

## Examining non-linear built environment effects on elderly's walking: A random forest approach

Long Cheng<sup>a</sup>, Jonas De Vos<sup>b</sup>, Pengjun Zhao<sup>c</sup>, Min Yang<sup>d,\*</sup>, Frank Witlox<sup>a,e</sup>

<sup>a</sup> *Department of Geography, Ghent University, Belgium*

<sup>b</sup> *Bartlett School of Planning, University College London, United Kingdom*

<sup>c</sup> *College of Urban and Environmental Sciences, Peking University, China*

<sup>d</sup> *School of Transportation, Southeast University, China*

<sup>e</sup> *Department of Geography, University of Tartu, Estonia*

\* *Corresponding author: Si Pai Lou #2, Nanjing, 210096, China; Email: [yangmin@seu.edu.cn](mailto:yangmin@seu.edu.cn)*

### Abstract

Previous studies often assume a pre-specified relationship between walking and the built environment. Using data from 702 older adults in Nanjing (China), this study employs a random forest method to scrutinise the refined associations between the built environment and walking time among older adults, while controlling for socio-demographics. Results show that all the analysed built environment attributes tend to impose salient non-linear and threshold effects on walking time. For example, population density and land use mix only increase older adults' walking at certain levels. Ultra-densely populated areas and excessively mixed land uses may even lead to a decline in walking. Built environment interventions would consequently only be effective in certain ranges. By quantifying the relative importance of built environment attributes, we further indicate the hierarchy of intervention priorities. These results offer nuanced and appropriate guidance for building age-friendly neighbourhoods.

**Keywords:** Built environment; Land use; Walking; Non-linear effect; Random forest; Active ageing

## 1. Introduction

Improved life expectancy over the past decades along with sustained low fertility is causing a remarkable shift in the distribution of population towards older ages in most regions of the world. According to the United Nations, older adults (65 years or above) consisted of 9% – 703 million – of the world population at the end of 2019. This share is projected to rise to 12% in 2030 and 16% in 2050 (United Nations, 2019). The increasingly ageing society has therefore highlighted the need for interventions which improve quality of life for older adults. Active travel, in particular walking, has been attracting policy attention as an important instrument to support health and wellbeing (Curl and Mason, 2019; Forsyth et al., 2007; Humphreys et al., 2013). Walking benefits active lifestyles and contributes to the functional capacity via the physical activity performed while moving. Since older adults are recommended to engage in at least 150 minutes of moderate physical activity per week in order to reduce the risk of non-communicable diseases (World Health Organisation, 2010), walking can be a good way for elderly people to stay healthy (Moniruzzaman et al., 2014). In addition, as a form of mobility, walking helps to access destinations/services and social engagement, which fosters a sense of independence, self-efficacy and wellbeing (Nordbakke and Schwanen, 2014).

Many studies have analysed the association between the built environment and walking among older adults. Higher exposure to well-designed urban environments – e.g. compact and mixed-use areas, parks and greenspaces, dedicated pedestrian infrastructure – increases the propensity to walk more (Cao et al., 2010; Cerin et al., 2014, 2017; Kemperman and Timmermans, 2009; Lee and Dean, 2018). However, despite a growing interest, most studies tend to involve a priori assumed relationships between the built environment and walking, with the application of statistical regression models (e.g. negative binomial regression (Cao et al., 2010), logistic regression (Barnes et al., 2016), joint discrete-continuous model (Moniruzzaman and Páez, 2016), and structural equation models (Leung et al., 2018)). In fact, their relationship may exist in a complicated non-linear manner. A recent review study indicates that in an ultra-dense city (e.g. Hong Kong), increasing residential density (beyond a certain threshold) may reduce walking frequency due to lower proximity to services and destinations, resulting in possible chaining of multiple trips in a single journey (Cerin et al., 2014).

Theoretically, two mechanisms may lead to the non-linear and threshold effects of the built environment on travel behaviour. The first relates to the peer effect (Walker et al., 2011) or collective socialisation (Galster, 2018) developed among travellers in a community. For density-related attributes, e.g. population density and residential density, only reaching some critical mass by a group could be effective in influencing others' travel behaviour (via imposing peer pressure). The second one refers to the effect of travel utility. Mokhtarian and Salomon (2001) noted that travel itself has a positive utility (e.g. movement control, a sense of speed, and enjoyment of scenery en-route) and may be desired for its own sake. As a result, a certain level of travel could be retained and these trips are difficult to be replaced by built environment interventions. Holden and Norland (2005) and Næss (2006) also observed a minimum level of travel with the compensatory hypothesis that reduced travel in compact areas will be compensated for by infrequent longer travel at other times. Against this backdrop, it is important to perform fine-grained analyses of the link between travel and the built environment to gain better insights into the optimal built environment profiles for effectively supporting elderly's walking. Overlooking the non-linear relationship may lead to ineffective and even erroneous policy interventions (Ding et al., 2018; Tao et al., 2020).

1  
2  
3  
4 Therefore, this study aims to disentangle the potential non-linearity of built environment effects on  
5 walking among older adults.  
6

7  
8 The present study contributes twofold to the literature. Empirically, it enriches the understanding of built  
9 environment-travel connections by examining non-linear and threshold effects between the built  
10 environment and older adults' walking. By revealing the refined relationships, we offer nuanced  
11 implications for appropriate environmental interventions and the efficient extent of policy initiatives to  
12 encourage walking among older people. Methodologically, the random forest method, a non-parametric  
13 approach, is adopted to capture complicated relationships of walking time with explanatory variables. It is  
14 proven to be more effective in revealing non-linear relationships compared to a conventional linear  
15 regression model with transformed independent variables.  
16  
17

18  
19 This paper is structured into six sections. Section 2 reviews the literature on the built environment effects  
20 on elderly's walking and the non-linear relationships between travel and the built environment. This is  
21 followed by Section 3 where the collected data and some descriptive statistics are presented. Section 4  
22 introduces the random forest method. Section 5 details research findings while Section 6 summarises the  
23 key results and discusses implications for planning practices.  
24  
25

## 26 27 28 **2. Literature review** 29

### 30 **2.1. Built environment and walking among older adults** 31

32 Walking is the most common form of physical activity for older adults, with substantiated benefits for  
33 health and wellbeing (Frank et al., 2010). The importance of the built environment in affecting walking  
34 among older people has been well documented in the literature (Gómez et al., 2010; Kemperman and  
35 Timmermans, 2009; Li et al., 2005; Nagel et al., 2008; Perchoux et al., 2019; van Cauwenberg et al., 2013).  
36 A meta-analysis shows a clear positive association between walking and various environmental elements,  
37 including the degree of urbanisation, residential density, land use diversity, street connectivity, pedestrian  
38 infrastructure, neighbourhood aesthetics and order, and access to services and destinations (e.g. public  
39 transport, parks/open spaces, and shops/commercial destinations) (Cerin et al., 2017).  
40  
41

42  
43 Neighbourhood-level population density is the most frequently analysed land use attribute in these studies  
44 and is found to increase the likelihood of choosing walking for transport (Böcker et al., 2017; Liao et al.,  
45 2018; Moniruzzaman et al., 2013). A higher level of land use mix is also shown to positively correlate with  
46 the propensity of older adults' walking in Canada (Moniruzzaman et al., 2013), the Netherlands (Böcker et  
47 al., 2017), and China (Leung et al., 2018). Hou (2019) analysed the influence of access to culture amenities  
48 (i.e. places of worship) on walking trips among seniors, revealing a positive association. There is evidence  
49 that older adults increase walking frequency and durations if their homes are nearby recreational  
50 amenities, e.g. park, square or greenspace (Cheng et al., 2019b; Leung et al., 2018). Infrastructure  
51 stimulating walking may comprise the presence of street lighting, walkway continuity, slip-resistant  
52 pavement, pedestrian crossing and signal, and protections from traffic with curb/barriers (Leung et al.,  
53 2018; Moniruzzaman and Páez, 2016). In addition, various built environment factors influencing older  
54 adults' walking may interact with each other, rather than exerting independent effects. For instance,  
55 walkability, a composite index, derived from multiple attributes including land use mix, residential density  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 and street connectivity is found to be a decisive predictor (Frank et al., 2010; Lee and Dean, 2018;  
5 Ramezani et al., 2019).  
6

