Enhanced Huff Model for Estimating Park and Ride (PnR) Catchment Areas in Perth, WA

Ting (Grace) Lin[a], Jianhong (Cecilia) Xia[a], Todd P. Robinson[a],

Doina Olaru[b], Brett Smith[b], John Taplin[b], Buyang Cao[c]

[a]Department of Spatial Sciences, Curtin University, Kent St Bentley, Western Australia

[b]Business School, University of Western Australia, WA, Australia

[c]School of Software Engineering, Tongji University, Shanghai, China

Corresponding author: Ting (Grace) Lin

Phone Number: +61 0420990820

Email: zju.grace@gmail.com

Abstract:
A train station catchment area delineates the spatial territory from which the users of a train station are drawn. The size and shape of this catchment can be influenced by a variety of factors, such as the transport network, the location of stations and the service quality they offer, as well as the land use density and diversity in the transport corridor. Although numerous studies have been conducted to understand the size of catchment areas, limited research has focused on determining the spatial boundary (shape) of train station catchments. This paper develops a framework for deriving a spatial boundary of a Park and Ride (PnR) catchment area by incorporating the Huff model and Geographic Information Systems (GIS) technologies. The approach is staged, firstly determining the PnR station choice as a function of the attractiveness of a train station and the cost of access between the origin (such as a suburb) and the destination of a trip (such as the Perth CBD). Linear referencing method is then applied to re-define the origins to train stations based on the derived station choice probability. Finally, the spatial boundary of a catchment area is determined according to the adjusted origins, using GIS technologies. The model outputs were evaluated against licence plate survey of station users, where the Kappa coefficient (0.74) and overall accuracy (0.88) statistic suggested that the model's results are robust. The paper then shows how catchment area data can be used to better manage travel demand and plan design solutions aimed at increased accessibility to train stations.

Keywords: urban rail transport, station choice, modified Huff model, linear referencing, station attractiveness

1. Introduction

A train station's catchment area refers to the areal extent from which the majority of users will typically be found (Dolega et al., 2016) . It is a prerequisite for the calculation of several fundamental statistics including latent demand (potential customers) (Banister, 1980), market share (the portion of a market) (Lee and Masao, 1988, p. 17-19) and accessibility (ability to reach) (El-Geneidy and Levinson, 2006). Catchment area estimation methods are numerous, range greatly in sophistication, and the choice between them largely depends on the complexity of the competitive forces involved, along with their computational complexity and data availability.

Proximity-only models include buffer rings and polygons depicting drive time along a network (i.e. service areas) from a point of interest (e.g. convenience store). Buffer rings are perhaps the simplest method to calculate, but assume distance from origin to destination is Euclidean and omnidirectional, whereas catchment areas can have a diverse shape affected by road alignments, natural features, property developments, zoning, parking capacity, location of train station, and surrounding land use (Cervero et al., 1995; Debrezion et al., 2009; Sanko and Shoji, 2009). Furthermore, these approaches make generalised, and strictly binary, decisions about maximum buffer distance (Upchurch et al., 2004) . For example, 800 m has been broadly accepted as a reasonable walking distance to a train station (Cervero, 2001; Cervero et al., 1995; El-Geneidy et al., 2010; Zhao et al., 2013). However, this distance varies spatially, with people living in the suburbs likely to accept larger distances than people living in the CBD (O'Sullivan and Morrall, 1996), for example. Service area polygons are a more realistic way of delineating the catchment area and are valid where patrons are expected to use the closest facility (Dolega et al., 2016; Landex and Hansen, 2009). However, like buffer rings, they can be poor predictors of catchment area where proximity is not the only consideration for selecting a particular service.

Proximity to residence is not necessarily the only factor for choosing a train station, as factors such as service quality, facilities available at station, total travel time, access time, service frequency, generalised cost, access mode, road congestion, network connectivity, parking search time, carriage crowding, and demographics all play a key role in station choice (Chen et al., 2014a; Chen et al., 2014b; Kastrenakes, 1988; Lin et al., 2014; Olaru et al., 2014; Ryan et al., 2016; Shao et al., 2015). For example, some users may choose a station nearer to their final destination, in order to save travel costs, while others may choose a station further away from their destination to secure a seat and improve the comfort of their travel, and get access to convenient parking. A study conducted by Debrezion et al. (2007) found that less than half (only 47%) of the passengers in a Dutch railway survey chose their nearest train station. Whilst this may be an extreme example, it does serve to illustrate that the size of catchment area can depend upon the interaction of travellers with facilities and services at stations. This cannot be well depicted with proximity-only models. Another concept uses the convex hull of geocoded trip data (origin to destination) after removal of outliers (Durr et al., 2010) which could present the real trip characteristics, but does require substantial sample sizes (both spatially and temporally) to be truly representative. In such cases, gravity models may be more appropriate as they include not only distance but attractiveness in their computation.

Hence some new perspectives are needed to define train station catchment areas that can incorporate the plethora of reasons affecting decisions made by travellers. In this study, we apply the Huff Model to define train station boundaries using the suburban railway service in Perth, Western Australia as a test case. The Huff model is a probabilistic retail gravity model originally used to predict consumer behaviour among competing retail stores (Huff, 1963). Its major advantage over proximity-only models and more simplistic retail models (e.g. Reilly (1931)) is the ability to simultaneously estimate a customer's patronage probability for many centres (e.g. retail locations) at once (Joseph and Kuby, 2011). Whilst originally developed for retail, the Huff model has been applied to many other areas including accessibility to health care (Luo, 2014) and healthy food (Kuai, 2015), and for choice based analysis like university campus or movie theatre to attend (Bruno and Improta, 2008; De Beule et al., 2014; Nakanishi and Cooper, 1974).

The aim of the paper is to develop a methodology for deriving the spatial boundary of the PnR catchment area of train stations, by incorporating the Huff model and Geographic Information Systems (GIS) technologies. Four objectives for achieving this aim are: a) adapt the Huff model by including additional factors that affect train station choice using the Multiple Criteria Decision Analysis (MCDA); b) determine the probabilities for PnR commuters to choose a parking station from the nearest three train stations to their origins; c) derive the spatial boundary of the PnR catchment area of train stations; d) validate the model with observed license plate survey of actual station PnR users.

The paper is structured as follows: Section 2 states the material used in the paper. Section 3 focuses on the framework and methodology of estimating catchment areas. The results are explained based on a case study of Perth, Western Australia in Section 4. Section 5 evaluates the results by two different methods and Section 6 exemplifies the methodology using two scenarios. The paper ends with a summary of findings and contributions, and a discussion of limitations and possible further developments.

