EJTIR

Issue 16(2), 2016
pp. 385-405
ISSN: 1567-7141
tlo.tbm.tudelft.nl/ejtir

Integration of unobserved effects in generalised transport access costs of cycling to railway stations

Lissy La Paix Puello[1]
Centre for Transport Studies, Faculty of Engineering Technology, University of Twente, the Netherlands

Karst T. Geurs[2]
Centre for Transport Studies, Faculty of Engineering Technology, University of Twente, the Netherlands

This paper examines the role of perceptions and attitudes in railway station accessibility. We add unobserved (latent) variables to the Generalised Transport Access Cost (GTAC) of cycling to Dutch railway stations in the metropolitan area of The Hague – Rotterdam. A hybrid discrete choice model was estimated for access mode and two latent variables which were obtained through factor analysis: perception of station environment (including factors such as the users' judgement of the station, assessment of travel information, presence of high speed trains) and perceived connectivity (including factor such as the evaluation of punctuality and the frequency of the train and quality of bicycle infrastructure). The estimated individual utility was applied to a station access cost index. A comparison between standard logit and hybrid utility functions identifies improvements in the utility-based measures by using discrete choice models. Utilities are computed by station departure, postcode of residence and neighbourhood. The results show, first, that omitting unobserved effect in utility-based measures tends to lead to overestimations of the accessibility levels. Secondly, different variations in accessibility levels are revealed, by size of railway stations and urban areas. Finally, the results highlight stronger effects of network connectivity impedances than station environmental impedances in generalised transport costs.

Keywords: *station area, accessibility, latent effects, generalised transport costs*

1. Introduction

Accessibility can be defined and measured in many different ways. Hansen (1959) was the first to define accessibility as 'the *potential* of opportunities of interaction'. Many researchers in transport research and different academic disciplines now operationalise accessibility as a function of the spatial distribution of activities and the ability and the desire of people or firms to overcome the spatial separation between activities. Geurs and van Wee (2004) distinguish four components of accessibility; a land use component, a transport component, a temporal component and an individual component. In this paper, we focus on the measurement of the transport component of

1 P.O. Box 217, 7500 AE Enschede, the Netherlands T:+ 31 53 489 2877 F: 31 53 489 4322 E: l.c.lapaixpuello@utwente.nl
2 P.O. Box 217, 7500 AE Enschede, the Netherlands T:+ 31 53 489 1056 F: 31 53 489 4322 E: k.t.geurs@utwente.nl

accessibility, which reflects the impedance of travellers to overcome the distance between spatially distributed activities.

In accessibility measures used in urban planning and geography, transport impedance is typically expressed in terms of travel distance or travel time. Applications of the Hansen-based potential accessibility measure, one of the most often applied accessibility measures; typically travel time is a proxy for impedance. In transportation research and planning more comprehensive transport impedance factors are used. Travel demand models (in particular discrete choice models) typically include a Generalised Transport Cost (GTC) function to measure impedance, represented as a set of impedance factors interacting in a GTC function and expressed in monetary or time units. In a cost index for an individual travelling between *i* and *j*, costs comprises the factors that reflect the 'effort' by the user. These are, for example, cost per kilometre, travel time per kilometre, value of unreliability and inconvenience (Koopmans *et al.*, 2013). However, transport impedance is not only related 'hard' impedance factors such as travel time and monetary cost elements, but also 'soft' impedance factors, such as (dis)comfort or physical effort of travelling, accident risk, availability of travel information. A body of choice modelling literature is emerging on the inclusion of unobserved ('latent') variables in choice models to capture attitudes and preferences co-determining travel behaviour, e.g. see applications of hybrid mode choice models by Paulssen *et al.* (2014); Espino *et al.* (2006) and La Paix and Geurs (2015). This line of research did not find its way into accessibility research. Up to now, accessibility measures only explicitly considered observable elements (Geurs and Van Wee, 2004). This implies that accessibility analysis is scarcely designed regarding individual's behaviour, preferences and attitudes. In this paper, we use an utility-based accessibility approach which allows, drawing on random utility theory, the inclusion of both 'hard' and 'soft' impedance factors in measuring accessibility, in this case to estimate an generalised public transport access cost index. To the authors' knowledge, this is the first exploratory study applying a hybrid choice modelling framework in an accessibility study.

In addition, we explore public transport accessibility from two approaches: local and regional. Many studies on public transport accessibility focus on the level of service of the main mode of transport (e.g., travel time and/or travel cost between origin and destination zones) and lack attention for the role of the quality of access and egress modes and transfer points. In accessibility studies, the importance of regional versus local accessibility and travel behaviour research has also been noted, examining local accessibility to activities within local communities and accessibility to regional centres of activities (Handy, 1993). Measures of local and regional accessibility have been developed separately, but the connection between these two has been scarcely studied. Here, we reconcile local and regional accessibility measures in a single accessibility measure. We represent local accessibility in terms of elements that determine the use of access modes, such as network infrastructure in the station area, bicycle facilities at the station, and liveliness and quality of the station. Regional accessibility is associated with farther destinations potentially reached from the station. Therefore, regional accessibility is represented by the factors that influence the use of main modes, i.e. frequency of trains, number of intercity trains stopping at the station, in-vehicle travel time, etc.

The rest of the paper is structured as follows. Section 2 discusses literature on measuring transport impedance in public transport accessibility models. Section 3 describes the study area and data used in our case study. Section 4 describes the hybrid choice modelling framework used to estimate a generalised transport access cost index. Section 5 presents the results, and Section 6 contains the conclusions of the paper.

2. Public transport accessibility and transport impedance

A large body of accessibility literature has been built from the 1950ies, and many different measures have been taken. The accessibility literature has been reviewed by several authors (e.g., see Handy and Niemeier (1997); Geurs and Van Wee (2004); Páez *et al.* (2012)), some of which focus specifically on public transport accessibility (Lei and Church, 2010; Mavoa *et al.*, 2012). Measuring public transport impedances is challenging. Ideally, the four components of accessibility as identified by Geurs and van Wee (2004) need to be taken into account: (1) a land-use component reflecting the amount, quality and spatial distribution of opportunities supplied at each destination (jobs, shops, health, social and recreational facilities, etc.); (2) a transportation component describes the transport system, expressed as disutility (or impedance) for an individual to cover the distance between an origin and a destination using a specific transport mode; (3) a temporal component reflects the temporal constraints and the time available for individuals to participate in certain activities; and (4) an individual component reflecting the needs, abilities and opportunities of individuals, e.g. depending on people's income, age, income, educational level and physical fitness.

A comprehensive measurement of public transport accessibility provides major challenges. In particular, many factors influence the transport component, and there are interactions between the four components of accessibility. Several impedance factors ideally need to be taken into account to measure the transport component. The transport component comprises time, cost and 'effort' factors (see van Wee *et al.* (2013) for an overview). A public transport trip may involve a (hidden) waiting time at origin location, access travel time and cost to get to a public transport stop, in-vehicle travel time and costs, transfer walking and waiting times, and egress travel time and costs to get to the final destination. Public transport trips also involve 'effort' impedance factors such as (mental) strain, stress, reliability, physical effort (stairs, carrying luggage, etc.) and feelings of safety. These 'effort' factors influence the perception of public transport trips and can also influence time valuations. Public transport waiting time can for example be perceived as especially burdensome when travellers have to wait in uncomfortable environments, such as in cold, warm or rainy weather, or in seemingly unsafe or insecure conditions. Individuals and specific population segments (e.g. commuters, elderly, man and women) will value these transport impedance factors differently, thus creating interactions between the transport and individual components of accessibility.