7  
8 Ghani et al. (2018) and Cheng et al. (2019c) argue that the built environment has a stronger effect on older  
9 adults' walking than on walking of younger people. This is partly because of shrinking activity spaces of  
10 older people – resulting from the decline of physical capabilities – making them spend more time in their  
11 neighbourhoods (Mercado and Páez, 2009). Another possible reason is that older adults have a different  
12 perception of the built environment than the younger population (Haselwandter et al., 2015). They are  
13 more susceptible to certain environmental attributes as a result of the increased prevalence of physical  
14 and cognitive constraints (van Cauwenberg et al., 2013; Yang et al., 2018), e.g. absence of crosswalks or  
15 poorly maintained footpaths may prevent them from walking. As a consequence, a special focus on the  
16 built environment enhancement is essential to support walking among older people.  
17  
18  
19  
20  
21

## 22 **2.2. Non-linear relationships of travel and the built environment**

23  
24 Several recent studies have detected the non-linear and threshold relationships of travel and the built  
25 environment (Ding et al., 2018, 2019; Gan et al., 2020; Lewis, 2018; Tao et al., 2020), although a predefined  
26 functional form of the relationship is often assumed in the empirical research. Non-linearity suggests that  
27 the marginal effect of an explanatory variable on the outcome is dependent on the value of that variable  
28 (Galster, 2018). A threshold effect is a particular type of a non-linear relationship where the magnitude of  
29 impact changes – either increases or decreases – notably past a cut-off point. A built environment attribute  
30 may not have equivalent effects across the whole range of that attribute and thus finding the appropriate  
31 range would be cost-effective (van Wee and Handy, 2016).  
32  
33  
34

35  
36 The strong non-linearity of built environment effects on driving distance was observed in an empirical  
37 study of Oslo, Norway (Ding et al., 2018). The identified effective ranges of distance to the city centre and  
38 population density for reducing driving on weekdays are 0-20 km and 0-3000 persons/km<sup>2</sup>. Beyond these  
39 intervals, their marginal effects become trivial. Ding et al. (2019) substantiated a threshold effect of access  
40 to public transport on ridership using Metrorail data collected in the Washington Metropolitan Area. In  
41 the analysis, population density, employment density, land use mix, intersection density, and distance  
42 from CBD also show threshold effects on public transport ridership. Their effects get saturated when  
43 certain points are reached. With data from the Minneapolis and St. Paul Metropolitan Area, Tao et al.  
44 (2020) explored the non-linear connections between spatial attributes and walking distance to public  
45 transport. They found that walking distance stays relatively stable when job and household entropy – a  
46 measure of land use diversity – is in the interval of 0 to 0.7, and increases remarkably when the entropy  
47 rises from 0.7 to 1.0. Consequently, highly diversified land use (entropy reaches at least 0.7) around public  
48 transport stops is important to enlarge catchment areas of public transport services.  
49  
50  
51  
52

53  
54 The non-linear effects of the built environment on elderly's walking have been implied by Cerin et al.  
55 (2014). They found that additional residential density in highly compact areas would not increase – but  
56 even decrease – the odds of walking among older adults. Through a dichotomous variable of street  
57 connectivity, Perchoux et al. (2019) observed that street connectivity with a value lower than eight  
58 intersections in a 30-metre buffer was positively associated with the probability of walking in Luxembourg,  
59 while a negative relationship was found above eight intersections. Their study acknowledges the  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 complexity of built environment effects on elderly's walking and suggests the potential benefits of  
5 exploring threshold effects to inform more effective environmental interventions. Particularly,  
6 understanding how the built environment affects the overall relationship provides critical knowledge  
7 concerning the success of walking-supportive neighbourhood improvements. However, relevant  
8 exploration of the non-linear associations between the built environment and walking among older adults  
9 is rather limited.  
10  
11

### 14 3. Data

15  
16 To reveal the non-linear relationships between the built environment and walking among older adults, this  
17 study investigates their daily walking time – the total amount of time spent walking on the reference travel  
18 day – based on the 2013 Nanjing Household Travel Survey.<sup>1</sup> Nanjing, the capital of Jiangsu Province in  
19 China, is an interesting case study as walking constitutes a dominant share (61%) of elderly's travel. The  
20 proportion of older adults has been rapidly increasing in Nanjing. In 2006, 13% of residents were aged 60  
21 or older and the percentage increased to 21% with a population of 1.31 million in 2017 (Nanjing Municipal  
22 Bureau Statistics, 2018). According to the projection, the proportion will exceed 30% in 2030. The main  
23 city area of Nanjing is partitioned into 49 sub-districts, which are the basic territorial units for urban  
24 governance and administration.  
25  
26  
27

28  
29 The survey was conducted by the Nanjing Transport Bureau every three to five years to establish snapshots  
30 of residents' travel patterns in one day. The home interview survey using structured questionnaires  
31 collected travel information, including trip origin and destination, start and end times, purpose, mode  
32 choice, as well as socio-demographics of households and individuals. Travel diaries during a designated  
33 24-hour period on a typical day – i.e. October 30<sup>th</sup> 2013, Wednesday – were obtained. Individuals were  
34 randomly selected from each of the 49 sub-districts according to their population; a sub-district with a  
35 higher population was targeted with a larger sample size. To seek the representativeness of the sample,  
36 people residing in each stratum were contacted via computer-aided random-digit-dialling phone calls.  
37 These recruitment calls were made by the staff members of sub-district offices on October 28<sup>th</sup> and 29<sup>th</sup>,  
38 2013. 5,562 residents were contacted, of which 5,172 provided complete answers. Since this study focuses  
39 on older people, only respondents who are aged 60 years or above are included for further analysis. A  
40 resulting sample size of 702 respondents consists of 376 men and 326 women.  
41  
42  
43  
44

45  
46 The profile of the sample is summarised in Table 1. A high proportion (48%) of the respondents was in the  
47 medium-income households (a family annual income of RMB 50,000 – 150,000), while only 22% belonged  
48 to the high-income households (RMB 150,000 or more). The majority (64%) had no access to a car, while  
49 71% indicated to own bicycles in their households. Only around one-eighth of the respondents completed  
50 higher education (bachelor degree or above). Holding a driving license was not common (less than one in  
51 ten of those interviewed). More than 80% of the respondents had a public transport pass to enjoy travel  
52 concessions (i.e. at a reduced price). In regard to their travel behaviour, we calculated the daily walking  
53 time. The mean and standard deviation were 40.2 minutes and 49.7.  
54  
55  
56

---

57  
58 <sup>1</sup> The built environment simultaneously affects different facets of walking behaviour, including walking frequency,  
59 distance, and time. Given that walking time is a more useful indicator representing the intensity of physical activity  
60 for older adults (WHO, 2010), our study emphasises on the built environment effects on walking time.  
61  
62  
63  
64  
65

Table 1. Profile of the respondents and the built environment data

	Variable	Description	Mean	SD
Travel behaviour	Walking time	Total amount of time for walking trips (min)	40.2	49.7
Socio-demographics	Household income	High annual household income (> 150,000 RMB) (0: no; 1: yes)	0.22	0.47
		Medium annual household income (50,000–150,000 RMB) (0: no; 1: yes)	0.48	0.49
		Low annual household income (< 50,000 RMB) (0: no; 1: yes)	0.30	0.46
	Car ownership	Car ownership in the household (0: no; 1: yes)	0.36	0.48
	Bicycle ownership	Bicycle ownership in the household (0: no; 1: yes)	0.71	0.45
	Gender	Male = 1, female = 0	0.54	0.50
	Education level	Highly educated respondent (0: no; 1: yes)	0.13	0.34
		Medium educated respondent (0: no; 1: yes)	0.41	0.49
		Low educated respondent (0: no; 1: yes)	0.46	0.50
	Driving license holding	Whether the respondent holds a driving license (0: no; 1: yes)	0.09	0.28
Public transport pass availability	Whether the respondent has public transport pass to enjoy travel concession (0: no; 1: yes)	0.86	0.35	
Built environment	Population density	Total number of residential population/Total built-up area in a buffer zone (persons/1000m <sup>2</sup> )	10.75	12.37
	Land use mix	Land use diversity index measured by equation (1)	0.61	0.14
	Street connectivity	Total length of sidewalks/Total built-up area in a buffer zone (km/km <sup>2</sup> )	3.72	4.36
	Number of bus stops	Number of bus stops in a buffer zone	5.02	2.84
	Number of bike-sharing stations	Number of bike-sharing stations in a buffer zone	4.03	2.89
	Distance to the nearest square/park	Distance to the nearest square/park in a neighbourhood vicinity (km)	1.50	0.83
	Distance to the nearest card/chess room	Distance to the nearest card/chess room in a neighbourhood vicinity (km)	0.61	0.48

Note: SD = standard deviation.