2. Materials

2.1 Study area

Perth has 70 train stations on 173 kilometres of track (Australian Department of Infrastructure and Transport, 2014). Figure 1 shows the overall design of Perth's rail network, which includes three long established lines (Midland, Fremantle and Armadale lines built before the 1900s) and two new lines crossing the city from the North to the South (Joondalup, 1992 and Mandurah, 2007 lines) (PTA, 2009). The radial system starts from Perth's CBD, which is the largest destination of the rail trips, and most railway trips in the morning peak hour are in-bound trips. There are significant differences among the train lines, stations and their surrounding land use, mainly because of their place in the urban development of Perth. (see Figure 1).

Perth (Western Australia (WA)) is a low-density city (310 people per km^2 for the Great Perth) with high car ownership (around 723 vehicles per 1, 000 people) (Australian Bureau of Statistics, 2012-2013; Curtis, 2008). This has made the delivery of a high-frequency public transport system a major challenge. The PnR system is widely recognised as having a positive influence on public transport demand in a low density city (Olaru et al., 2013). Perth has developed a 17,203 parking bays (including 270 short-term parking bays and 16,983 long-term parking bays) by 2010 and plan to add additional 17,000 by 2021 as part of its $3.8billion METRONET plan (McGowan, 2013).

Figure 1: Perth railway network and location of intercept surveys (Lin et al., 2014)

(Surveys were conducted at the seven labelled stations)

2.2 Data collection

This research used multiple sources of primary and secondary data. Primary data refer to data that observed and collected from first-hand experience whist secondary data refer to the data that previously gathered by someone else for some other purposes (Stevens, 2006, p. 90). Table 1 summarises the data collected by this study. In order to understand commuters' station choice behaviour, intercept surveys were conducted to collect travel data of all public transport riders (including all travel modes) and their satisfaction with train services and facilities. We randomly chose train users at station platforms and asked them to fill in the questionnaire designed for our study. The intercept surveys were conducted on all five rail lines, at seven stations (Figure 1) on two occasions: 31 July - 1 August 2012 (between 6:00AM and 4:00PM) and 19 - 20 September 2013 (between 7:00AM and 12:30PM). A total of 1,263 responses were collected. The purpose of the intercept surveys is to understand the commuters' travel behaviour. For this study, we used the PnR component of the survey. Prior to that, a PnR facilities survey was conducted in April 2012 to understand train station service quality. Twenty-seven types of facilities (12 categories of facilities within and around the train stations) were audited at all train stations in the Perth Metropolitan Area. Finally, the licence plate survey data and the public transport timetable information were obtained from the state government agencies (e.g. Public Transport Authority, PTA, Department of Planning, DoP, and Department of Transport, DoT). The licence plate survey provided the home address for the PnR users at the train stations, based on their number plate and vehicle registration information. The home location (randomly shifted within a 50 m buffer in order to protect individual privacy) was then geocoded and mapped. Although this procedure aimed to ensure anonymity it did introduced some locational errors, though they are were deemed small enough and within the range of confidence for model validation.

Information from the intercept surveys indicates that work and education represent the dominant trip purposes for the morning peak travel (over 80%) and one-third of commuters use PnR (32.65%). In addition, results show that over 70% of the PnR travellers accessed stations at distances less than 8 km, which corresponds to an average of three stations (see cumulative function Figure 2).

Table 1 Data collection summary table

	Primary data source			
	Name	**Num Stations**	**Num Samples**	**Time period**
1	**Intercept survey 1**	7	940	31/07/ 2012- 1/8/2012 (6:00AM - 4:00PM)
2	**Intercept survey 2**	7	323	19-20/ 09/ 2013 (7:00AM-12:30PM)
3	**Facilities survey**	69[1]		April 2012
4	**Factors importance ranking survey**[2]		17	December 2013

	Secondary data source			
	Name	**Num Stations**	**Time period**	**Source**
1	**Network data**			Provided by DoP
2	**Walk Score**	69		Work Score (https://www.walkscore.com/)
3	**Licence plate survey**	22	2006-2008	Provided by DPI
4	**TransPerth timetable**	69		PTA (2015)
5	**Car park full time survey**		2014	Parliament of WA (2014)
6	**Statistical boundaries**		2011	Australian Bureau of Statistics (http://www.abs.gov.au)
7	**Journey to work**		2011	Australian Bureau of Statistics (2011)

[1] A new station (Bulter) opened in Perth, 2014

[2] 17 policy makers from government agencies

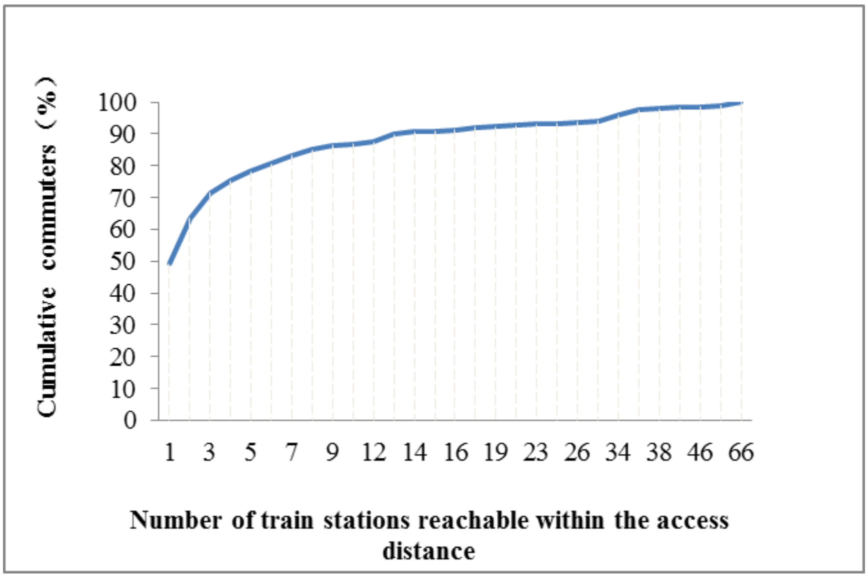


Figure 2: Cumulative distance function

3. Methods

Starting from the Huff model, this section describes the steps required to model the catchment area of railway stations. The first step applied a modified Huff model to determine the probability of a station was chosen by PnR users. Linear referencing was used to calibrate the points of the origins for trips. Finally, the spatial boundary of a station catchment area was delineated according to adjusted points using ArcGIS™ software. Perth, Western Australia, was selected as case study given the prominence of PnR in the public transport mode share and the local knowledge of the researchers. For simplicity, only morning commuting trips to CBD were analysed, as they represent more than 60% of the total trips done in the morning peak.