Comfortable environments can also affect time valuations. Cascetta and Cartenì (2014), for example, show a significant impact of stations architectural quality on the valuation of waiting times at stations; a commuter is willing to wait up to 7 min more, or to spend 10 min more to reach an high aesthetic station. These results suggest that stations architectural quality affects both accessibility by public transport (lower perceived travel time) and accessibility to public transport (a greater catchment area). Cartenì *et al.* (2014) also show that the catchment area of high architectural quality of metro Line 1 stations in Naples (also called "Metrò dell'arte") is twice as large as traditional metro stations. In addition, Van Hagen (2011), for example shows that waiting experience is also a cognitive and affective process which can be influenced. By adding environmental stimuli in the shape of music, advertising, infotainment and coloured light, passengers find the wait more enjoyable, useful and pleasant. In a visually stimulating (busy) environment, however, that same music affords too much arousal, which can lead to mental overload and a more negative station evaluation.

Applications of public transport accessibility are typically GIS-based and focus on travel distance and travel times as observed impedance factors. For example, Gutiérrez (2001) estimated gravity-based accessibility measure to estimate the effects of a high-speed train line (Gutiérrez, 2001; Linneker and Spence, 1992), distinguishing distance and time. Gutiérrez *et al.* (2011) studied the accessibility of transport nodes, called nodal accessibility, based on travel times between nodes in the network. Hadas and Ranjitkar (2012) estimate a transit connectivity index based the transport costs on travel time between stations and transfer time between transport modes. The literature only recently acknowledges that not only the places and opportunities that can be reached by transit (i.e. accessibility by public transport) need to be taken into account in accessibility studies but also accessibility to public transport (walking, biking, etc.). See for example Mavoa *et al.* (2012) for an example of a GIS-based public transport accessibility measure, in which a walkability and public transport accessibility index are combined.

The concept of generalised transport costs was introduced as a function of time and distance (Nichols, 1975), as representation of the main forces restricting transport. The concept of transport cost has evolved over the years to more elaborated measures, such as direct costs, indirect and economic costs (Combes and Lafourcade, 2005; Zofío *et al.*, 2014), and it represents an important step towards a more comprehensive approach to measure the accessibility impacts of policy strategies (Koopmans *et al.*, 2013). De Keizer *et al.* (2012), for example, used transfer time, waiting time and frequency to represent the connectivity of train stations. Accessibility studies which are based on outputs from transport demand models or land-use/transport interaction models typically use the generalised transport cost concept, using value-of-time values by trip purpose (e.g., see Wang *et al.* (2015); Zondag *et al.* (2015). However, to the authors' knowledge, accessibility studies have so far focussed on time and cost factors as 'observable' factors of generalized cost. Many effort impedance factors, attitudes to transport mode and perceptions of transport impedance factors are not included and interpreted as 'unobservable' elements.

The challenge is the number of difficulties experienced to represent the '*soft*' elements in the generalised transport cost, precisely because they are not observed. One cannot account units of comfort, units of quality of a train station, or units or connectivity. By definition, gravity-based and network-based accessibility measures are restrictive to parameters different than travel time, transfer time and cost. Utility-based accessibility measures, based on discrete choice models, do allow the inclusion of the individual's perspective to the generalised transport cost. One can measure transport impedance as the 'dis*utility*' of reaching a destination *j* from an origin *i*. The utility of *j* can be weighted by the users who use *j* as destination from different *i* points. As any function, it can be estimated via a set of coefficients, following a predefined form, i.e. logit. An advantage is that the value of utility is unitless. Then, the function can incorporate zonal elements, level of service and attitudes of users (*soft* elements). Therefore, one tacitly states that users may select the alternative (destination, station, transport mode, etc.) that maximises their benefit considering both *hard* and *soft* elements according to random utility theory.

The main issue with utility-based measures is the specification of the utility equation, which means the composition of impedance factors (both *observable* and *soft* factors). Earlier works on public transport access can serve to determine these factors. For example, the transport modes used to get the station, i.e. shares of bike-and-ride, can vary substantially, even between stops or stations of the same type of public transport (Martens, 2004). The built environment of the station (density, diversity, etc.) can also vary strongly, which is an important factor in defining the access mode choice to public transport (Monteiro and Campos, 2012). Furthermore, service and quality of the railway station and trains play an important role in customer satisfaction, railway station choice and propensity to use rail (Brons *et al.*, 2009; Debrezion *et al.*, 2007). The challenge, however, is an efficient and consistent inclusion of quality, and other '*soft*' elements, as an impedance factor in the generalised transport access costs. An extended framework for integrated discrete choice and latent variables models, called hybrid choice model, was proposed by Ben-Akiva *et al.* (1999) and generalised by Walker and Ben-Akiva (2002). Especially during the last few years, an increasing number of hybrid choice models have been developed in the choice modelling literature, e.g. see

Paulssen *et al.* (2014) and Glerum *et al.* (2014). Hybrid choice models have proven to be more accurate than standard choice models by incorporating *unobserved* effects. However, Chorus and Kroesen (2014) also note the data are almost without exception cross-sectional as far as the latent variable is concerned, and as such do not allow for claims concerning changes in the variable at the individual level.

An hybrid choice model for access mode choice was developed in earlier work by La Paix and Geurs (2015). The discrete choice was a binary choice for access by bicycle to the train station. The latent variable model, representing the '*soft*' elements, had two forms: *perception* of network connectivity and *attitude towards station environment*. The hybrid choice modelling framework is able to cope with the psychological interpretation of such elements (quality, connectivity, etc.) that vary across individuals. This paper builds upon this work, aiming to explore the added value of incorporating these 'soft' impedance factors in an access cost index of bicycle accessibility to railway stations.

3. Study area, data and survey

Our study covers 35 stations in the wider metropolitan area of Rotterdam - The Hague in the Netherlands. This area is also known as Randstad South in Dutch policy and planning documents. The metropolitan region of The Hague-Rotterdam and surroundings comprise 3 million residents and is one the most urbanised area in the Netherlands (see the demarcation of the area in Figure 1). The analysis in this paper is mainly based on data from the Netherlands Railways (NS) customer satisfaction survey, covering the years 2009-2011. The database comprises survey data of about 12,000 train passengers in the study area. Figure 1 shows the distribution of the respondents over the station areas.

The variables considered comprise transport and land use variables, socio-economic variables and psychometric variables. These variables are briefly explained as follows.

Transport and land-use variables
Three types of transport and land use variables are used: travel related variables, station related variables and station areas variables:

- Travel related variables: this information comes from the NS customer satisfaction survey. The travel-related attributes include journey characteristics, such as travel time, rush hour and payment method.
- Station characteristics: this information corresponds to the station of departure, such as: the number of bicycle parking facilities at the station, number of high-speed trains stopping at the station, lighting, availability of travel information, the number of intercity trains and the station size based on the Netherlands Railways classification.
- Zonal variables of station area: number of job positions available within a radius of 5 kilometres of network distance, and status of cycle routes. The information of job positions is obtained from LISA database, which contains employment details on all individual firms in the Netherlands (LISA, 2012). The characteristics of the cycle route are obtained from a databased from the Dutch Cycling Union (Fietsersbond, 2011); which contains information about quality and nuisance of the cyclist route to the station, among others.