Data for measuring the built environment was sourced from the Nanjing Urban GIS Database. Exposure to the neighbourhood environment is assessed within a 500-metre buffer zone (Cerin et al., 2017) around the residence of respondents.<sup>2</sup> Based on a review of relevant studies (e.g. Cheng et al., 2019b; Gómez et al., 2010; Leung et al., 2018; Liu et al., 2020; Perchoux et al., 2019), seven variables which can be considered to have important impacts on the time older adults spent walking are included in the analysis. They are: (1) population density; (2) land use mix, this measure considers five types of land use: residential, commercial, educational, entertainment and public services. It is calculated as:

$$H = -(\sum_{i=1}^n p_i * \ln(p_i))/\ln(n) \quad (1)$$

where  $H$  represents the land use mix score (ranging from 0 (complete dominance of one type of land use) to 1 (all types of land use are equally distributed)),  $p_i$  is the proportion of the  $i^{\text{th}}$  land use, and  $n$  is the number of land use types; (3) street connectivity – a measure of walkability – is computed as the density

<sup>2</sup> Neighbourhood environmental attributes within 500-m from the place of residence appear to be strong predictors of physical activity (including active travel) for older adults (Cerin et al., 2017).

of sidewalks (Hou, 2019); (4) number of bus stops; (5) number of bike-sharing stations; (6) distance to the nearest square/park; and (7) distance to the nearest card/chess room. Distances are calculated based on the shortest route along the street network. Table 1 presents the descriptive statistics of built environment characteristics.

#### 4. Methodology

This study applies a random forest method to disentangle the complex relationships between the built environment and walking among older adults. Rather than a priori assuming a particular relationship which is fitted to the data, the random forest method – a data mining approach – explores more refined associations between the outcome and explanatory variables. It is a tree-based ensemble approach, combining multiple individual decision trees to optimise model fit and prediction (Breiman, 2001). The objective is to minimise a loss function and approximate the true value by recursively assessing prediction errors until the loss function reaches a minimum or remains stable. A decision tree is the basic algorithm for the random forest method, whose work process is illustrated in Figure 1. Two treatments of randomness are incorporated to ensure the diversity of decision trees. First, bootstrapping the training data so that each tree grows with a different sub-sample. Second, selecting feature randomly to result in different subsets of explanatory variables for splitting nodes in a tree. In an individual decision tree, splitting continues until a maximum tree depth is reached. A single prediction is made based on the mean response of each region. The final result is obtained by averaging the predictions from all individual trees.

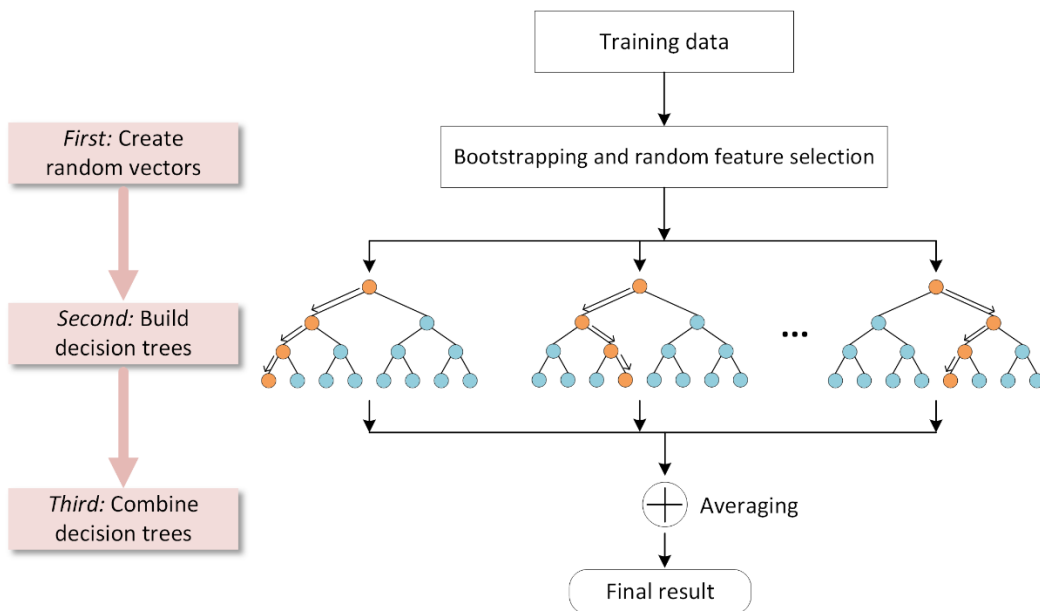


Figure 1. An illustration of the random forest method (Source: Cheng et al. (2019a))

Three aspects have important influences on the predictive performance of the random forest method (Cheng et al., 2019a): (1) the performance of an individual tree needs to be improved; (2) the correlation between trees needs to be reduced; and (3) the number of trees needs to be large. Accordingly, we need

to tune three parameters for model specification: maximum depth of the tree, number of splitting variables, and the number of trees. The random forest is able to quantify the relative importance of explanatory variables in predicting the outcome, which increases model interpretability. The permutation importance measure, introduced by Breiman (2001), is commonly used to compute variable importance. The importance for variable  $x_i$  is formulated as:

$$VI_{x_i} = \frac{1}{N} \sum_t (OOB_{MSE}^t - OOB_{MSE,perm_i}^t) \quad (2)$$

where  $VI_{x_i}$  is the variable importance;  $N$  is the total number of decision trees in the forest;  $OOB_{MSE}^t$  denotes the mean square error before permutation for tree  $t$ ; and  $OOB_{MSE,perm_i}^t$  represents the mean square error after permutation of variable  $x_i$ .<sup>3</sup>

More importantly, the partial dependence plot produced by the random forest visualises the relationships between the outcome and explanatory variables (Hastie et al., 2009). These relationships are not refrained by any pre-specified (e.g. linear or log-linear) assumptions, and illustrates the data-based estimate of the built environment on walking. In this research, it depicts the marginal effect of a built environment attribute on walking time while controlling for the average effects of all other explanatory variables in the given model. Mathematically, the partial dependence of  $\hat{f}(x)$  on  $x_s$  is formulated as follows:

$$\hat{f}_s(x_s) = \frac{1}{N} \sum_{i=1}^N \hat{f}(x_s, x_{ic}) \quad (3)$$

where  $\{x_{1c}, x_{2c}, \dots, x_{Nc}\}$  are the values of other variables  $x_c$  in the dataset,  $N$  is the number of instances. Partial dependence works by marginalising the model output over the distribution of other variables. On the basis of the plot, we can see how the effects of the built environment on walking time (in)stabilise within certain ranges, i.e. the non-linear relationships.