3.1 Modified Huff model

The original Huff model was developed by Huff in 1963 (Huff, 1963) for understanding the popularity of shopping centres based on a spatial interaction theory. It has endured for more than 50 years and has been widely used by business analysts and academicians all over the world (Huff and McCallum, 2008). For this study, the original Huff model was modified for application to choice of train station as follows through provided equation (2) and (3) regarding to how to define A_j and T_{ij}:

$$P_{ij} = \frac{\frac{Aj}{T_{ij}^{\lambda}}}{\sum_{j=1}^{n} \frac{Aj}{T_{ij}^{\lambda}}} \tag{1}$$

$$A_j = \omega_1 F_1 + \omega_2 F_2 + \ldots + \omega_l F_l \tag{2}$$

$$T_{ij} = TOT_{ij} + TTD_j \tag{3}$$

where:

P_{ij} is the probability of travelling from origin suburb *i* to Perth CBD, through train station *j*;

T_{ij} is network based travel time from origin suburb *i* to Perth CBD through train station *j*;

λ is a distance decay exponent, indicating the effect of travel time on station choice (here λ =2);

A_j is the attractiveness of train station *j*;

F_l is the factor *l* that contributes to the train station's attractiveness, such as parking availability index or land use diversity index;

W_l is the weight of the factor *l* that contributes to the train station's attractiveness;

TOT_{ij} is network based travel time from origin suburb *i* to train station *j* (access time)*;* and

TTD_j is travel time from train station *j* to Perth CBD (here it means in-vehicle time).

In the most recent form of Huff model, the attractiveness was measured in a multiplicative form and the weight or parameter for the sensitivity of a choice associated with a factor was estimated and calibrated statistically using the actual shopping preference survey data (Huff and McCallum, 2008). In our study, rather than the multiplicative form, we adopted the additive form to derive weights by conducting an extra survey for understanding policymakers' opinions on the importance of train station choice factors. For the detailed information, see section 3.2.

Dolega et al. (2016) reported the distance decay parameter usually takes a value of between −1 and −2, depending on factors such as the types of retail centres or competition between centres; Dramowicz (2005) also noted the distance decay parameter as a value of 2. In transport, it is reported that 2 is usually used for the distance decay of a power function (Cambridge Systematics et al., 2012, p. 44). Through modelling the relationship between distance and the percentage of PnR trips within a *1* km buffer ring using a power function in the Matlab™, the distance decay parameter was estimated to be about -1.87 (R^2 = 0.67) (see figure 3). Based on literature and our model, two, therefore, was adopted for the modified Huff model. The robustness of model was thoroughly evaluated in Section 5.

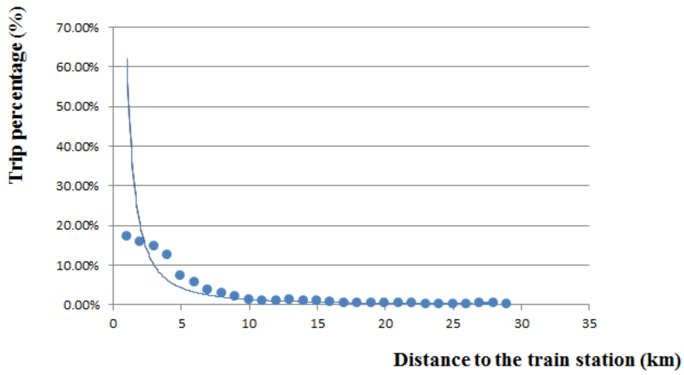


Figure 3: Distance decay of commuters' station choice

Because Perth's CBD is the largest employment centre and the largest destination in Perth, we simplified the analysis by considering only trips to the city. Therefore, the travel time includes access time from home to a station and travel time from the station to Perth's CBD. While the travel time from home to a station is populated through network analysis in ArcGIS™, the travel time from station to Perth's CBD is directly extracted from Transperth timetables (PTA, 2015). As indicated, given the low density and longer travel distances in Perth, the choice set is reduced from 70 stations to the closest three stations to home location. Therefore, instead of calculating the probability of accessing any of the 70 stations, the nearest three stations to the centroid of a suburb were considered as candidate stations.

3.2 Attractiveness of a train station

The attractiveness of station can be determined using both a multiplicative form (Huff, 2003) and an additive form (Haimes and Steuer, 2012, p. 333). This study adopted the additive form based on the MCDA model in order to incorporate the experts' opinions on the importance of factors affecting station choices. The attractiveness of a train station was measured using four indices:

- Parking capacity (the number of available parking bays at the train stations);
- Street parking availability (dummy variable, indicating whether street parking is available around a station *1* or not *0*);
- Land use diversity index, and
- Service and facility quality index.

This research adopted the Walk Score for assessing land use diversity (Leslie et al., 2007), as it represents a good proxy for land use mix. Walk Score was calculated based on "distance to 13 categories of amenities (e.g., grocery stores, coffee shops, restaurants, schools, parks, libraries); and each category was weighted equally and summarized scores were then normalized to yield a score of 0-100" (Carr et al., 2010). Finally, the train station service and facility quality index include two components: facilities and frequency of services. Frequency was measured by the average number of trains serving the train station on a working day (using the Transperth timetables). The facilities index was calculated as a weighted sum of 12 facilities and its components is shown in Figure 4.

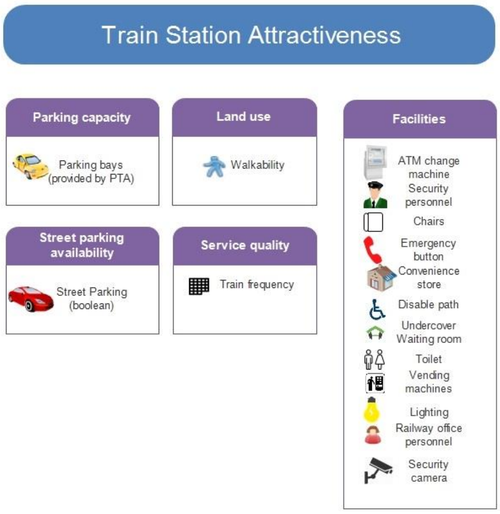


Figure 4: Components of a Train Station's attractiveness

As these factors are measured in different units, in order to combine them into one attractiveness index, the factors were "standardised" using the score range (benefit criteria) method (Malczewski, 1999):

$$X'_{ij} = \frac{X_{ij} - X_j^{\min}}{X_j^{\max} - X_j^{\min}} \tag{4}$$

where:

X'_{ij} is the normalised value for item *i* in *jth* attribute;

$X_j^{\min}$ is the minimum score for the j^{th} attribute;

$X_j^{\max}$ is the maximum score for the j^{th} attribute; and

$X_j^{\max} - X_j^{\min}$ is the range of a given criteria.