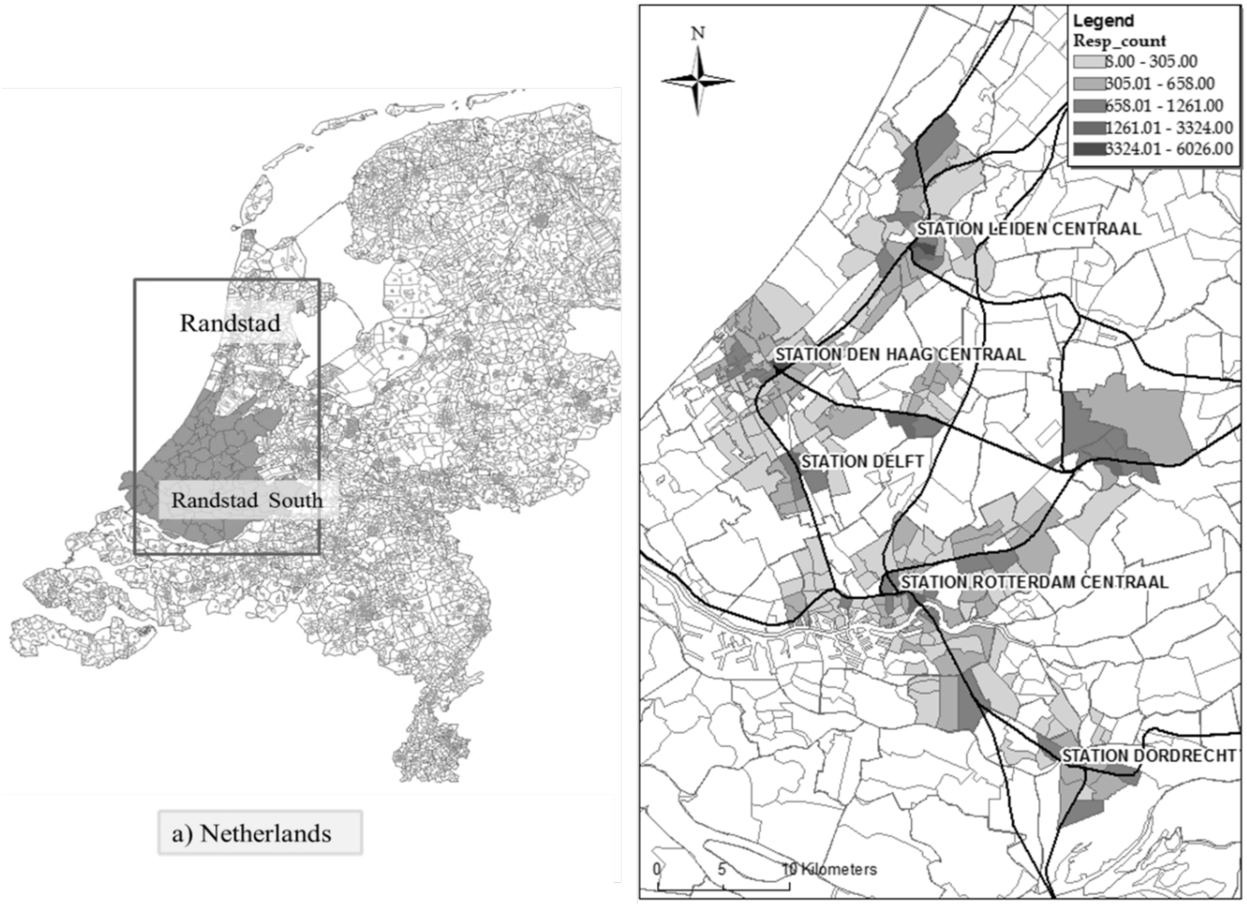


Figure 1. Stations in the survey area of The Hague – Rotterdam (Randstad South)

The station characteristics and zonal variables were grouped in the 5Ds of transit oriented development (TOD concept), developed by Ewing and Cervero (2010), namely: density, diversity, design, distance to transit and access to destinations. The TOD concept can be considered a well-known and complete framework for the analysis of public transport accessibility. Figure 2 shows the distribution of these variables in the model specification, which is explained in the following section.

Socioeconomic variables
This group of variables considers the traditional socio-demographic variables, such as age, gender and occupation, which were obtained from the NS survey.

Psychometric indicators
This data come from the Netherlands Railways (NS) customer satisfaction survey, which is carried out annually. The survey contains information about the travellers' valuation of train journeys and station facilities. The individual replies on a 10-point Likert scale, in which 0 means 'cannot be worse', and 10 means 'excellent'. These statements try to capture the individual's impression, positive or negative. Two latent variables are estimated via a factor analysis, see Section 4.

4. Hybrid choice models for generalised transport access (GTAC)

In this case, the probability function follows the structure of a binary logit model of the logit family. Three models were estimated: (1) standard logit model of cycling, (2) a hybrid model including '*perceived connectivity*' as latent variable, and (3) a hybrid model including '*attitude towards station environment*' as latent variable. Generalised transport cost are represented by the utility of cycling to the station, U_{ijk}. Utility can be treated as the sum of the systematic, representative or observable part (U_{ijk}), which is a function of the attributes, and the random component:

$$U_{ijk}^{n} = V_{ij} + \varepsilon_{nk} \tag{1}$$

Therefore, the utility of choosing bicycle (alternative k) is a function of the explanatory variables that vary individually and across the station of departure (*j*), and zonal elements Z_{jn}, both socioeconomic and trip related characteristic S_n, and the *latent* variable (X_n^*) are expressed as follows:

$$f_T\,(U_{ijk}^{n}\,|S_n, Z_i, X_n^*\,;\,\beta_k) \tag{2}$$

Here, U_{ijk} is the utility for individual n for the alternative k, from the residence postcode i, to the station of departure j. Note that n is omitted to simplify the equations. Z_i is the vector of parameters associated to the characteristics of station (i).

The latent variables, *soft* factors, are a linear function of both transport service aspects X_n and satisfaction statements S_n. Two latent variables are estimated (1) *attitude towards station environment* (X_n^s), which is; (2) *perception of connectivity* (X_n^c). The latent variables are given by the formula:

$$X_n^* = h(S_n\,, X_n; \lambda_n) + \omega_n \tag{3}$$

In this equation, X_n^* is the generic expression of latent variables. λ_n is a vector of unknown parameters to be estimated, and ω is the random disturbance term, normally distributed, with variance. The estimated vectors of parameters λ_n and β_k allow to calculate the utility. As in any hybrid choice model, for the latent variable model, we need the distribution of the latent variables given the observed variables, X_n, which can be measured by a *measurement model* with the corresponding indicators. Those indicators were obtained via factor analysis, and discussed in La Paix and Geurs (2015). The α parameters in table 1 belong to the measurement model.

If we assume that the error terms are identically and independently distributed (i.i.d.) and are of extreme value type 1 (EV1), then the typical logit model is obtained, as follows:

$$\boldsymbol{P_{ni}} = \frac{\boldsymbol{e}^{\theta U_{ijk}}}{\sum_{\boldsymbol{m}\in \boldsymbol{c_n}} \boldsymbol{e}^{\boldsymbol{\theta} U_{ijm}}} \tag{4}$$

Here, θ >0 is the scale parameter of the EV1 distribution. This model or equivalent variants of it can be derived in a large number of ways. And m is the vector of alternatives in the choice set C_n.

In Figure 2, terms in ellipses represent unobservable (latent) constructs, whereas those in rectangles represent observable variables. The right portion of Figure 2 is the latent variable model. The latent variable is denoted by X_n^* for individual n. X_n^* is unobservable, but the observable variable indicator (I_n) is the materialisation of the latent variable. The dashed arrow from the latent variable to the indicator is the measurement model. The indicator is only used to test the estimation of the latent variable; it is not used in the model estimation itself. Thus, the indicator is used to identify the latent variable, and is introduced as unobserved construct in the discrete choice model by the structural equation, represented by the solid arrow from X_n^* to choice model. The latent variable is predicted via a set of parameters.

As can be seen in Figure 2, the groups of Design, Distance and Diversity enter in both choice model and latent model. Via the simultaneous estimation of both latent and choice model, we allow the latent variable model to estimate the best parameter for each explanatory variable as function of the psychometric indicators (F1 and F2).

The variables were distributed between both latent and choice models according to, firstly, their nature and location. Specific variables can affect both choice and latent model. For example, variables related to connectivity with other public transport modes (BTM lines, sprinter trains, etc.) might affect the bicycle choice directly (competitors of bicycle) but it could also affect the perceived connectivity of the station, manifested through the indicators. At the same time, connectivity variables are used as proxy of station type and size (e.g. larger stations receive more intercity trains, and are also endowed by better BTM connection).