## 5. Results

Before building the random forest, we performed a variance inflation factor (VIF) analysis to detect possible multicollinearity between the independent variables. All the socio-demographics and built environment attributes in Table 1 have a VIF value of less than five – a threshold for excluding variables (Craney and Surles, 2002) – and thus are retained in the final model. In order to obtain the optimal specification of random forest, we first developed the model with the number of trees ranging from 10 to 1,000 in increments of 10. It is found that building trees beyond 500 produces a negligible decrease in root mean square error while consumes much more time. Therefore, the number of trees is set as 500. Then, we experimented with different combinations of the number of splitting variables (3, 5, 7, 9)<sup>4</sup> and maximum depth of the tree (from 10 to 2000 with an interval of 20). The correlation of individual decision trees may be strengthened if we choose a higher number of splitting variables. Results indicate that 5 splitting variables lead to the best prediction model when the maximum tree depth is 1,200. The root

<sup>3</sup> In the random forest, bootstrapping the data – sampling with replacement – for training an individual tree leads to some observations left out, known as out-of-bag (OOB) sample (Breiman, 2001). Each tree has its own OOB sample which is not used during training. This sample is used to quantify the prediction performance of the model.

<sup>4</sup> Breiman (2001) suggested that the number of splitting variables should be  $p/3$ , which is 5 ( $p = 14$  is the total number of explanatory variables, Table 1).



mean square error (RMSE) is 32.20 and pseudo-R<sup>2</sup> is 0.313. This final model is used for further analysis, quantifying variable importance and creating partial dependence plot.

### 5.1. Relative importance of explanatory variables

The relative importance of explanatory variables in predicting walking time among older adults is presented in Table 2. The relative importance represents the relative improvement in the reduction of prediction error. Among the considered explanatory variables in Table 1, the collective contribution of the built environment is higher than that of socio-demographics. This result reinforces earlier research that land use variables exert more influences on travel than demographics (Ewing and Cervero, 2001; Gim, 2013) and indicates the crucial role of land use interventions encouraging elderly’s walking. In addition, the relative importance of built environment attributes informs useful planning practice because not all interventions can be implemented for limited resources and there is usually a priority ranking among these attributes. Population density is the most important attribute, with a contribution of 11.5%. This is followed by distance to the nearest square/park and street connectivity, with non-trivial effects of more than 10%. Although land use mix, number of bus stops, and number of bike-sharing stations have less predictive power, they still rank higher compared to socio-demographic variables. Among socio-demographics, household income is the most important variable in predicting walking time, accounting for 6.9% of the total effect. By contrast, gender contributes the least, with only 3.4% in the model.

Table 2. Relative importance of explanatory variables

	Variable	Relative importance (%)	Total (%)
Built environment	Population density	11.5	68.0
	Distance to the nearest square/park	11.3	
	Street connectivity	11.2	
	Land use mix	9.9	
	Number of bus stops	9.1	
	Number of bike-sharing stations	8.8	
	Distance to the nearest card/chess room	6.2	
Socio-demographics	Household income	6.9	32.0
	Public transport pass availability	6.6	
	Car ownership	4.2	
	Education level	3.8	
	Driving license holding	3.6	
	Bicycle ownership	3.5	
	Gender	3.4	
Total relative importance			100

## 5.2. Non-linear effects of the built environment on walking time

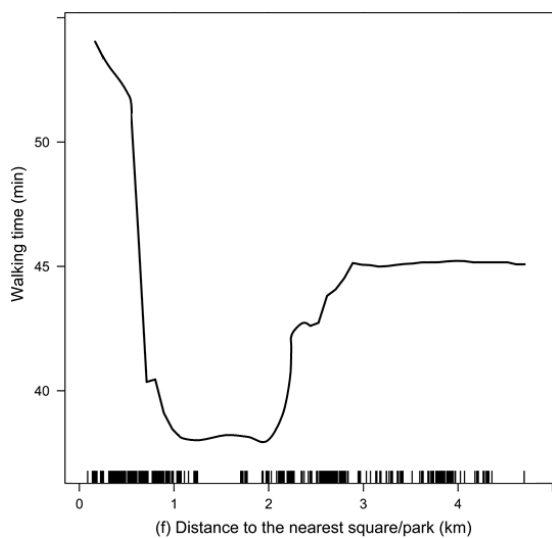
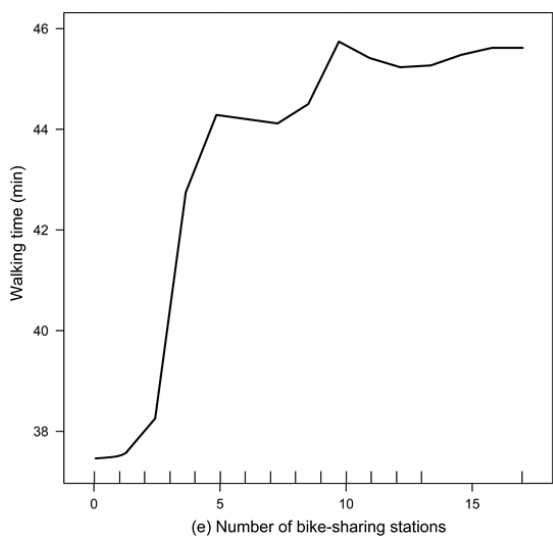
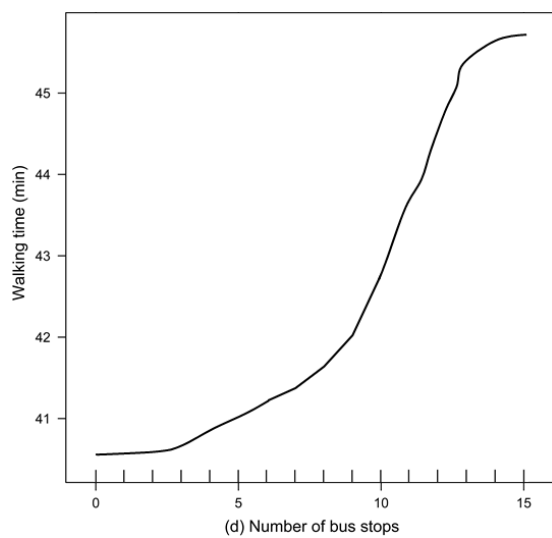
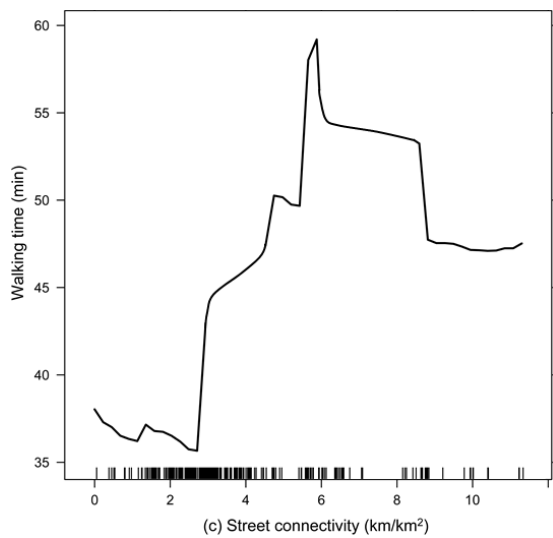
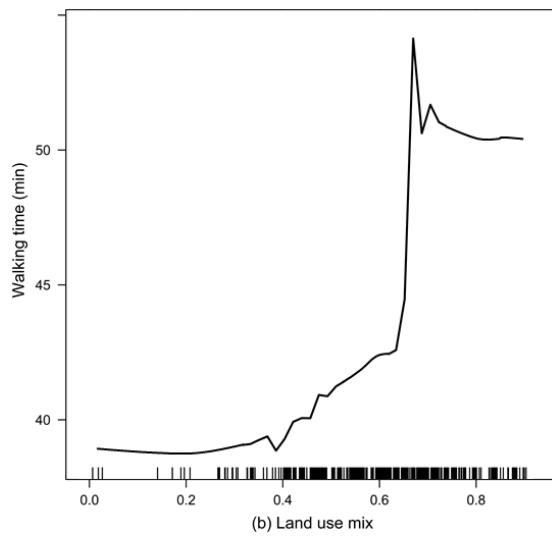
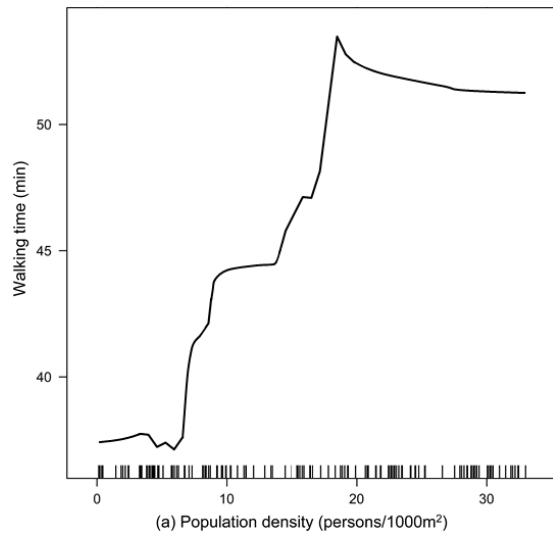
The partial dependence plot provides a fine-grained analysis of the complicated relationships between the built environment and walking time among older adults (Figure 2). The x-axis of each plot indicates the data distribution. It is noted that partial dependence estimates are less reliable in regions where there is not much training data. Probably, the random forest method could not learn a meaningful prediction within these regions; hence estimates are vulnerable to noise.