Then, the overall attractiveness of a train station was calculated according to equation (2). These weights were determined through ranking the importance of factors from 1-7, 7 is the most important by policy makers. Seventeen officers from government agencies (such as DoP, PTA and DoT) were interviewed. The average of ranked values of each factor was calculated and rescaled into weights using a comparison weighting matrix. These weights, therefore, can be added up to one for deriving the attractiveness of a train station using the MCDA model (See Table 2). For services and facilities index (SQI), there are two main components (frequency and facilities). However, we didn't separate them in the questionnaire. Based on the research from Chen et al. (2014b), the frequency is twice more important than facilities. The weight of SQI was re-distributed.

Table 2: The weights of factors that contribute to attractiveness of a train station

Factor	***Weights***		
Parking capacity	0.29		
Street parking availability	0.24		
Land use diversity index	0.23		
Services and facilities quality index	0.24	Facilities	0.16
		Frequency	0.08

3.3 Linear referencing and origin calibration for deriving spatial boundary of catchment area

The purpose of linear referencing and origin calibration is to define the spatial boundary of the catchment area of a train station. The modified Huff model outputs the probabilities of a station being chosen from a particular location, such as a centroid of a suburb (Figure 5). One suburb can then be allocated to three train stations with different probabilities (section 3.1). Once these probabilities are calculated, the next step is to determine the spatial boundary of the catchment area for each train station. In order to make a fair allocation, the centroid of the suburb is relocated using the linear referencing method. The underlying principle is that the probability of a station being chosen is inversely proportional to the distance between a suburb and a station. If the probability of a station being chosen is lower, the centroid of a suburb will be moved away from its original location and get closer to the station. The lower the probability P_{ij}, the more the adjustment of the centroid of suburb i and the shorter the distance D'_{ij}. We call this process linear referencing and origin calibration. It can be formalised in the following.

$$D'_{ij} = D_{ij} * \frac{P_{ij}}{P_i^{highest}} \quad (5)$$

where:

D'_{ij} is the adjusted distance from the centroid of a suburb (origin) i to station j which will determine the calibrated origin;

D_{ij} is the distance from the centroid of a suburb (origin) i to station j;

P_{ij} is the probability of choosing station j from the centroid of a suburb (origin) i to Perth CBD; and

$P_i^{highest}$ is the highest probability of a station being chosen from the centroid of a suburb (origin) i to Perth CBD.

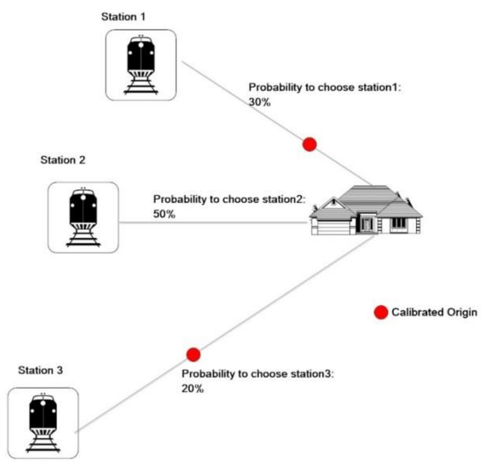


Figure 5: A diagrammatic sketch of calibrating origins for the nearest three train stations from the centroid of a suburb

The spatial boundary of the catchment area of a station was determined using the linear referencing method. As each calibrated origin point represents a suburb, the spatial boundary of a train station was drawn by selecting the intersected suburbs of a station and dissolving or aggregating the boundary of selected suburb's polygons into one area of the station using the ArcGIS™ software. Figure 6 illustrates the process of how the boundaries were drawn using Model Builder™ in the ArcGIS™.

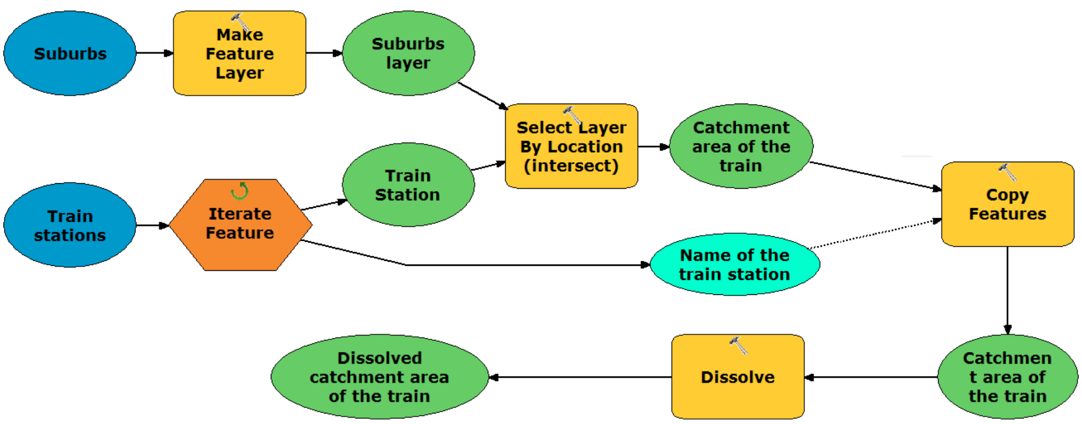
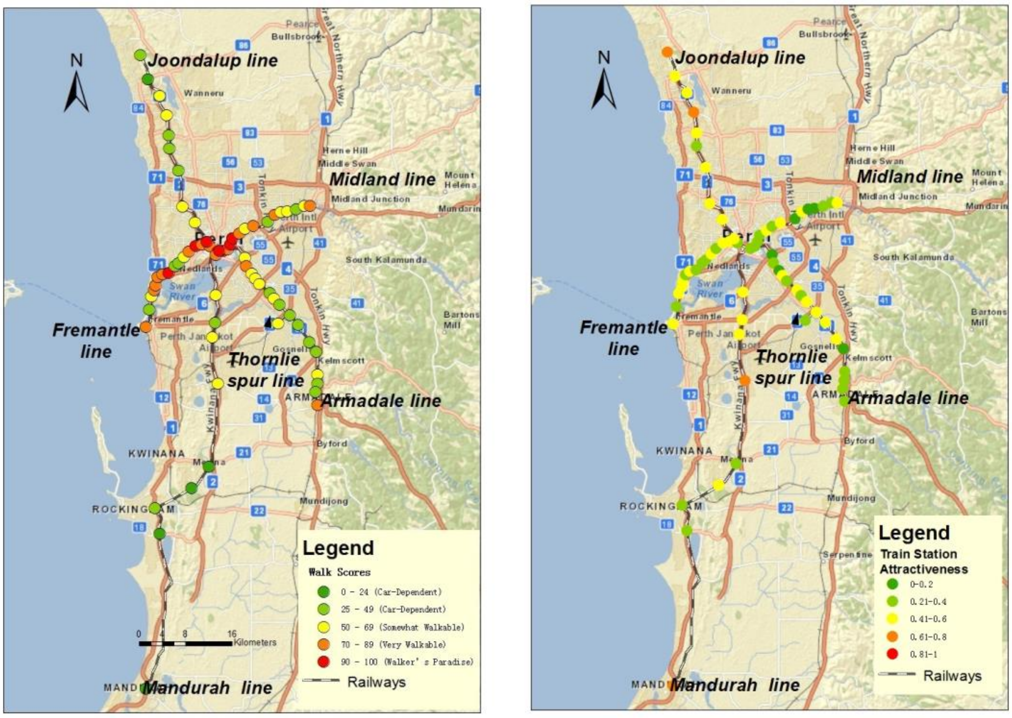