The difference between including a parameter in the choice model or latent model is: the first one influences the choice directly, while the second one is influencing the choice through multiple indicators, which is more powerful. Table 1 shows the distribution of each specific variable within the models.

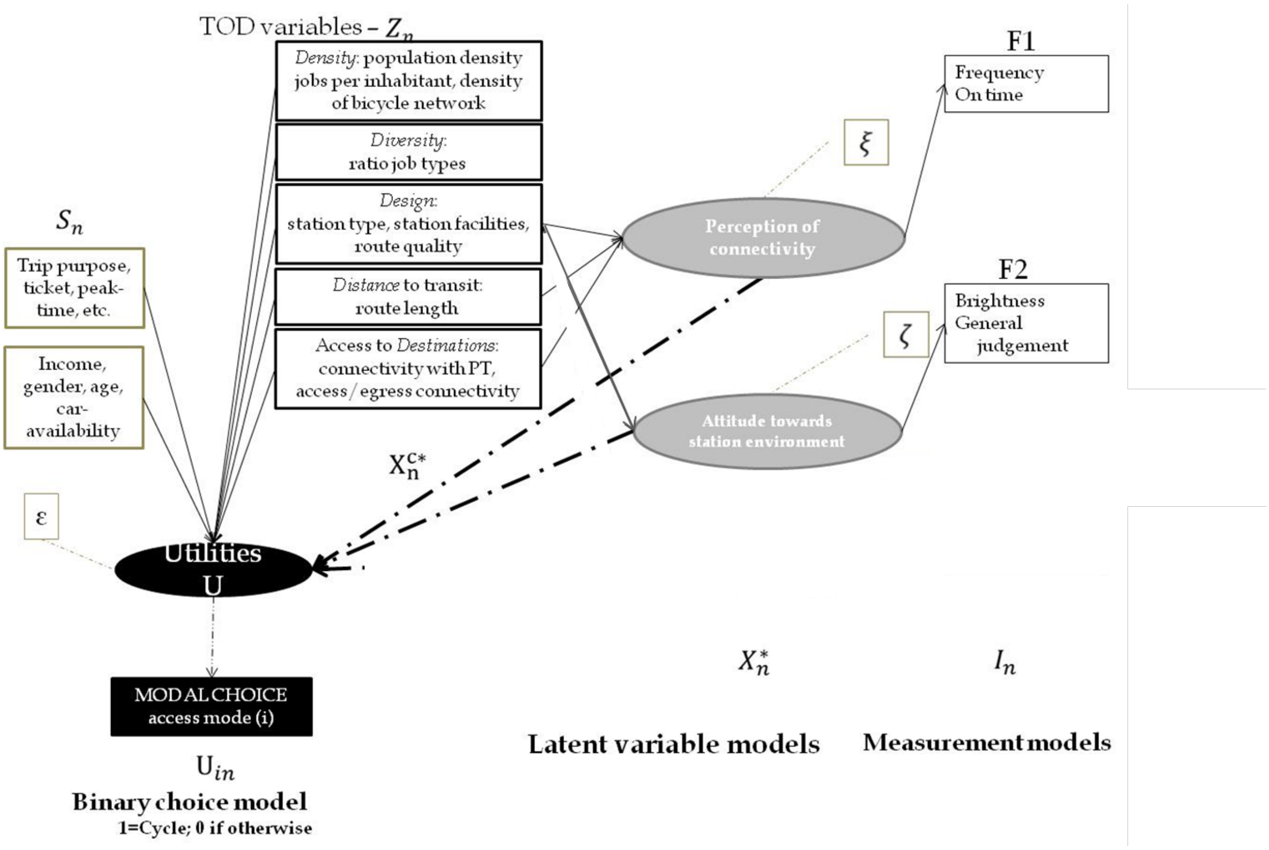


Figure 2. Analytical framework of the hybrid choice model

4.1 Calculation of GTAC index

The access cost accessibility means the generalised costs to reach a station by bicycle, weighted by the probability of cycling. This measure is calculated by station, the rank of $GTAC_i^l$ is called access cost index. This measure represents the transport costs faced by residents in postcode j to access the station i (Eq. 2), weighted by the probability of a cyclist journey between i and j. The probability is calculated with the discrete choice model (Eq. 4). Therefore, the GTAC index is expressed as follows:

$$\mathrm{GTAC}\,_i^l = \sum_j p_{ij} U_{ijk}^l \qquad (5)$$

In this equation, p_{ij} is the probability of cycling (alternative k) to station j from the origin i, and. U_{ijk} is the utility or generalised cost of cycling between i and j. The variations of l represents the three types of utility functions where l takes the value 0 if the utility is taken from the binary logit model; and l takes the values 1 and 2 if the utility is calculated via the hybrid choice models of perceived network station environment and connectivity, respectively.

4.2 Estimation of hybrid choice models

As explained in the introduction, the modelling framework of this paper concerns the access mode choice to railway stations. The choice set is binary, composed of either using the bicycle to access the train station or accessing the station otherwise. The utility function of the choice model consisted of variables related to the journey (purpose, time, payment method for train fare, length of the access route to the station, etc.), the station (station type, BTM lines, etc.), and station area (i.e. types of job positions available in the station area). The two latent variables are integrated in two separated choice models. The measurement model of the latent variables consists of psychometric indicators. The latent models are described as follows:

- The latent variable of *perceived connectivity* is estimated via the zonal variables such as: quality of access roads; and station characteristics related to connection with other transport modes, i.e. number of intercity trains, quality of unguarded and guarded bicycle parking, and number of local trains stopping at the station. The psychometric indicators were two: evaluation of *punctuality* of the train and the evaluation of *frequency* of the train.
- The latent variable of *attitudes towards station environment* is estimated via the explanatory variables related to the station characteristics. For example: lighting, number of bicycle places, availability of travel information. We used the *general judgement* of the station as the psychometric indicator in the measurement model.

Table 1 shows the model results. The first section presents the estimated parameters in the utility function of the binary logit model; second and third section presents the estimated parameters in the latent variable models. As can been, all the variables have the expected sign. Variables related to the journey (rush hour, education or working purpose of the trip, etc.) are positively associated with bicycle use. By contrast, length of the access journey and connection with other public transport modes (BTM lines, intercity and HST trains) have a negative effect on cycling to the station, as shown in Table 1. Also, the station size (type 1) has a negative effect on both the cycling choice and latent model. The station Type 1 defines the largest station size, which are endowed with better BTM connection, therefore a negative association with bicycle choice is expected. HST is added in both connectivity and station latent models. As we can observe, there is a negative influence since HST trains characterize large stations.

The goodness of fit of the estimated models is shown at the bottom of Table 1. The best goodness of fit is reached in the HCM-connectivity model. Also, this model is better than the standard logit model. However, the performance of HCM-station model is the lowest one. For further details, the model results are discussed in La Paix and Geurs (2015).

Table 1. Variables included in the utility functions Source: La Paix and Geurs (2015).