Figure 2a presents the non-linear effects of population density on walking, controlling for all other explanatory variables. Population density ranging between 6 and 20 persons/1000m<sup>2</sup> is observed to positively and monotonically influence walking. Higher population density is likely to create a critical mass of pedestrians – people walk more, more road space and infrastructure will be allocated to pedestrians, and walking will become more established, safe and pleasant. Traffic congestion in densely populated areas also makes walking a more convenient way to travel compared to motorised transport (Oakes et al., 2007). In addition, higher population density tends to support the presence of various services and shops, which has an indirect influence on walking (Transportation Research Board, 2005). When the population density exceeds 20 persons/1000m<sup>2</sup>, it has a slightly negative impact. This might relate to the higher risk of injuries in crowded places, in particular for older people, some of whom may have reduced physical capabilities.

For land use mix, walking time remains stable within the interval of 0 and 0.4 (Figure 2b). But when it changes from 0.4 to 0.7, walking time increases rapidly and reaches the highest level around 0.7. Diversity of land use (residential, commercial, educational, entertainment and public services) provides opportunities for local activities, and thus lowers the need for motorised travel to access (distant) activities (Kamruzzaman et al., 2016). A neighbourhood with a good mixture of land use is often related with a wide range of non-residential destinations for walking trips, which is also conducive to walking (Duncan et al., 2010). However, when land use mix exceeds the score of 0.7, walking time decreases. One possible reason for this result may be that older people chain multiple trips in a single journey due to easy access to destinations and services in highly mixed areas.

The partial dependence curve in Figure 2c indicates a positive correlation between street connectivity and walking time among those with a lower level of street connectivity (below six km/km<sup>2</sup>). Better street connectivity facilitates walking through the provision of more (direct) routes within the neighbourhood and shorter distances to destinations. It makes walking more efficient and increases the number of alternative walking routes with implications for security and interest (Oakes et al., 2007). A negative association is observed with a higher level of street connectivity (above six km/km<sup>2</sup>). This effect, however, needs to be interpreted with caution due to the absence of enough data. These mixed results are also found in other studies on walking among older people (e.g. Barnett et al., 2017; Satariano et al., 2010; Wang and Lee, 2010).

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



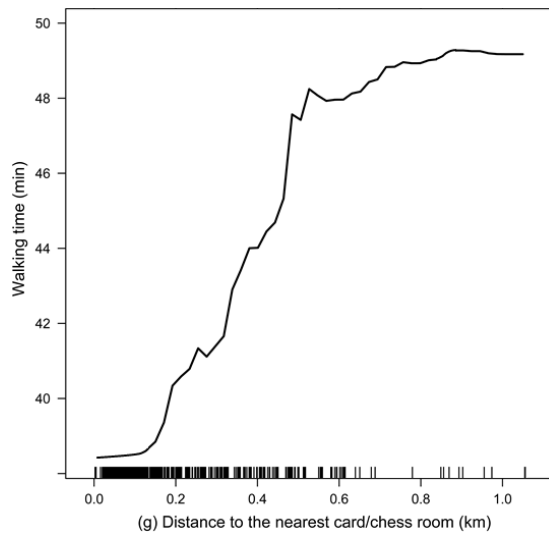


Figure 2. Non-linear effects of the built environment on walking time

Walking time among older adults increases with access to public transport – number of bus stops and number of bike-sharing stations – as shown in Figures 2d and 2e. These results align with existing literature (Barnes et al., 2016; Rodriguez et al., 2008; Suminski et al., 2005). As indicated by Freeland et al. (2013), travelling by public transport is always coupled with walking to and from bus stops/bike stations (since walking forms a portion of most public transport trips). It could also possibly be attributed to the transit-oriented developments (TODs) in the neighbourhood. They are manifested in two ways: fewer opportunities for car use foster active mobility (Perchoux et al., 2019), and the provision of various services offers attractive walking destinations (Hou, 2019). Furthermore, neighbourhoods with better public transport accessibility may also facilitate walking among non-users of public transport. For instance, a large share of people walking to public transport can create a walk-friendly atmosphere and the idea that walking can be considered as an important way of travelling. This in turn promotes walking through the mechanism of perceived safety and social norms (Forsyth et al., 2007). It is worth highlighting that the number of bus stops has threshold effects on the amount of walking. Only when the number is above three, a positive relationship with walking time becomes clear. The threshold effect is also observed for the number of bike-sharing stations. Walking time increases substantially as the number of bike-sharing stations grows from zero to ten; then the curve becomes almost horizontal (with ten as the cut-off point).

We observe different overall impacts of the two analysed recreational facilities: Walking time is negatively correlated with distance to a square/park (if the distance is shorter than one kilometre) while positively correlated with distance to a card/chess room. For the distance to the nearest square/park, the distribution of the data in Figure 2f shows that we could only make reliable interpretations in the ranges of 0 to 1 km and 2 to 4.3 km. It is negatively correlated with walking time when the distance is less than one kilometre. Living farther from urban parks, plazas, and green spaces is unsupportive of walking among older people, which concurs with earlier empirical findings (Cheng et al., 2019b; Julien et al., 2015). But when the distance changes from two to three kilometres, walking time begins to increase. This probably occurs because older people living in this area (whose homes are two to three kilometres far away) specially walk to squares/parks for exercise/leisure. If the distance becomes too large and goes beyond

the daily activity space of older people (i.e. three kilometres in this study), its effect will be trivial. To put it another way, the provision of a square/park located more than three kilometres far away has no important effects on walking. In terms of distance to card/chess rooms (Figure 2g), it seems to be positively associated with the amount of walking if the distance ranges between 0.1 and 0.6 km. The potential rationale is that senior people tend to be more attracted by card/chess rooms if these facilities are available in the nearby vicinity and spend more time engaging such activities, which would reduce the time allocated to walking in a day. It is worth noting that it is difficult to accurately interpret the effects in the ranges of above 0.6 km because of the sparse distribution of the data.