Figure 6: The process drawing a boundary in ArcGIS

4. Results

4.1 The attractiveness of train stations

Figure 7a shows the Walk Score, which represents the land use diversity around the train stations. From the map, it is seen that Perth CBD has higher walkability. Perth and Esplanade stations get the highest values; and the further from the CBD, the lower the Walk Score. Among the train lines, Fremantle and Midland lines receive higher Walk Scores than other train lines as the surrounding areas are well-developed along these two train lines. Although Perth City got higher Walk Score, the final attractiveness was not necessarily the highest when all factors (Figure 7b) were taken into account. It is mainly due to the limited parking capacity in the CBD area.



a) The map of Walk Score

b) The map of train stations' attractiveness

Figure 7: The map of Walk Scores of train stations and train station's attractiveness

4.2 Origin calibration

To illustrate how the method was applied, Figure 8 provides an example of output from the modified Huff model (Alexander Heights is the suburb name). The nearest three stations to Alexander Heights are Warwick, Greenwood and Whitfords stations (see Figure 8). The probabilities of these three stations being chosen from the suburb are 0.41, 0.31 and 0.28. Warwick station has the highest probability; therefore, the centroid of Alexander Heights will remain unchanged on the line to the Warwick station. However, the centroid of the suburb will move towards the Greenwood and Whitfords stations by 1,881.55 m and 2,585.67 m respectively.

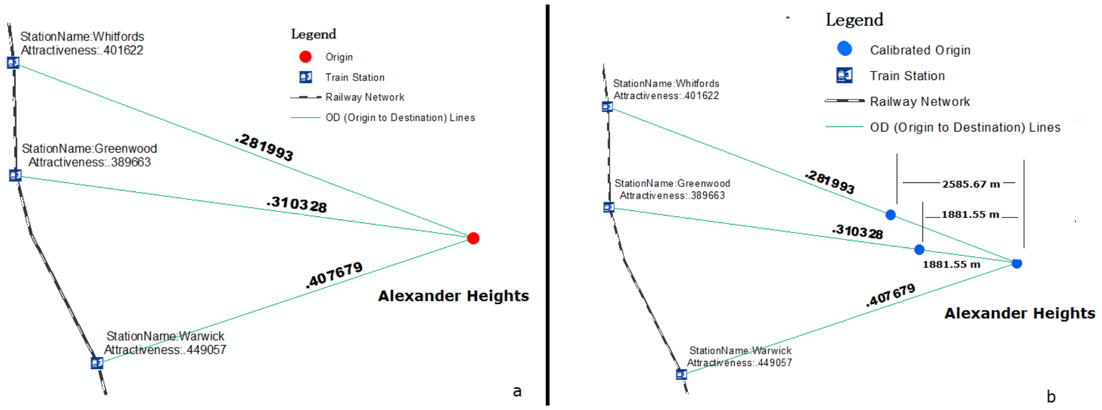


Figure 8: The outputs from the Modified Huff Model and origin calibration

5. Evaluation of the methodology

Two approaches were applied to assess the performance of the methodology: direct comparison and using the Kappa test, described in sections 5.1. Both of them used license plate survey data (section 2.2). The approximate home locations of PnR users as observed data served for validating the accuracy of derived spatial boundary of the catchment area of a station. The license plate survey is a good source for understanding the train station catchment areas (See Figure 9). The yellow points represent the approximate locations of the PnR users' travel origin. Buffer rings of 1, 3, 5, and 10 km were drawn around the train stations to illustrate the size of their catchment areas.

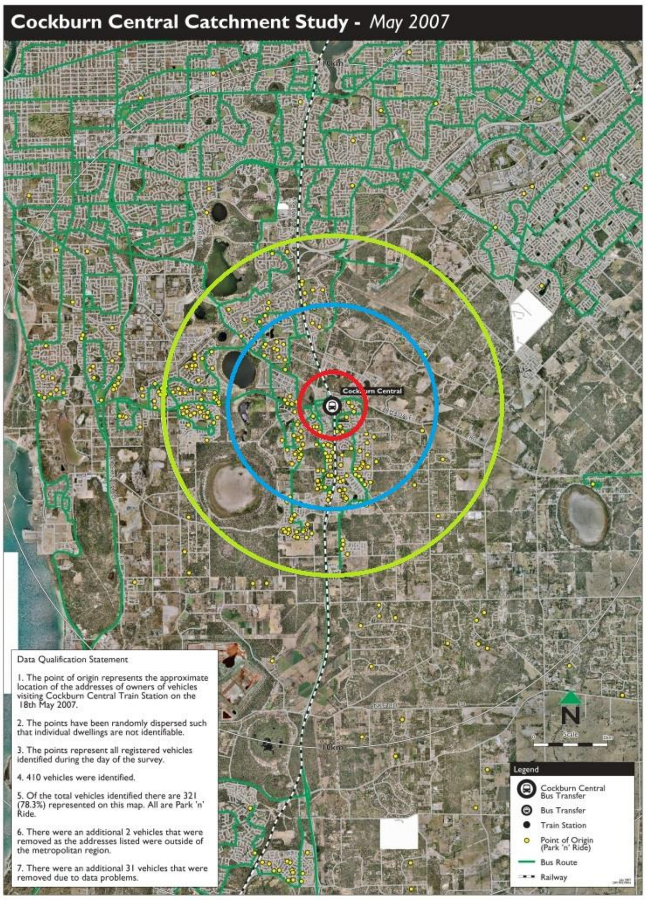


Figure 9:Cockburn central catchment area study based on 2007 plate survey data (DPI, 2007)

