Chifeng Road without land use mix change scenario were predicted to have a positive influence on pedestrian flows of Chifeng Road in the three time periods although

the difference varies across time periods. Moreover, the changes in the Chifeng Road built environment decreased the pedestrian flow in Harbin Road which indicates a competitive relationship between these two roads. The most important reason might be the spatial location of the two roads. Links H-1, H-2, H-3 of Harbin Road are not directly linked to the metro station; pedestrians need to pass through the links of Chifeng Road. The routes passing the links of Harbin Road, especially H-1 and H-2, have to pass at least C-2 or C-3. However, the routes from/to the northern part of Chifeng Road do not necessarily include Harbin Road.

The small differences between the two scenarios can be interpreted directly from the relative small parameter

Table 9 Coefficients of variation and confidence intervals of each link in the morning period by number of runs for the baseline scenario

Link	50 runs	100 runs	150 runs	200 runs	250 runs	300 runs
C-1	2.27% (0.0182 ~ 0.0271)	1.89% (0.0163 ~ 0.0215)	1.87% (0.0166 ~ 0.0208)	1.81% (0.0163 ~ 0.0199)	1.77% (0.0162 ~ 0.0193)	1.86% (0.0171 ~ 0.0201)
C-2	4.67% (0.0376 ~ 0.0559)	3.99% (0.0343 ~ 0.0454)	3.79% (0.0336 ~ 0.0421)	3.78% (0.0341 ~ 0.0415)	3.76% (0.0343 ~ 0.0409)	3.96% (0.0365 ~ 0.0428)
C-3	9.32% (0.0749 ~ 0.1114)	8.24% (0.0710 ~ 0.0939)	7.70% (0.0683 ~ 0.0857)	7.82% (0.0705 ~ 0.0859)	7.86% (0.0717 ~ 0.0854)	8.15% (0.0750 ~ 0.0880)
H-1	0.64% (0.0051 ~ 0.0076)	0.70% (0.0060 ~ 0.0080)	0.70% (0.0062 ~ 0.0078)	0.79% (0.0072 ~ 0.0087)	0.86% (0.0078 ~ 0.0093)	0.83% (0.0076 ~ 0.0089)
H-2	0.87% (0.0070 ~ 0.0104)	0.86% (0.0074 ~ 0.0098)	0.82% (0.0073 ~ 0.0091)	0.87% (0.0078 ~ 0.0095)	0.86% (0.0079 ~ 0.0094)	0.87% (0.0080 ~ 0.0094)
H-3	1.53% (0.0123 ~ 0.0183)	1.72% (0.0148 ~ 0.0196)	1.77% (0.0157 ~ 0.0197)	1.78% (0.0161 ~ 0.0196)	1.75% (0.0160 ~ 0.0191)	1.71% (0.0157 ~ 0.0185)

Note: confidence intervals in parentheses

Fig. 8 Uncertainty analysis of the baseline scenario in the morning period

for land use mix and the small differences in land use mix. It should be realized, however, that in this simulation we only capture the influence of varied land use on route choice and that the change in land use mix is very small. Past studies only found that walking trip frequency and destinations have a positive relationship with the higher land use mix (Day, 2016; Saelens & Handy, 2008) but have less evidence on land use mix and pedestrian route choice.

5 Discussion

In completing this paper, a number of limitations should be mentioned. First, the simulation only concerned changing the route choice of pedestrian from/to the metro station. It did not consider pedestrians that enter or leave the study area from other locations. Of course, the route choice behavior of these other pedestrians may also be affected by the changes in the built environment. Thus, the simulated number of pedestrians for each link and their shares of the pedestrian volume is not representative of the area. This is also the reason why the simulations in the baseline scenario cannot be validated against pedestrian counts. If the interest would be in the total pedestrian system in the study area, the data should be collected at every entry/exit point in the study area.

Second, we only considered access and egress trips. In reality, however, a pedestrian may be involved in multi-stop (and multi-purpose) trips. This will influence the distribution of pedestrian flows in the study area. The inclusion of multi-stop, multi-purpose trips means that appropriate data on such trips need to be selected and that the current simple route choice model should be replaced with a much more complex model of route choice behavior under multi-stop, multi-purpose trips.