### 5.3. Model comparison with linear regression

We estimated a conventional linear regression model with transformed independent variables. A natural log transformation for distance-related independent variables is performed to capture the diminishing effect of these variables with distance. Previous studies have shown that the log transformation of the distance variables fits the model better than a simple linear functional form (Bin and Polasky, 2004; Morency et al., 2011). We use quadratic specifications for other built environment variables such as population density, land use mix, and street connectivity. We experimented with several alternative specifications and find that the primary results are robust to alternative functional forms. A commonly-used 10-fold cross-validation was applied to compare and evaluate the predictive capability (Zhao et al., 2020). The results in terms of mean absolute error (MAE) and root mean square error (RMSE) for out-of-sample are shown in Table 3. The two metrics are formulated as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (5)$$

where  $N$  is the total number of respondents in the validation set,  $\hat{y}_i$  is the predicted walking time for the  $i^{th}$  respondent,  $y_i$  is the observed walking time for the  $i^{th}$  respondent. The partial dependence plots in the random forest and parameter estimates in the linear regression both demonstrate the non-linear relationship between the built environment and walking time (Table 3). However, the  $R^2$  of the linear regression is 0.144, much smaller than the pseudo- $R^2$  of the random forest method (0.313). It is also evident that the random forest model has much lower prediction error than the linear regression. The values of out-of-sample MAE and RMSE for the random forest are smaller. This is not surprising because the random forest is able to capture more complex relationships by letting the data speak more freely.

We also compared the relative importance of independent variables to walking time. In the linear regression, adding one independent variable can improve the model goodness-of-fit. The proportion of this improvement in the overall goodness-of-fit ( $\%R^2$ ) is counted as the relative importance of the independent variable. Table 3 shows that, in general, the relative importance of independent variables is different between the random forest and the linear regression. The Spearman's rank correlation of the relative importance (random forest) and  $\%R^2$  (linear regression, excluding the non-linear terms) is 0.310. Because the random forest is more flexible in terms of revealing non-linear relationships, the pre-specified relationship between independent variables and walking time in the linear regression is flawed.

Table 3. Comparison between random forest and linear regression

Variable	Random forest		Linear regression		
	Relative importance (%)	Rank	Coefficient	T-statistic	%R <sup>2</sup>
<i>Built environment</i>					
Population density	11.5	1	56.04	2.15	8.9
(Population density) <sup>2</sup>			-1.48	-1.87	6.4
Distance to square/park	11.3	2			
Ln (Distance to square/park)			-17.01	-2.89	4.5
Street connectivity	11.2	3	27.13	2.32	6.2
(Street connectivity) <sup>2</sup>			-1.94	-2.86	3.0
Land use mix	9.9	4	44.84	2.38	7.1
(Land use mix) <sup>2</sup>			-42.77	-2.89	6.3
Number of bus stops	9.1	5	4.06	2.22	3.4
(Number of bus stops) <sup>2</sup>			-0.37	-1.91	7.0
Number of bike-sharing stations	8.8	6	3.31	1.22	1.7
(Number of bike-sharing stations) <sup>2</sup>			-0.23	-1.21	0.8
Distance to card/chess room	6.2	9			
Ln (Distance to card/chess room)			8.84	2.11	0.7
<i>Socio-demographics</i>					
Household income	6.9	7	-4.06	-1.97	9.5
Public transport pass availability	6.6	8	10.76	2.14	8.8
Car ownership	4.2	10	-5.25	-2.38	5.8
Education level	3.8	11	2.55	1.23	6.3
Driving license holding	3.6	12	-9.31	-2.04	4.4
Bicycle ownership	3.5	13	3.31	1.53	6.1
Gender	3.4	14	2.79	1.28	3.1
R <sup>2</sup>	0.313		0.144		
Out-of-sample predictive capability (10-fold cross-validation)					
	Mean	SD	Mean	SD	
MAE	21.23	0.91	27.45	1.18	
RMSE	32.15	1.93	39.64	2.21	

## 6. Discussion and conclusions

In order to build an age-friendly society – where elderly people can easily perform preferred out-of-home activities and remain physically active – understanding how the built environment can support active travel of older adults is of particular interest in urban and transport planning. In this study, we have shed light on the non-linear associations between the built environment and walking time among older adults. The non-linear patterns are demonstrated to be prevalent and vary among built environment attributes. The random forest method is employed as no predefined functional form between the outcome and explanatory variables is required. The uncovered relationship is a refined expression deriving from the data itself. A comparison with the conventional linear regression shows that the random forest is more effective in revealing the complicated relationships. To the best of our knowledge, this is the first study adopting a random forest method in the travel behaviour-built environment field, providing an important contribution to existing literature.

The salient non-linear effects of the built environment on walking can assist the design of more effective policies promoting walking among older adults. In general, efforts need to be directed towards nuanced environmental interventions to ensure that the facilitating effects of the built environment are realised.

1  
2  
3  
4 Specifically, the effective ranges of population density and land use mix stimulating walking are 6-20  
5 persons/1000m<sup>2</sup> and 0.4-0.7 score, respectively. The ultra-densely populated areas and excessively mixed  
6 land uses may lead to negative impacts on the amount of elderly's walking. Strategies for designing street  
7 networks should also account for the non-linear relationships found in this research, i.e. the optimal  
8 setting found is around six km road per km<sup>2</sup>. These recommended values are important with respect to  
9 improving the cost-effectiveness of land use interventions. Access to public transport fosters walking and  
10 TODs can therefore have important implications for achieving regular engagement in walking among older  
11 adults. More importantly, lower bound and upper bound threshold effects are observed, respectively, for  
12 the number of bus stops and the number of bike-sharing stations. Roughly speaking, the number of bus  
13 stops would not have pronounced effects on walking until it exceeds three in the neighbourhood.  
14 Pertaining to the number of bike-sharing stations, when it is above ten in the neighbourhood, the  
15 association becomes negligible. The proximity to a square/park is positively related to walking, but it  
16 should be borne in mind that the effective distance is within one kilometre. On the other hand, the  
17 proximity to card/chess rooms may lead to reduced walking time and a more sedentary lifestyle, although  
18 it may offer leisure and social interactions. Thus, policies to increase the provision of recreational facilities  
19 have to be implemented with caution.  
20  
21  
22  
23  
24  
25

26 The built environment collectively accounts for the vast majority of power describing walking time, relative  
27 to socio-demographics. The substantial built environment effects indicate that land use interventions can  
28 effectively influence elderly's walking. The relative importance of each built environment attribute further  
29 implies the hierarchy of intervention priorities. Population density plays the most important role in walking  
30 time among older people. Therefore, urban densification should be the top consideration. Proximity to a  
31 square/park and street connectivity are also strongly and positively associated with elderly's walking. This  
32 informs policy initiatives in terms of improving the provision of public venues for walking (e.g. park, square,  
33 or plaza) and the density of walkways when designing pedestrian-friendly neighbourhoods.  
34  
35  
36

37 This study has some limitations resulting in several avenues for future research. First, whether the non-  
38 linear effects of the built environment on seniors' walking are comparable to other places is unclear due  
39 to the focus on a single city in a Chinese context (i.e. Nanjing). Since this research is the first to  
40 quantitatively look at non-linear relationships between the built environment and walking, the results are  
41 difficult to generalise. Future studies in other regions, but using a similar method, should be conducted  
42 before any generalisation can be made. Second, as a data-driven approach, the random forest produces  
43 fluctuating results for some ranges of explanatory variables. This could be related to outliers, noise, sparse  
44 data, and/or the spatially uneven distribution of the explanatory variables. Increasing the amount and  
45 density of data would be a way out. In addition, the random forest method used in this study depends on  
46 orthogonal decision boundaries for regression, which may fail to find optimal partitions. Therefore, it  
47 would be interesting to apply non-orthogonal and non-linear decision boundaries in decision trees (e.g.  
48 via interactive basis functions), which could improve the performance of the technique (Páez et al., 2019).  
49 Third, an extension of this work might be the mapping of the potential for compliance with physical activity  
50 guidelines (i.e. 150 minutes per week) in relation to walking for transport. Compliance potential mapping  
51 can offer valuable information for transport and public health planning (Moniruzzaman et al., 2014).  
52 Nonetheless, the non-linear and threshold effects revealed in this study yield important insights for  
53 deliberating land use and transport policies that seek to encourage walking among older adults.  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 **Declarations of interest**  
5