5.1 Direct evaluation

Direct evaluation was conducted by overlaying the catchment area derived from our method over the observed PnR users' origins on a map and calculating the percentage of PnR users within the catchment area boundary. Table 3 shows the direct evaluation results for 22 train stations where the license plate survey was conducted. The overall accuracy of the model is satisfactory by capturing around 73% of patronage, given the nearest three train stations considered in the methodology. Maylands, Cannington and Claremonts stations, which are on heritage train lines, have lower performance which is probably due to a combination of the small spacing between stations on these lines and the land use diversity around those stations, attracting for commuters from beyond the three nearest stations to come. For example, Cannington station remains 51.4% driving longer distances to board the train at the Cannington station. It is also consistent with the results in Shao's research that it is only 26.9% commuters in Cannington station chose Cannington station because it is the nearest station, which means 73.1% people didn't choose their nearest station to their origin instead of driving a longer distance to use Cannington station (Shao et al., 2015).

Table 3: Evaluation table using plate survey provided by PTA

Station Name	***Percentage[1] (%)***	***Station Name***	***Percentage[1] (%)***	***Station Name***	***Percentage[1] (%)***
Cannington	48.59	Maylands	58.18	Bassendean	60.55
Thornlie	66.73/70.89[2]	Meltham	59.62	Midland	60.55
Armadale	76.34	Bayswater	60.83	Claremont	36.00
Fremantle	89.66	Stirling	62.45	Glendalough	74.92
Warwick	87.22	Greenwood	86.34	Whitfords	76.74
Edgewater	89.66	Currambine	85.38	Clarkson	85.64
Bull Creek	72.90	Murdoch	75.75	Cockburn Central	79.15
Mandurah	87.99			**Average**	**72.85**

1 Percentage of survey commuters covered by catchment area generation algorithm

2 PTA conducted two number plate surveys at the station

5.2 Kappa statistic test

Although the direct evaluation provided a simple way to assess the performance of the model, it only counts points of origins inside the catchment area. Kappa statistic test can evaluate performance by thoroughly considering origin locations both within and outside the boundaries of the station catchment areas – LOFI (little out from inside) and LIFO (little in from outside) (Huff and McCallum, 2008).

Kappa test, introduced by Cohen (1960), is the most commonly used index for analysing agreement on a binary outcome between two observers or two classification methods (McLsaac and Cook, 2014). It is frequently used to test reliability.

Figure 10 illustrates how Kappa coefficient was calculated. The colours denote three different train stations. The circles indicate the modelled catchment areas and the points indicate the origins of travel (observed data). To conduct the Kappa test for the blue station, we need to also include the data from adjacent stations (yellow and purple stations). According to the Kappa test, the records are grouped into four categories depending on the agreement between catchment areas and the vehicle registration plate survey data: observed presence and modelled presence (PoPm), observed absence and modelled absence (AoAm), observed presence and modelled absence (PoAm), observed absence and modelled presence (AoPm). PoPm counts all the observed license plate points inside the modelled catchment area (see four blue points inside the blue circle on Figure 9; AoAm counts for all the observed points outside the modelled catchment area (six yellow and purple points outside the blue circle on Figure 9). AoPm counts the yellow and purple dots inside the blue circle, but outside their own colour circles. These locations are important because the catchment areas of the train stations can substantially overlap. Finally, PoAm counts for all blue points outside the blue circle. AoPm and PoAm indicate the errors of the model. Then the Kappa coefficient and the accuracy of the model could be determined as (Viera and Garrett, 2005) :

$$m_1 = P_o P_m + A_o P_m \tag{6}$$

$$m_0 = P_o A_m + A_o A_m \tag{7}$$

$$n_1 = P_o P_m + P_o A_m \tag{8}$$

$$n_0 = A_o P_m + A_o A_m \tag{9}$$

$$n = P_o P_m + A_o A_m + A_o P_m + P_o A_m \tag{10}$$

$$AG_o = \frac{P_o P_m + A_o A_m}{n} \tag{11}$$

$$AG_m = [\frac{n_1}{n} * \frac{m_1}{n}] + [\frac{n_0}{n} * \frac{m_0}{n}] \tag{12}$$

$$K = \frac{AG_o - AG_m}{1 - AG_m} \tag{13}$$

where:

AG_o represents the observed agreement;

AG_m = modelled agreement; and

K = Kappa index;

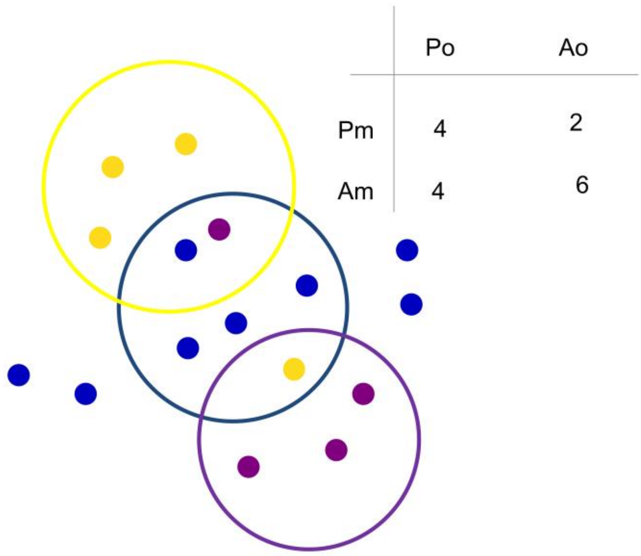


Figure 10: The illustration of Kappa test

Like most correlation statistics, the Kappa can range from −1 to +1. Cohen suggested that Kappa statistic could be interpreted as follows: values ≤ 0 indicate no agreement, 0.01–0.20 show weak/slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 as almost perfect agreement.

As described, Kappa calculation requires the presence of adjacent stations. Although 22 stations have the license plate data, only six stations satisfied the criterion of three consecutive train stations available. The results are shown in Table 4. The results from Kappa index show the overall accuracy of the model is satisfactory and the proportions are higher than in the direct evaluation method. The overall accuracy here refers to the ratio that is correctly modelled. To example shown in the Figure 10, it can be expressed as (PoPm+ AoAm)/(PoPm+ AoAm+ PoAm+ AoPm).