	Binary logit: Bicycle U_0		HCM Station U_1		HCM Connectivity U_2	
Name	**Value**	**Robust t-test**	**Value**	**Robust t-test**	**Value**	**Robust t-test**
Variables in binary choice model: 'cycling'						
ASC1	-0.457	-1.43	-1.06	**-2.96**	-3.67	**-17.20**
$\beta_{attConnectivity}$					0.25	**8.99**
$\beta_{attstation}$			0.10	**3.76**		
β_{age}	-0.01	**-4.28**	-0.01	**-3.88**	-0.01	**-4.87**
β_{male}	0.04	0.29	0.04	0.97	0.02	0.53
β_{work_motive}	0.48	**8.03**	0.51	**9.15**	0.55	**9.78**
$\beta_{business}$	0.63	**7.68**	0.63	**7.59**	0.61	**7.41**
β_{School_study}	0.41	**5.72**	0.27	**5.17**	0.27	**5.15**
$\beta_{discount}$	0.26	**4.92**	0.36	**5.04**	0.44	**6.25**
$\beta_{studentcard}$	-0.16	**-2.53**	-0.21	**-3.19**	-0.22	**-3.39**
$\beta_{Rush\ hour}$	0.27	**5.26**	0.28	**5.67**	0.27	**5.52**
β_{car}	-0.08	-1.86	-0.08	-1.62	-0.10	**-2.23**
$\beta_{value\_time\_2}$	0.07	1.55	0.08	1.55	0.00	0.68
Density						
$\beta_{bicycle\ density}$	-0.24	**-4.61**	-0.24	**-4.62**	-0.20	**-3.92**
$\beta_{population}$	-0.01	-1.16	-0.01	-1.19	-0.03	**-4.05**
$\beta_{dwellings}$*(per squared km)*	-0.01	-1.21	-0.52	-1.19	0.002	0.57
β_{jobs}	0.40	**3.26**	0.38	**3.14**	0.20	**2.51**
Diversity						
β_{WP_Health}	-0.21	-0.44	-0.238	-0.49	-0.185	-0.38
β_{WP_public}	-3.23	**-4.98**	-3.32	**-5.08**	-2.95	**-5.02**
β_{WP_retail}	1.18	1.25	1.05	1.12	0.605	0.66
Design						
$\beta_{Averageroad\ quality}$	2.40	**6.63**	2.40	**7.78**	2.75	**9.15**
$\beta_{Averagetraffic\ nuisance}$	-1.04	-0.26	-1.04	**-4.60**	-1.00	**-4.49**
Distance (to public transport)						
$\beta_{length3km}$	-0.35	**-6.21**	-0.35	**-6.12**	-0.33	**-5.97**
Destinations (accessibility)						
β_{BTM_lines}	-0.03	**-5.24**	-0.03	**-6.30**	-0.03	**-6.45**

	Binary logit: Bicycle U_0		**HCM Station U_1**		**HCM Connectivity U_2**	
Name	**Value**	**Robust t-test**	**Value**	**Robust t-test**	**Value**	**Robust t-test**
$\beta_{station\_Type\_1}$	-0.27	**-4.33**	-0.37	**-8.29**	-0.11	**-2.69**
$\beta_{Bicycle\ parking\ spaces)}$	0.002	**9.35**	0.002	**2.08**	0.002	**9.67**
Variables in latent models						
$\beta_{meanAtt}$			6.15	**43.83**	5.87	**42.52**
Design						
$\lambda 2_{NSstationType1}$			-0.24	-1.83		
$\lambda_{easytofindtravelinformation}$			0.03	0.37		
$\lambda_{Bicycleparking}$			0.06	**21.03**		
$\lambda_{lightingstationenvironment}$			-0.003	-0.04		
$\lambda_{qualitybikeaccessroads}$					-0.02	-1.26
$\lambda_{qualityguarded\ parking}$					0.003	0.79
$\lambda_{qualityunguarded}$					0.08	**15.29**
Destinations (accessibility)						
$\lambda_{IC_{trains}}$					-0.01	**-2.10**
$\lambda_{Sprintertrains}$					0.05	**2.96**
λ_{HST_trains}			-0.15	**-10.31**		
σ_{LV}			0.06	**2.98**	0.23	**11.59**
Measurement model						
$\sigma_{punctuality}$					0.199	**12.71**
$\sigma_{frequency}$					0.234	**13.14**
$\alpha_{frequency}$					0.368	**20.14**
$\sigma_{safety\ night}$			0.805	**72.54**		
$\alpha_{safety\ night}$			0.297	**18.33**		
σ_{light}			0.182	**10.11**		
$\sigma_{gral\ judgment}$			0.241	**11.75**		
Fit measures						
Number of estimated parameters	24		36		35	
Rho-bar	0.157		0.215		0.330	

5. Applications of the GTAC

This section analyses the utilities and the GTAC from each model. Since in any discrete choice model the absolute value of utility is irrelevant; to account for this fact, the researcher must normalize the scale of utility (Train, 2003). To overcome this inherent limitation, we firstly stablish natural breaks to compare the levels of utilities within each model in section 5.1. We avoid the comparison between absolute values of utilities from different models. Secondly, section 5.2 analyses the GTAC by which the probabilities obtained from eq. 4 consider the differences between alternatives in the choice set (bicycle or otherwise), according to *eq*. 5. Additionally, in the section 5.3, we normalize the GTAC as percentage over the average in the network to compare the results within models, instead of across models. Section 5.4 analyses the rank of the GTAC over the 31 stations in the sample.

5.1 Weighted utilities by postcode: analysis station catchment area

Figure 3 shows the average utility by postcode level according to the closest station of departure. The shortest route from the residential postcode of each respondent was calculated. Then, each postcode is linked to one station, assuming that respondents would use the closest station. The utility by postcode is normalized as percentage over the total. Also, a weight is applied according to the number of population.

Figure 3 is a clear representation of the distribution of transport costs over the station influence area. As the figure shows, distance from the station is not the only element that influences transport impedances of accessing the station by bicycle. Therefore, a cost function including only cost or time would be biased. Several postcodes located close to the station show a high level of cost penalties. By contrast, other catchment areas present the same impedance levels from the different postcodes in the area. Therefore the impedance is independent of the distance from the postcode to the station. However, in a number of cases, the postcodes located farther from the station present high impedance factors.