6 None.  
7  
8  
9

10 **Acknowledgements**  
11

12 The authors would like to gratefully acknowledge the support by the National Key Research and  
13 Development Program of China (No. 2018YFB1600900), the National Natural Science Foundation of China  
14 (No. 71801041 and No. 71771049), the Research Foundation – Flanders (FWO, grant 12F2519N), and the  
15 Estonian Research Council (PUT PRG306). Thanks to extensive comments from the anonymous reviewers,  
16 they have significantly improved the paper.  
17  
18  
19  
20

21 **References**  
22

23 Barnes, R., Winters, M., Ste-Marie, N., McKay, H., Ashe, M.C., 2016. Age and retirement status differences  
24 in associations between the built environment and active travel behaviour. *Journal of Transport & Health*,  
25 3(4), 513-522.  
26

27 Barnett, D.W., Barnett, A., Nathan, A., van Cauwenberg, J., Cerin, E., 2017. Built environmental correlates  
28 of older adults' total physical activity and walking: a systematic review and meta-analysis. *International*  
29 *Journal of Behavioral Nutrition and Physical Activity*, 14, 103.  
30

31 Böcker, L., van Amen, P., Helbich, M., 2017. Elderly travel frequencies and transport mode choices in  
32 Greater Rotterdam, the Netherlands. *Transportation*, 44(4), 831-852.  
33

34 Bin, O., Polasky, S., 2004. Effects of flood hazards on property values: Evidence before and after Hurricane  
35 Floyd. *Land Economics*, 80(4), 490-500.  
36

37 Breiman, L., 2001. Random forests. *Machine Learning*, 45, 5-32.  
38

39 Cao, X., Mokhtarian, P.L., Handy, S.L., 2010. Neighborhood design and the accessibility of the elderly: An  
40 empirical analysis in Northern California. *International Journal of Sustainable Transportation*, 4(6), 347-  
41 371.  
42

43 Cerin, E., Nathan, A., van Cauwenberg, J., Barnett, D.W., Barnett, A., 2017. The neighbourhood physical  
44 environment and active travel in older adults: a systematic review and meta-analysis. *International Journal*  
45 *of Behavioral Nutrition and Physical Activity*, 14, 15.  
46

47 Cerin, E., Sit, C.H., Barnett, A., Johnston, J.M., Cheung, M.C., Chan, W.M., 2014. Ageing in an ultra-dense  
48 metropolis: perceived neighbourhood characteristics and utilitarian walking in Hong Kong elders. *Public*  
49 *Health Nutrition*, 17(1), 225-232.  
50

51 Cheng, L., Chen, X., De Vos, J., Lai, X., Witlox, F., 2019a. Applying a random forest method approach to  
52 model travel mode choice behavior. *Travel Behaviour and Society*, 14, 1-10.  
53

54 Cheng, L., Chen, X., Yang, S., Cao, Z., De Vos, J., Witlox, F., 2019b. Active travel for active ageing in China:  
55 The role of built environment. *Journal of Transport Geography*, 76, 142-152.  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



- 1  
2  
3  
4 Cheng, L., De Vos, J., Shi, K., Yang, M., Chen, X., Witlox, F., 2019c. Do residential location effects on travel  
5 behavior differ between the elderly and younger adults? *Transportation Research Part D*, 73, 367-380.  
6  
7  
8 Curl, A., Mason, P., 2019. Neighbourhood perceptions and older adults' wellbeing: Does walking explain  
9 the relationship in deprived urban communities? *Transportation Research Part A*, 123, 119-129.  
10  
11 Craney, T.A., Surlles, J.G., 2002. Model-dependent variance inflation factor cutoff values. *Quality*  
12 *Engineering*, 14(3), 391-403.  
13  
14 Ding, C., Cao, X., Liu, C., 2019. How does the station-area built environment influence Metrorail ridership?  
15 Using gradient boosting decision trees to identify non-linear thresholds. *Journal of Transport Geography*,  
16 77, 70-78.  
17  
18 Ding, C., Cao, X.J., Næss, P., 2018. Applying gradient boosting decision trees to examine non-linear effects  
19 of the built environment on driving distance in Oslo. *Transportation Research Part A*, 110, 107-117.  
20  
21 Duncan, M.J., Winkler, E., Sugiyama, T., Cerin, E., Leslie, E., Owen, N., 2010. Relationships of land use mix  
22 with walking for transport: do land uses and geographical scale matter? *Journal of Urban Health*, 87(5),  
23 782-795.  
24  
25  
26 Ewing, R., Cervero, R., 2001. Travel and the built environment: a synthesis. *Transportation Research Record*,  
27 1780, 87-114.  
28  
29 Forsyth, A., Oakes, J.M., Schmitz, K.H., Hearst, M., 2007. Does residential density increase walking and  
30 other physical activity? *Urban Studies*, 44(4), 679-697.  
31  
32  
33 Frank, L., Kerr, J., Rosenberg, D., King, A., 2010. Healthy aging and where you live: community design  
34 relationships with physical activity and body weight in older Americans. *Journal of Physical Activity and*  
35 *Health*, 7(s1), S82-S90.  
36  
37 Freeland, A.L., Banerjee, S.N., Dannenberg, A.L., Wendel, A.M., 2013. Walking associated with public  
38 transit: moving toward increased physical activity in the United States. *American Journal of Public Health*,  
39 103(3), 536-542.  
40  
41 Galster, G.C., 2018. Nonlinear and threshold effects related to neighborhood: Implications for planning  
42 and policy. *Journal of Planning Literature*, 33(4), 492-508.  
43  
44 Gan, Z., Yang, M., Feng, T., Timmermans, H.J.P., 2020. Examining the relationship between built  
45 environment and metro ridership at station-to-station level. *Transportation Research Part D*, 82, 102332.  
46  
47 Ghani, F., Rachele, J.N., Loh, V.H., Washington, S., Turrell, G., 2018. Do differences in built environments  
48 explain age differences in transport walking across neighbourhoods? *Journal of Transport & Health*, 9, 83-  
49 95.  
50  
51  
52 Gim, T.H.T., 2013. The relationships between land use measures and travel behavior: A meta-analytic  
53 approach. *Transportation Planning and Technology*, 36(5), 413-434.  
54  
55 Gómez, L.F., Parra, D.C., Buchner, D., et al., 2010. Built environment attributes and walking patterns among  
56 the elderly population in Bogotá. *American Journal of Preventive Medicine*, 38(6), 592-599.  
57  
58  
59  
60  
61  
62  
63  
64  
65