Table 4: Kappa coefficient table

Station Name	***Kappa index***	***Overall Accuracy***	***Station Name***	***Kappa index***	***Overall Accuracy***
Meltham	0.70	0.91	Stirling	0.61	0.80
Warwick	0.86	0.95	Greenwood	0.84	0.92
Whitfords	0.78	0.89	Murdoch	0.62	0.83
Average	**0.74**	**0.88**			

6. Implementation of the methodology and policy implications

Two case studies are presented next to understand the catchment area and the supply-demand relationships for Perth train stations.

6.1 Changes after Mandurah line expansion

Figure 11 and Table 5 show the variation of catchment areas of train stations after Mandurah line expansion. It appears that the operation on the newest rail line to Mandurah had a significant influence on the Fremantle and Armadale lines. The largest catchment areas correspond to the stations located near to the end of the train line, such as Mandurah, Rockingham. At the same time, for Fremantle and Armadale lines, most of the train stations have decreased their catchment areas after Mandurah train line expansion (Table 5). The reason of the big catchment area variations in Table 5 is due to lack of train services in the south/southwest suburbs of Perth before train line expansion and train users live far away from Mandurah train station still need to access train services (Figure 12). Another possible reason is the long station spacing on the Mandurah line. For example, the spacing between Warnbro and Mandurah is around 23.5 km. Figure 10b shows the substantial changes in the catchment area of Fremantle train station before and after Mandurah line expansion. The results are also due to some changes in the suburb boundaries occurred between 2006 and 2011 (the line started operation in December 2007). From the modelling result, the average rate of Fremantle catchment decrease is 126.65% whist for

Armadale line, the average decrease rate is around 90.95% after Mandurah line expansion.

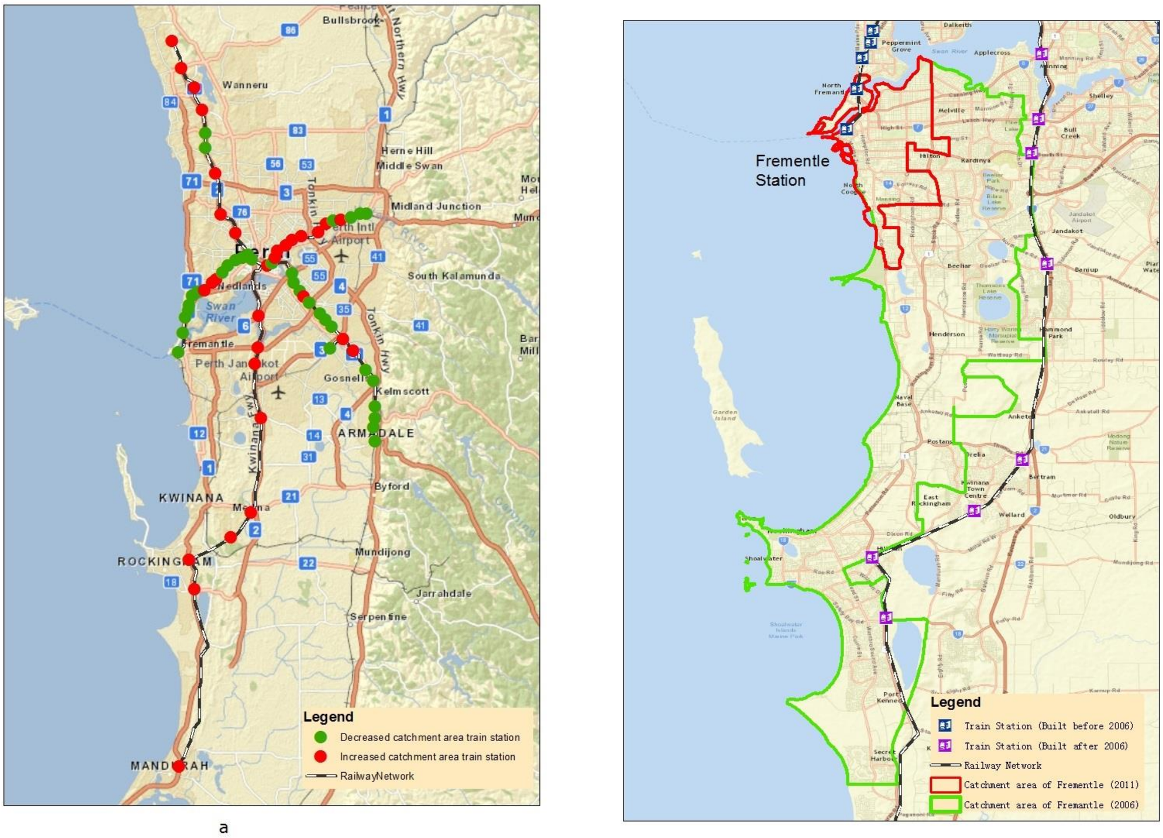


Figure 11: Catchment area variation after Mandurah line expansion

Table 5: Catchment area variation rank table

Station Name	***Catchment Area (km^2)***	***Rank***	***Station Name***	***Catchment Area Variation - Decreased (km^2)***	***Rank***
Mandurah	2,129.99	1	Sherwood	-3,352.54	1
Rockingham	817.98	2	Armadale	-2,261.31	2
Warnbro	622.21	3	Challis	-924.21	3
Wellard	370.95	4	Seaforth	-486.57	4
Kwinana	315.47	5	Fremantle	-222.27	5