Third, we did not consider the possible effect of changing land use and street characteristics on destination

choice, not at the city level, not within the study area. It means that the results only depict the changing route choice behavior of a selected group of pedestrians conditional on invariant destination choice.

Fourth, it should be realized that simulation error is only one source of error. An error may also be introduced by model specification. Assessing its impact requires the comparison of different model specifications, based on different assumptions underlying pedestrian route choice behavior. Another source of error is the demarcation of the choice sets. Future research should examine the effect of different specifications of choice set on simulated pedestrian flows.

Finally, the estimated results of the route choice model emphasize the relative importance of the presence of a fence. Consequently, some distinctive simulation results are the direct consequence of this parameter. However, when we look into the data, it became clear that most links directly surrounding the metro station where we observed most pedestrians had a fence. Hence, the distribution of data points in the space spanned by the presence of fence and the number of observed pedestrians on a route not far from uniformly distributed. Consequently, the estimated parameter for the presence of fence is likely biased. This is a clear limitation of revealed preference models.

Acknowledgements

The authors wish to acknowledge the valuable help and support from Dr. Jie He, Tianjin University; data collection effort of Yu Wang and the student members in the Spatial Humanities and Place Computation Laboratory (SHAPC Lab) and School of Architecture in Tianjin University. Yanan Liu holds a scholarship from the China Scholarship Council.

Code availability

The code used in the analyses during the current study is available from the corresponding author on a reasonable request.

Authors' contributions

I confirm that all authors listed on the title page have contributed significantly to the work, have read the manuscript, attest to the validity and legitimacy of the data and its interpretation, and agree to its submission. All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

Funding

The corresponding author is financially supported by China Scholarship Council. This research is not supported by any other funding.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue. The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript. The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent/licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Received: 10 November 2020 Accepted: 8 December 2020