- 1  
2  
3  
4 Haselwandter, E.M., Corcoran, M.P., Folta, S.C., Hyatt, R., Fenton, M., Nelson, M.E., 2015. The built  
5 environment, physical activity, and aging in the United States: A state of the science review. *Journal of*  
6 *Aging and Physical Activity*, 23(2), 323-329.  
7  
8  
9 Hastie, T., Tibshirani, R., Friedman, J., 2009. *The elements of statistical learning: data mining, inference,*  
10 *and prediction, second edition.* Springer.  
11  
12 Holden, E., Norland, I.T., 2005. Three challenges for the compact city as a sustainable urban form:  
13 household consumption of energy and transport in eight residential areas in the greater Oslo region. *Urban*  
14 *Studies*, 42(12), 2145-2166.  
15  
16 Hou, Y., 2019. Polycentric urban form and non-work travel in Singapore: A focus on seniors. *Transportation*  
17 *Research Part D*, 73, 245-275.  
18  
19  
20 Humphreys, D.K., Goodman, A., Ogilvie, D., 2013. Associations between active commuting and physical  
21 and mental wellbeing. *Preventive Medicine*, 57(2), 135-139.  
22  
23 Julien, D., Richard, L., Gauvin, L., et al., 2015. Transit use and walking as potential mediators of the  
24 association between accessibility to services and amenities and social participation among urban-dwelling  
25 older adults: insights from the VoisiNuAge study. *Journal of Transport & Health*, 2(1), 35-43.  
26  
27 Kamruzzaman, M., Washington, S., Baker, D., Brown, W., Giles-Corti, B., Turrell, G., 2016. Built  
28 environment impacts on walking for transport in Brisbane, Australia. *Transportation*, 43(1), 53-77.  
29  
30  
31 Kemperman, A., Timmermans, H., 2009. Influences of built environment on walking and cycling by latent  
32 segments of aging population. *Transportation Research Record*, 2134, 1-9.  
33  
34 Lee, E., Dean, J., 2018. Perceptions of walkability and determinants of walking behaviour among urban  
35 seniors in Toronto, Canada. *Journal of Transport & Health*, 9, 309-320.  
36  
37 Leung, K.M., Chung, P.K., Wang, D., Liu, J.D., 2018. Impact of physical and social environments on the  
38 walking behaviour of Hong Kong's older adults. *Journal of Transport & Health*, 9, 299-308.  
39  
40  
41 Lewis, S., 2018. Neighborhood density and travel mode: new survey findings for high densities.  
42 *International Journal of Sustainable Development & World Ecology*, 25(2), 152-165.  
43  
44 Li, F., Fisher, K.J., Brownson, R.C., Bosworth, M., 2005. Multilevel modelling of built environment  
45 characteristics related to neighbourhood walking activity in older adults. *Journal of Epidemiology &*  
46 *Community Health*, 59(7), 558-564.  
47  
48 Liao, Y., Shibata, A., Ishii, K., Koohsari, M.J., Inoue, S., Oka, K., 2018. Can neighborhood design support  
49 walking? Cross-sectional and prospective findings from Japan. *Journal of Transport & Health*, 11, 73-79.  
50  
51  
52 Liu, Z., Kemperman, A., Timmermans, H., 2020. Correlates of older adults' walking trip duration. *Journal of*  
53 *Transport & Health*, 18, 100889.  
54  
55 Mercado, R., Páez, A., 2009. Determinants of distance traveled with a focus on the elderly: a multilevel  
56 analysis in the Hamilton CMA, Canada. *Journal of Transport Geography*, 17(1), 65-76.  
57  
58 Mokhtarian, P.L., Salomon, I., 2001. How derived is the demand for travel? Some conceptual and  
59 measurement considerations. *Transportation Research Part A*, 35(8), 695-719.  
60  
61  
62  
63  
64  
65

- 1  
2  
3  
4 Moniruzzaman, M., Páez, A., 2016. An investigation of the attributes of walkable environments from the  
5 perspective of seniors in Montreal. *Journal of Transport Geography*, 51, 85-96.  
6  
7 Moniruzzaman, M., Páez, A., Habib, K.M.N., Morency, C., 2013. Mode use and trip length of seniors in  
8 Montreal. *Journal of Transport Geography*, 30, 89-99.  
9  
10 Moniruzzaman, M., Páez, A., Morency, C., 2014. Compliance potential mapping: a tool to assess potential  
11 contributions of walking towards physical activity guidelines. *BMC Public Health*, 14, 511.  
12  
13 Morency, C., Páez, A., Roorda, M.J., Mercado, R., Farber, S., 2011. Distance traveled in three Canadian  
14 cities: Spatial analysis from the perspective of vulnerable population segments. *Journal of Transport*  
15 *Geography*, 19(1), 39-50.  
16  
17 Næss, P., 2006. Are short daily trips compensated by higher leisure mobility? *Environment and Planning B*,  
18 33(2), 197-220.  
19  
20 Nagel, C.L., Carlson, N.E., Bosworth, M., Michael, Y.L., 2008. The relation between neighborhood built  
21 environment and walking activity among older adults. *American Journal of Epidemiology*, 168(4), 461-468.  
22  
23 Nanjing Municipal Bureau Statistics, 2018. *Statistical Yearbook of Nanjing 2018*. China Statistics Press,  
24 Beijing, China.  
25  
26 Nordbakke, S., Schwanen, T., 2014. Well-being and mobility: A theoretical framework and literature review  
27 focusing on older people. *Mobilities*, 9(1), 104-129.  
28  
29 Oakes, J.M., Forsyth, A., Schmitz, K H., 2007. The effects of neighborhood density and street connectivity  
30 on walking behavior: the Twin Cities walking study. *Epidemiologic Perspectives & Innovations*, 4, 16.  
31  
32 Páez, A., López, F., Ruiz, M., Camacho, M., 2019. Inducing non-orthogonal and non-linear decision  
33 boundaries in decision trees via interactive basis functions. *Expert Systems with Applications*, 122, 183-  
34 206.  
35  
36 Perchoux, C., Brondeel, R., Wasfi, R., et al., 2019. Walking, trip purpose, and exposure to multiple  
37 environments: a case study of older adults in Luxembourg. *Journal of Transport & Health*, 13, 170-184.  
38  
39 Ramezani, S., Laatikainen, T., Hasanzadeh, K., Kytä, M., 2019. Shopping trip mode choice of older adults:  
40 an application of activity space and hybrid choice models in understanding the effects of built environment  
41 and personal goals. *Transportation*. <https://doi.org/10.1007/s11116-019-10065-z>  
42  
43 Rodriguez, D.A., Aytur, S., Forsyth, A., Oakes, J.M., Clifton, K.J., 2008. Relation of modifiable neighborhood  
44 attributes to walking. *Preventive Medicine*, 47(3), 260-264.  
45  
46 Satariano, W.A., Ivey, S.L., Kurtovich, E., et al., 2010. Lower-body function, neighborhoods, and walking in  
47 an older population. *American Journal of Preventive Medicine*, 38(4), 419-428.  
48  
49 Suminski, R.R., Poston, W.S.C., Petosa, R.L., Stevens, E., Katzenmoyer, L.M., 2005. Features of the  
50 neighborhood environment and walking by US adults. *American Journal of Preventive Medicine*, 28(2),  
51 149-155.  
52  
53 Tao, T., Wang, J., Cao, X., 2020. Exploring the non-linear associations between spatial attributes and  
54 walking distance to transit. *Journal of Transport Geography*, 82, 102560.  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 Transportation Research Board, 2005. Does the built environment influence physical activity? Examining  
5 the evidence. TRB Special Report 282, Washington, D.C., United States.  
6

7 United Nations, 2019. World population ageing 2019: Highlights. New York, United States.  
8

9 Van Cauwenberg, J., Clarys, P., De Bourdeaudhuij, I., et al., 2013. Older adults' transportation walking: a  
10 cross-sectional study on the cumulative influence of physical environmental factors. International Journal  
11 of Health Geographics, 12, 37.  
12

13 Van Wee, B., Handy, S., 2016. Key research themes on urban space, scale, and sustainable urban mobility.  
14 International Journal of Sustainable Transportation, 10(1), 18-24.  
15

16 Walker, J.L., Ehlers, E., Banerjee, I., Dugundji, E.R., 2011. Correcting for endogeneity in behavioral choice  
17 models with social influence variables. Transportation Research Part A, 45(4), 362-374.  
18

19 Wang, Z., Lee, C., 2010. Site and neighborhood environments for walking among older adults. Health &  
20 Place, 16(6), 1268-1279.  
21

22 World Health Organisation, 2010. Global recommendations on physical activity for health. Geneva,  
23 Switzerland.  
24

25 Yang, Y., Xu, Y., Rodriguez, D.A., Michael, Y., Zhang, H., 2018. Active travel, public transportation use, and  
26 daily transport among older adults: the association of built environment. Journal of Transport & Health, 9,  
27 288-298.  
28

29 Zhao, X., Yan, X., Yu, A., van Hentenryck, P., 2020. Prediction and behavioral analysis of travel mode choice:  
30 A comparison of machine learning and logit models. Travel Behaviour and Society, 20, 22-35.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

### **Credit authorship contribution statement**

Long Cheng: Conceptualisation; Literature review; Analysis and interpretation of results; Writing – original draft. Jonas De Vos: Analysis and interpretation of results; Writing – original draft. Pengjun Zhao: Analysis and interpretation of results; Writing – review & editing. Min Yang: Funding acquisition; Data curation. Frank Witlox: Funding acquisition; Writing – review & editing.