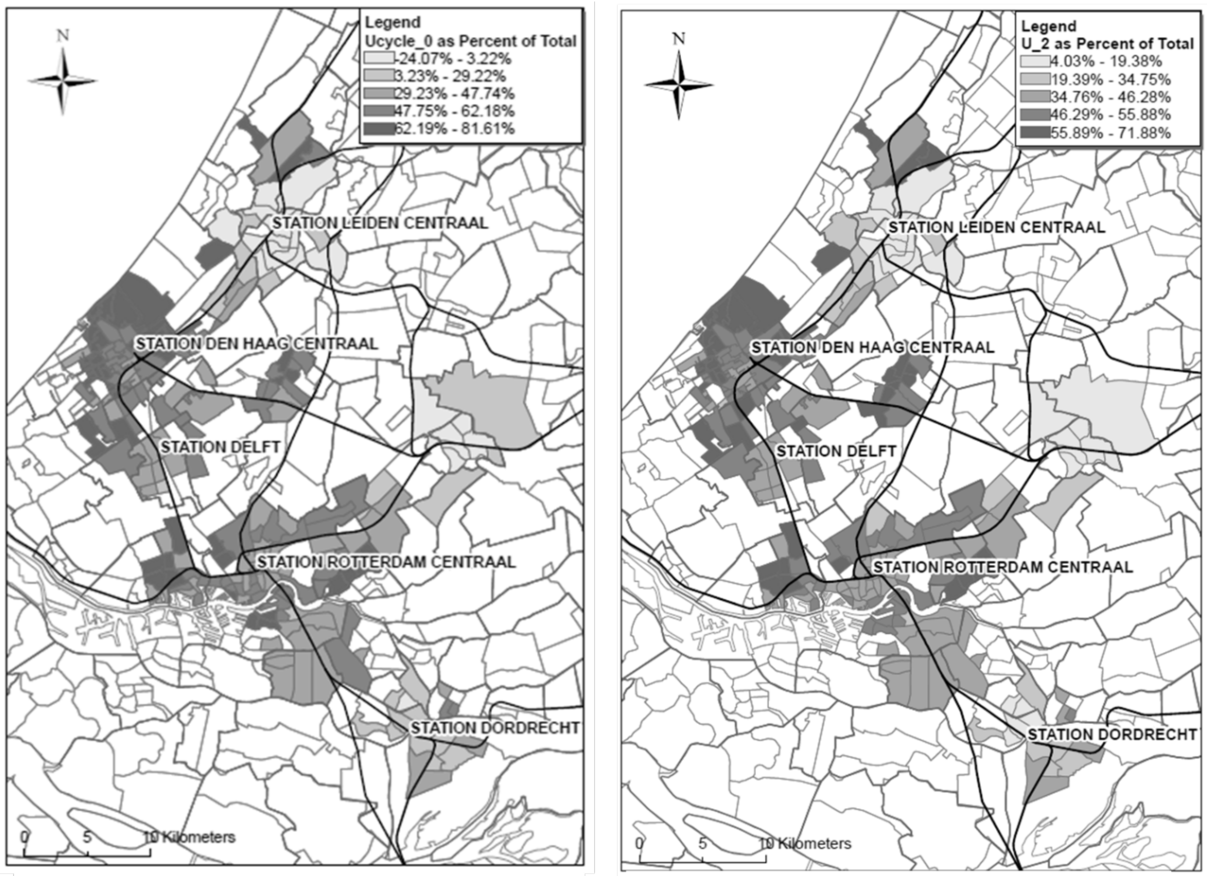
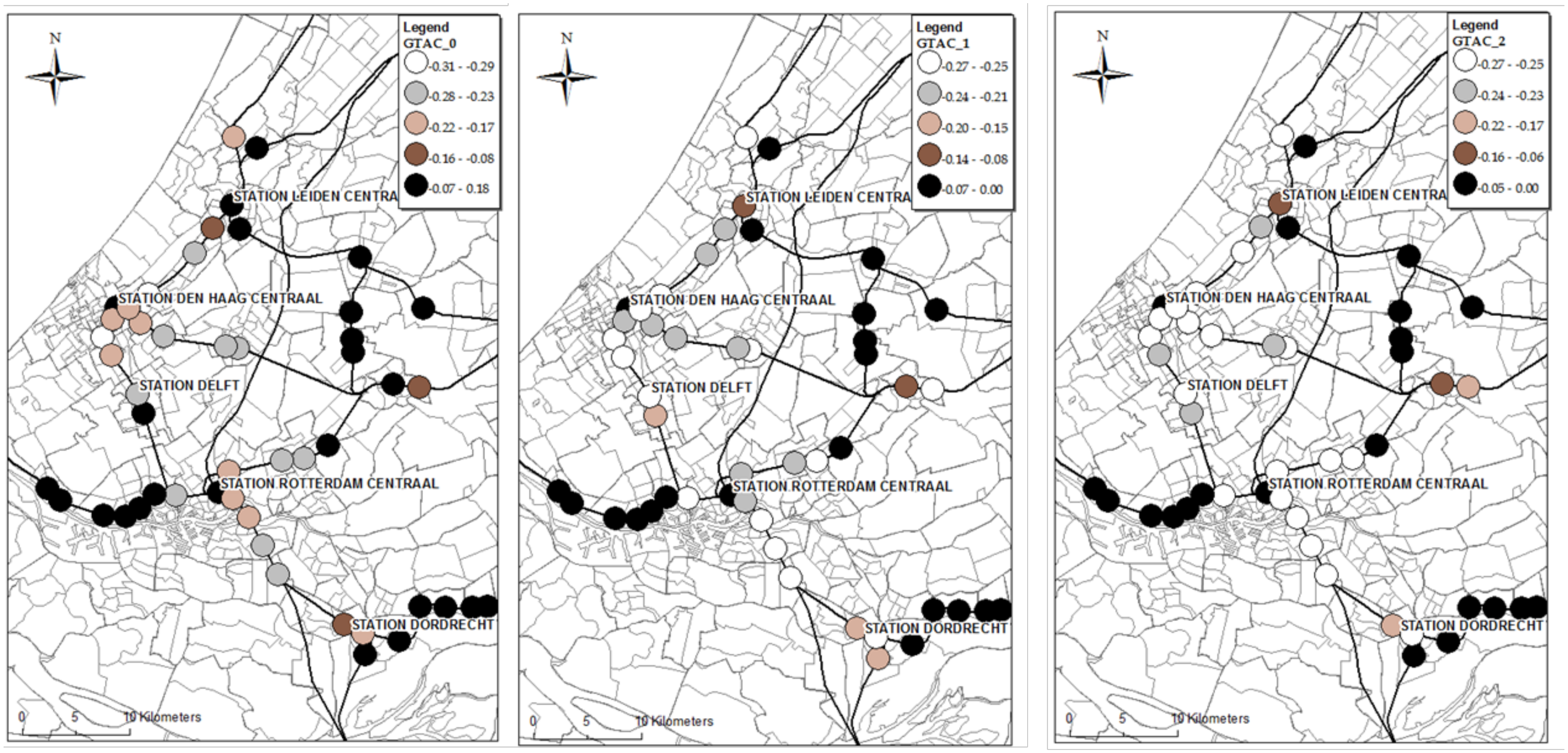


Figure 3. Utilities of accessing the railway station by postcode, calculated with binary and hybrid choice models. Normalization as percentage over the total.

5.2 GTAC by station and station type

Figure 4 shows the average GTAC for the stations in the study area, weighted by the probabilities of cycling. The results show that the impedances produced by a combination of observed and unobservable elements, i.e. network connectivity (U2) and station environment (U1), substantially differ from the impedances produced by only observable elements. The GTAC calculated by the hybrid choice model including 'perceived connectivity' is on average higher than the GTAC calculated by the hybrid choice model including 'attitudes towards station environment'. The differences between stations can be further illustrated by classifying the GTAC by station type. We use six station types, following the classification of stations from Netherlands Railways by the number of train users, city size and location of the station in the city.

Table 2 shows the total utility by station type, weighted by the number of respondents. The table shows that larger station size offer lower accessibility levels for bicycle journeys. According to the modelling work, the inclusion of the station assessment tends to decrease the value of the utility. Lower GTAC values are measured for smaller stations, such suburban stations which mainly act as a departure point (station type 5).



a) GTAC U_0 b) GTAC U_1 c) GTAC U_2

Figure 4. Comparison of average GTAC from U_0 (logit), U_1 (station) and U_2 (connectivity)

Table 2. Average transport costs (disutility) by station type

	Examples of stations	U_0 % over Average	Station U_1 % over Average	Connectivity U_2 % over Average	Count N
1- Very large station in the centre of a large city	Rotterdam CS, the Hague CS	163%	130%	134%	4323
2- Large station in the centre of a medium-sized city	Leiden CS, Den Haag HS	24%	53%	56%	3983
3- Suburban station	Rotterdam Alexander	146%	164%	132%	902
4- Station in a small town or village	Rijswijk, Zoetermeer Delft Zuid	107%	103%	104%	618
5- Suburban station with departure function	Rotterdam Noord, *Den Haag Moerwijk*	83%	97%	92%	884
6- Station outside a small town or village	Barendrecht, Voorschoten	104%	96%	102%	182
Total					**10892**

Table 2 shows that most of the larger stations are linked with high GTAC levels, but we can observe a few exceptions when other stations are at close distance and catchment areas are overlapping. Stations in medium-sized cities are more attractive for cyclists than stations in big cities. For example, the probability to cycle to the station in The Hague is 15%, which is less than half of the percentage for Leiden (44%). This is consistent with the impedance factors. The average disutility faced by travellers from The Hague Central station is -1.8, which is more than the average disutility in the network captured by Leiden (-0.27). As can be observed in Table 2 and Figure 4, large stations in the centre of medium-sized cities (i.e. *Gouda* and *Leiden*) are the most accessible stations by bicycle in our sample. Similarly, suburban stations present high accessibility levels by bicycle.