Published online: 29 March 2021

References

- Borgers, A. W. J., & Timmermans, H. J. P. (2015). Modeling pedestrians' shopping behavior in downtown areas. In *CUPUM 2015 -14th International Conference on Computers in Urban Planning and Urban Management* http://web.mit.edu/cron/project/CUPUM2015/proceedings/Content/modeling/202_borgers_h.pdf.
- Borgers, A. W. J., & Timmermans, H. J. P. (2014). Indices of pedestrian behavior in shopping areas. *Procedia Environmental Sciences*, 22, 366–379.
- Cerin, E., Macfarlane, D. J., Ko, H. H., & Chan, K. C. A. (2007). Measuring perceived neighbourhood walkability in Hong Kong. *Cities*, 24(3), 209–217.
- Clifton, K. J., Smith, A. D. L., & Rodriguez, D. (2007). The development and testing of an audit for the pedestrian environment. *Landscape and Urban Planning*, 80(1–2), 95–110.
- Chen, X., Li, H., Miao, J., Jiang, S., & Jiang, X. (2017). A multiagent-based model for pedestrian simulation in subway stations. *Simulation Modelling Practice and Theory*, 71, 134–148.
- Daamen, W., & Hoogendoorn, S. P. (2003a). Controlled experiment to derive walking behavior. *European Journal of Transport and Infrastructure Research*, 3(1), 39–59.
- Daamen, W., & Hoogendoorn, S. P. (2003b). Experimental research of pedestrian walking behavior. *Transportation Research Record*, 1828, 20–30.
- Day, K. (2016). Built environmental correlates of physical activity in China: a review. *Preventive Medicine Reports*, 3, 303–316.
- Dijkstra, J., Jessurun, A. J., Timmermans, H. J. P., & de Vries, B. (2011). A framework for processing agent-based pedestrian activity simulations in shopping environments. *Cybernetics and Systems*, 42(7), 526–545.
- Guo, Z., & Loo, B. P. Y. (2013). Pedestrian environment and route choice: evidence from New York City and Hong Kong. *Transport Geography*, 28, 124–136.
- Lau, S. S. Y., Giridharan, R., & Ganesan, S. (2005). Multiple and intensive land use: case studies in Hong Kong. *Habitat International*, 29, 527–546.
- Li, F., Fisher, K. J., Brownson, R. C., & Bosworth, M. (2005). Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults. *Journal of Epidemiol & Community Health*, 59, 558–564.
- Lin, H. (2015). *The influence of built environment on walking behavior: measurement issues, theoretical considerations, modeling methodologies and Chinese empirical studies*. Dordrecht: Springer.
- Liu, Y., Yang, D., Timmermans, H. J. P., & de Vries, B. (2020). The impact of the street-scale built environment on pedestrian metro station access/egress route choice. *Transportation Research Part D: Transport and Environment*, 87, 1–9.
- Lue, G., & Miller, E. J. (2019). Estimating a Toronto pedestrian route choice model using smartphone GPS data. *Travel Behaviour and Society*, 14, 34–42.
- Miller, V. (1992). Inflation uncertainty and the disappearance of financial markets: the Mexican example. *Journal of Economic Development*, 17, 131–152.
- Moran, M. R., Rodriguez, D. A., & Corburn, J. (2018). Examining the role of trip destination and neighborhood attributes in shaping environmental influences on children's route choice. *Transportation Research Part D: Transportation and Environment*, 65, 63–81.
- Nuworsoo, C., Cooper, E., Cushing, K., & Jud, E. (2013). Considerations for integrating bicycling and walking facilities into urban infrastructure. *Transportation Research Record*, 2393(1), 125–133.
- Prato, C. G. (2009). Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, 2(1), 65–100.
- Rodriguez, A. D., Merlin, L., Prato, G. C., Conway, L. T., Cohen, D., Elder, P. J., Evenson, R. K., McKenzie, L. T., Pickrel, L. J., & Mortenson, V. S. (2015). Influence of the built environment on pedestrian route choices of adolescent girls. *Environment and Behavior*, 47(4), 359–394.
- Rose, J., Ligtenberg, A., & van der Sperk, S. (2014). Simulating pedestrians through the inner-city: an agent-based approach. In Q. Miguel, F. Amblard, J. Barceló, & M. Madella (Eds.), *Advances in computational social science and social simulation*. Barcelona: September 1–5. <http://ddd.uab.cat/record/125597>.
- Saelens, B. E., & Handy, S. L. (2008). Built environment correlates of walking: a review. *Medicine & Science in Sports & Exercise*, 40(7), 550–566.
- Schelhorn, T., O'Sullivan, D., Haklay, M., & Thurstain-Goodwin, M. (1999). STREETS: an agent based pedestrian model. *Proceedings of Computers in Urban Planning and Urban Management*.
- Song, Y., Merlin, L., & Rodriguez, D. (2013). Comparing measures of urban land use mix. *Computers, Environment and Urban Systems*, 42, 1–13.
- Tianjin Bureau of Commerce (2019). Notice of projects promoting "Jinjie" shopping area. <http://tianjin.mofcom.gov.cn/article/sjtongzhigg/201905/20190502867069.shtml>.
- Tilahun, N., & Li, M. (2015). Walking access to transit stations: evaluating barriers with stated preference. *Transportation Research Record: Journal of the Transportation Research Board*, 2534, 16–23.
- Tribby, C. P., Miller, H. J., Brown, B. B., Werner, C. M., & Smith, K. R. (2017). Analyzing walking route choice through built environments using random forests and discrete choice techniques. *Environment and Planning B: Urban Analytics and City Science*, 44(6), 1145–1167.
- Ukkusuri, S., Miranda-Moreno, L. F., Ramadurai, G., & Isa-Tavarez, J. (2012). The role of built environment on pedestrian crash frequency. *Safety Science*, 50(4), 1141–1151.
- Yang, J., Chen, J., Le, X., & Zhang, Q. (2016). Density-oriented versus development-oriented transit investment: decoding metro station location selection in Shenzhen. *Transport Policy*, 51, 93–102.
- Zacharias, J. (2001). Pedestrian behavior and perception in urban walking environments. *Journal of Planning Literature*, 6(1), 1–18.
- Zhu, W., & Timmermans, H. J. P. (2011). Modeling pedestrian shopping behavior using principles of bounded rationality: model comparison and validation. *Journal of Geographical Systems*, 13(2), 101–126.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.