Another way to compare the different GTAC indices by station is to provide a ranking, instead of comparing utility values directly. Table 2 provides this ranking by station type, Figure 5 shows the rank of *access cost index* for 31 train stations in the network, ordered from largest to smallest in GTAC-U_0. The smaller the rank values, the better accessibility. As can be observed in the figure, the rank of access cost fluctuates from one index to another. We can also deduct that, in a number of stations, the rank based on GTAC-U_1 is more pessimistic than the ranks based on GTAC-U_0 and GTAC-U_2. Similarly, the rank based on U_2 is the strictest one. The main conclusion that emerges from here is that *perceived connectivity* is an important attribute of accessibility measures. The ranking process allows comparing the different access cost indices. Therefore the differences highlight the potential biases that one could incur latent perceptions are excluded from accessibility measures.

Table 3. Rank of GTAC by station type

Station Type	N	Rank with U_0 - logit	Rank with $U_{1\text{- station}}$	Rank with $U_{2\text{- connectivity}}$
1- Very large station in the centre of a large city	2	6	6	4
2- Large station in the centre of medium-sized city	5	1	1	1
3- Suburban station	4	5	3	5
4- Station in small town or village	4	2	2	2
5- Suburban station with departure function	13	3	5	3
6- Station outside a small town or village	2	4	4	6

Furthermore, Table 3 clearly illustrates that the relative position of station types varies with perceived station access and perceived connectivity. The largest stations in the network, the Hague CS and Rotterdam CS, have better network connectivity but a poorer perceived station environment, which is partly due to poor bicycle parking facilities in the period 2009-2011. The NS customer satisfaction survey also confirms that train users are not satisfied with bicycle parking facilities at large railway stations as parking demand at these stations exceeds supply. In contrast, large stations scored better once considering the connectivity of the station in the calculations of *GTAC* from the Eq.5. In contrast, smaller stations (type 6) have lower connectivity by public transport and are better accessible by bicycle; therefore those stations scored lower in the GTAC-U_2 than GTAC-U_0.

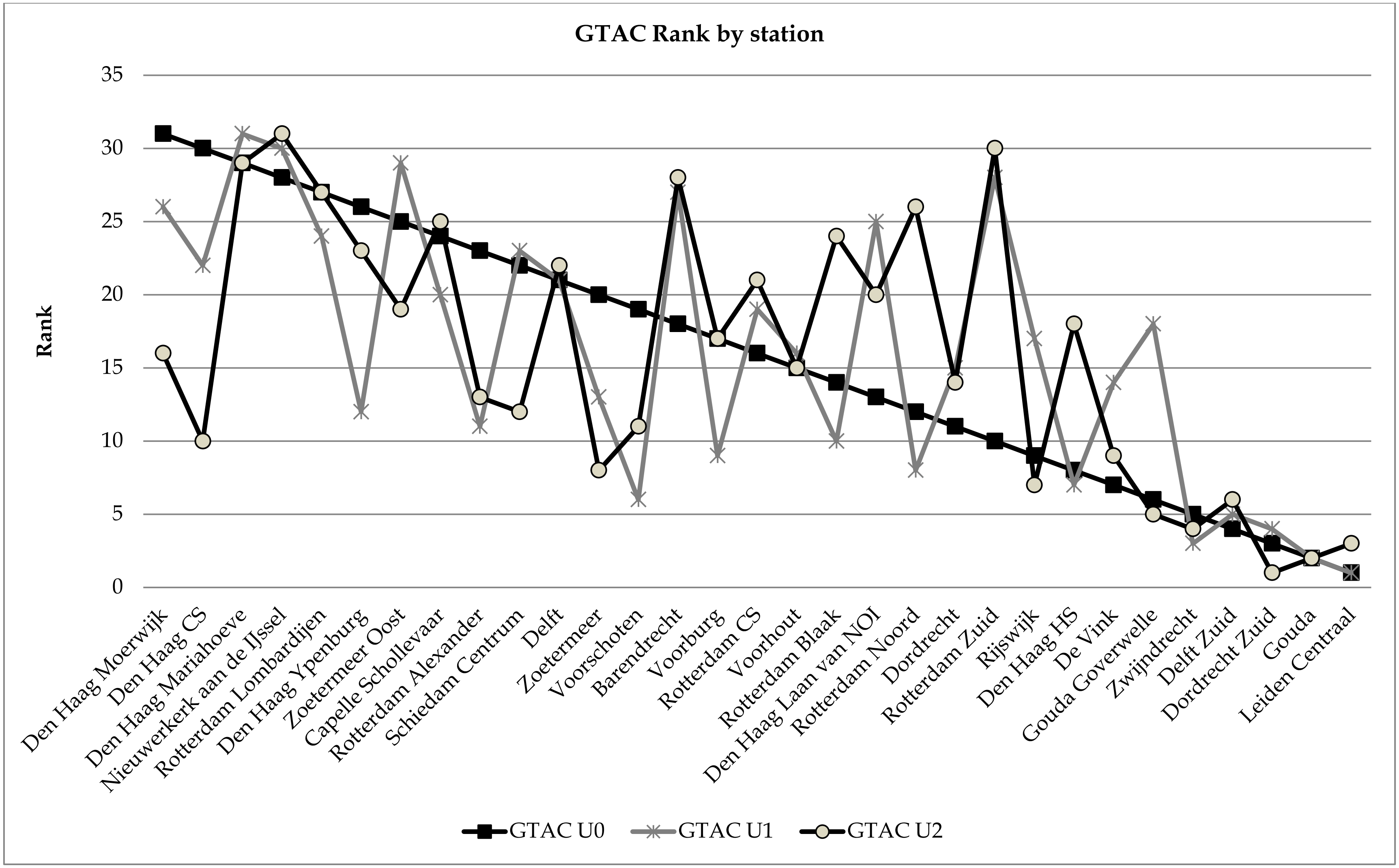


Figure 5. Rank of train stations by GTAC index

6. Conclusions and discussion

In accessibility literature, most of the studies use travel time as impedance factor, or a generalised cost combining travel time and travel cost. In this exploratory paper, we examined the added value of incorporating 'soft' impedance factors in railway station accessibility, i.e. 'perceived network connectivity' and 'perception of station environment', using a hybrid choice modelling framework. In addition, we examined the role of public transport accessibility at the regional level (represented by factors such as number of intercity trains stopping at the station, and in-vehicle travel time) on local station accessibility by bicycle.

The results show that, firstly, user perceptions of network connectivity and, to a somewhat smaller extent, the quality of station environments have a clear impact on Generalised Transport Access Cost levels. The analysis of GTAC by station type shows that bigger stations have better network connectivity but have larger impedances for cyclists. This is mainly due to poor quality of parking facilities provided at these stations. The spatial representation by neighbourhood and postcode allows demonstrating the variations within the catchment area of the stations. For example, there is observed heterogeneity in travel cost among neighbourhoods located in the same distance band from stations of departure. Therefore, the results show the potential bias of basing cost functions in accessibility modelling only on distance or cost.

In future research, this work could be extended to egress and to verification of the direct implications of these impedances on transit share. Secondly, a joint accessibility measure could be elaborated. The accessibility from *i* to *j* would include the impedances given by access and egress, additional to the transport cost given by in-vehicle time. Similarly, based on this choice model a log sum measure would be a step forwards this measure, which would provide the utilities from an origin *i* to all the possible destinations *j*. Fourthly, the ranking of the access cost index developed in this paper allows the analysing in relative terms, but the monetization of the utility would provide more clear interpretation.

In this paper we illustrate that accessibility models can be more theoretically and behaviourally sound by including both hard and soft impedance factors, but using complex hybrid choice models at the same time it increases its complexity, cost of calculation, data need and the difficulty of interpretation. Balancing rigour and usability is a well-known issue in accessibility studies, e.g. see for discussions also Handy and Niemeier (1997); Geurs and van Wee (2004); Hull *et al.* (2012). Geurs *et al.* (2010) concluded earlier that complex utility-based accessibility measures are an elegant and convenient solution to measuring accessibility when a discrete travel demand model is already available. The barriers of including soft factors in accessibility models will be probably only be overcome if hybrid choice models are becoming the state of the practice in transport demand modelling.

Acknowledgments

The authors want to thank two anonymous reviewers for valuable comments on an earlier version of this paper. We also thank the Netherlands Railways for allowing us to use Customer Satisfaction Survey data. This research was funded by the Netherlands Organisation for Scientific Research (NWO) as part of the Sustainable Accessibility of the Randstad (SAR) programme.

References

Ben-Akiva, M., McFadden, D., Gärling, T., Gopinath, D., Walker, J., Bolduc, D., et al. (1999). Extended Framework for Modeling Choice Behavior. *Marketing Letters, 10*(3), 187-203.

Brons, M., Givoni, M., and Rietveld, P. (2009). Access to railway stations and its potential in increasing rail use. [doi: DOI: 10.1016/j.tra.2008.08.002]. *Transportation Research Part A: Policy and Practice, 43*(2), 136-149.

Cartenì, A., Galante, G., and Henke, I. (2014). The catchment area of high architectural railway stations: An application of the Graham scan algorithm. *WIT Transactions on the Built Environment*, 463-474.

Cascetta, E., and Cartenì, A. (2014). The hedonic value of railways terminals. A quantitative analysis of the impact of stations quality on travellers behaviour. *Transportation Research Part A: Policy and Practice, 61*(0), 41-52.

Chorus, C. G., and Kroesen, M. (2014). On the (im-)possibility of deriving transport policy implications from hybrid choice models. *Transport Policy, 36*(0), 217-222.

Combes, P.-P., and Lafourcade, M. (2005). Transport costs: measures, determinants, and regional policy implications for France. *Journal of Economic Geography, 5*(3), 319-349.

De Keizer, B., Geurs, K. T., and Haarsman, G. H. (2012). *Interchanges in timetable design of railways: A closer look at customer resistance to interchange between trains*. Paper presented at the European Transport Conference, Glasgow, 8-10 October 2012.

Debrezion, G., Pels, E., and Rietveld, P. (2007). The Impact of Railway Stations on Residential and Commercial Property Value: A Meta-analysis. *Journal of Real Estate and Finance Economics, 35*, 161-180.

Espino, R., Román, C., and De Ortúzar, J. (2006). Analysing Demand for Suburban Trips: A Mixed RP/SP Model with Latent Variables and Interaction Effects. *Transportation, 33*(3), 241-261.

Fietsersbond. (2011, December 2013). Fietsersbond http://www.fietsersbond.nl/fietsersbond-routeplanner

Geurs, K., Zondag, B., De Jong, G., and Bok, M. d. (2010). Accessibility appraisal of integrated land-use/transport policy strategies: more than just adding up travel time savings. *Transportation Research Part D, 15*, 382-393.

Geurs, K. T., and Van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport geography, 12*, 127-140.

Glerum, A., Atasoy, B., and Bierlaire, M. (2014). Using semi-open questions to integrate perceptions in choice models. *Journal of Choice Modelling, 10*(0), 11-33.

Gutiérrez, J. (2001). Location, economic potential and daily accessibility: An analysis of the accessibility impact of the high-speed line Madrid-Barcelona-French border. *Journal of Transport Geography, 9*(4), 229-242.

Gutiérrez, J., Cardozo, O. D., and García-Palomares, J. C. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography, 19*(6), 1081-1092.

Hadas, Y., and Ranjitkar, P. (2012). Modeling public-transit connectivity with spatial quality-of-transfer measurements. *Journal of Transport Geography, 22*(0), 137-147.

Handy, S. (1993). Regional versus local accessibility: implications for non-work travel. *Transportation Research Record, 1400*(Journal Article), 58-66.

Handy, S. L., and Niemeier, D. A. (1997). Measuring accessibility: an exploration of issues and alternatives. *Environment and Planning A, 29*, 1175-1194.

Hull, A., Silva, C., and Bertolini, L. (2012). Accessibility Instruments for Planning Practice. In C. Office (Ed.).

Koopmans, C., Groot, W., Warffemius, P., Annema, J. A., and Hoogendoorn-Lanser, S. (2013). Measuring generalised transport costs as an indicator of accessibility changes over time. *Transport Policy, 29*(0), 154-159.

La Paix, L., and Geurs, K. (2015). Modelling observed and unobserved factors in cycling to railway stations: application to transit-oriented-developments in the Netherlands. *European Journal of Transport and Infrastructure Research, 15*(1), 27-50.

Lei, T. L., and Church, R. L. (2010). Mapping transit-based access: integrating GIS, routes and schedules. *International Journal of Geographical Information Science, 24*(2), 283-304.

Linneker, B. J., and Spence, N. A. (1992). Accessibility measures compared in an analysis of the impact of the M25 London Orbital Motorway on Britain. *Environment and Planning A, 24*, 1137-1154.

LISA. (2012). Lisa Foundation (Dutch employment database). Enschede, the Netherlands. http://lisa.nl/home

Martens, K. (2004). The bicycle as a feedering mode: experiences from three European countries. *Transportation Research Part D: Transport and Environment, 9*(4), 281-294.

Mavoa, S., Witten, K., McCreanor, T., and O'Sullivan, D. (2012). GIS based destination accessibility via public transit and walking in Auckland, New Zealand. *Journal of Transport Geography, 20*(1), 15-22.

Monteiro, F. B., and Campos, V. B. G. (2012). A Proposal of Indicators for Evaluation of the Urban Space for Pedestrians and Cyclists in Access to Mass Transit Station. *Procedia - Social and Behavioral Sciences, 54*(0), 637-645.

Nichols, A. J. (1975). Standard generalized cost parameters for modelling interurban traffic and evaluating interurban road schemes. Note 255: Mathematical Advisory Unit, Department of the Environment, London.

Páez, A., Scott, D. M., and Morency, C. (2012). Measuring accessibility: positive and normative implementations of various accessibility indicators. *Journal of Transport Geography, 25*(0), 141-153.

Paulssen, M., Temme, D., Vij, A., and Walker, J. (2014). Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. *Transportation, 41*(4), 873-888.

Train, K. (2003). *Discrete choice methods with simulation*. New York: Cambridge University Press.

Van Hagen, M. (2011). *Waiting experience at train stations. Phd Thesis.* Phd thesis, University of Twente, the Netherlands.

van Wee, B., Banister, D., and Annema, J. A. (2013). The Transport System and Transport Policy: An Introduction. In J. A. A. B. van Wee, D. Banister (Ed.), (pp. 227-253). Cheltenham, Northampton: Edward Elgar.

Walker, J., and Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical Social Sciences, 43*(3), 303-343.

Wang, Y., Monzon, A., and Ciommo, F. D. (2015). Assessing the accessibility impact of transport policy by a land-use and transport interaction model – The case of Madrid. *Computers, Environment and Urban Systems, 49*(0), 126-135.

Zofío, J. L., Condeço-Melhorado, A. M., Maroto-Sánchez, A., and Gutiérrez, J. (2014). Generalized transport costs and index numbers: A geographical analysis of economic and infrastructure fundamentals. *Transportation Research Part A: Policy and Practice, 67*(0), 141-157.

Zondag, B., de Bok, M., Geurs, K. T., and Molenwijk, E. (2015). Accessibility modeling and evaluation: The TIGRIS XL land-use and transport interaction model for the Netherlands. *Computers, Environment and Urban Systems, 49*(0), 115-125.