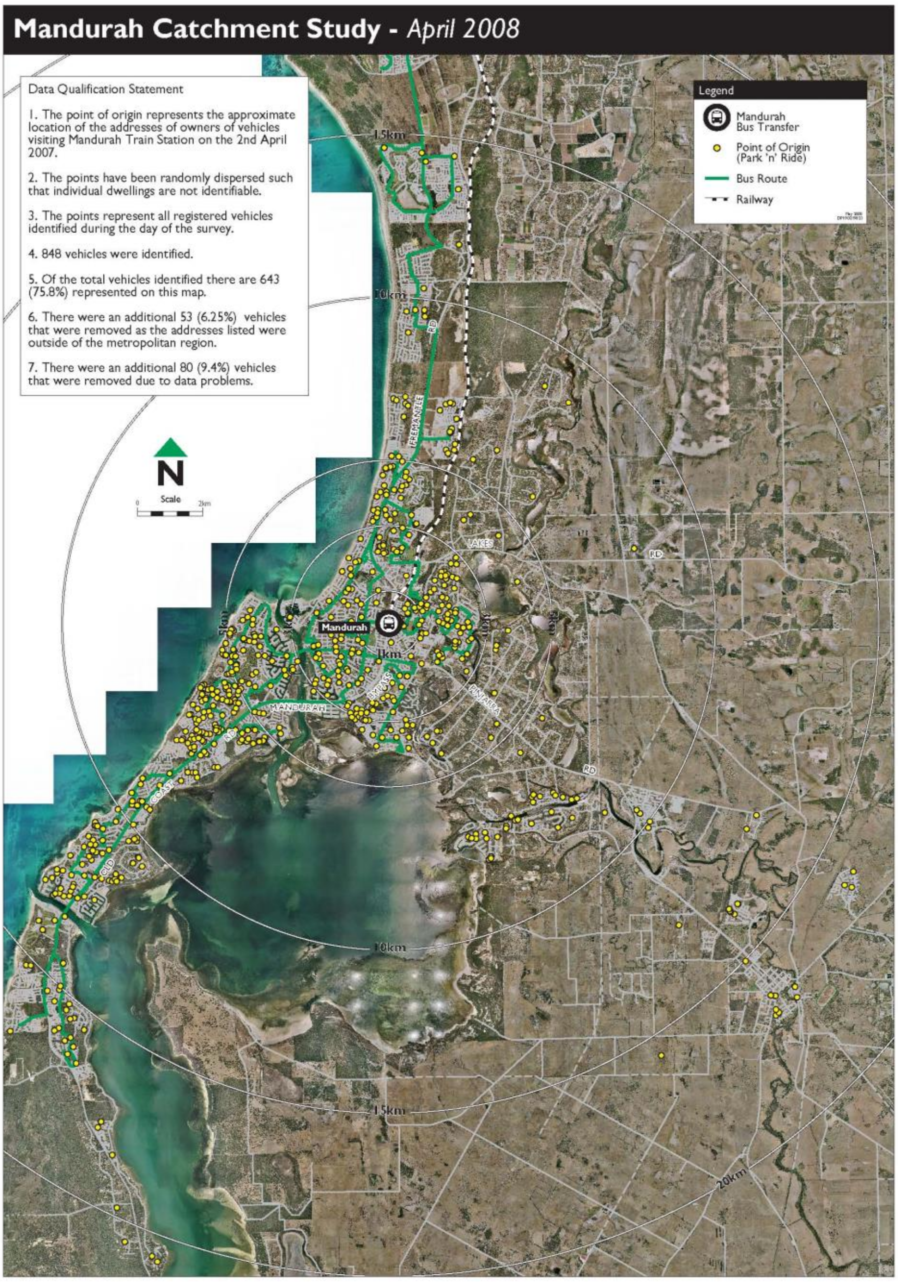


Figure 12 :Mandurah catchment area study based on 2007 plate survey data (DPI, 2007)

6.2 Latent PnR demand and supply

In Perth, Parking supply is an important determinant of the travel mode choice. As it delineates where travellers come from, the estimated catchment area of a train station can be used for estimating the parking demand and thus helping to support parking supply decisions. The survey of car parks found most of them are full before 8:00am, indicating that most train stations have insufficient parking bays (Parliament of WA, 2014). Combining this information with journey to work data from Census 2011 (Australian Bureau of Statistics, 2011), Table 6 shows the estimated parking demand compared to the current parking supply, highlighting stations with unsatisfied parking demand.

Table 6: Estimated PnR demand vs provided parking bays

Station Name	***Estimated PnR Demand***	***Ranking by Demand***	***Long-term bays***	***Current Capacity Status***
Edgewater	1,596	1	887	Full by 7:45 am
Joondalup	1,391	2	225	Full by 5:40 am
Murdoch	1,256	3	1,152	Full by 7:50 am
Greenwood	1,147	4	931	Full by 7:50 am
Whitfords	1,035	5	866	Full by 7:30 am

It is found that the highest PnR demand is at train stations where there is already a large supply of parking bays. Still the supply shortage exists. Most of the train stations identified in table 6 have parking areas full in the early hours of the morning and earlier than at other train stations (according to the car park full time survey conducted in 2014). This is especially the case of the Joondalup train station, where the car park is full before 5:40 am.

7. Concluding remarks

This paper has reported on the development of a modelling tool to forecast the catchment area of a train station, which is useful for understanding the potential travel demand of transit stations, and for subsequent planning of parking space supply. The developed model delineated the catchment area based on the attractiveness of the train station and the distribution of residential origins. This is novel and enriches the existing catchment area measures. By combining an enhanced Huff model with linear referencing modelling, we now better understand the competition between train stations, as well as the role of train station attractiveness on the catchment areas. Using a case study for Perth, Western Australia, the model was tested using direct evaluation and Kappa statistic. The results confirm the robustness of the model.

The method we developed has several advantages and benefits: 1) It is simple to calculate and provides not only the size of a catchment area, but also the spatial boundary (extent) of a catchment area of a transit station. Therefore, it can be easily used to link to other issues of concern to transit policy and planning (Dolega et al., 2016). 2) Using catchment areas can provide better estimates of latent demand for a transit station, and account for competition between stations. 3) The tool is useful for both long-term and short-term planning. For example, catchment area can be used to test various infrastructure scenarios. These include the impact of adding a new station, or even a new train line, on the catchment area of the station or line itself (local effect) or the catchment area of other stations and lines (global effect) (Rietveld, 2010). Therefore, decision-makers can develop more sensitive long term plans for transport supply and demand. 4) This methodology can also be useful for operational and practical purposes such as the effect of improvements to a train station infrastructure or the quality of services offered, or to assess the impact of changing accessibility to train stations (Cervero et al., 1995). 5) By incorporating parking supply, a land use index and a services and facilities quality index the tool developed here can evaluate any change to the overall attractiveness of the station following proposals such as to adding new parking facilities or increasing frequency of train services. 6) Although this tool was developed for understanding the catchment area of PnR users, the method can be transferred to estimate catchment areas based on other modes such as, walking, cycling or buss connections.

We implemented the methodology into two scenarios in Perth: one involving train line expansion and the other latent demand analysis. It is found that the some train line catchment areas experienced a great decrease after the Mandurah line expansion, which it is as expected as the new line improved accessibility from some residential areas. The average decrease in the size of catchment areas of train stations along the Armadale line is about 90.95%, whilst it is about 126.65% for the Fremantle line. The latent demand analysis confirmed the robust nature of the model as the estimated demand is consistent with the levels of car park pressure.

As with any research, there are limitations. Firstly, we didn't calibrate the modified Huff model to determine the distance decay parameter in a traditional manner. Rather a widely accepted value was adopted in the paper. However, we did use reliable benchmark data collected by PTA to validate the accuracy of that estimate, which was found to be 0.88. In future developments of this approach we will calibrate the distance decay parameter systematically in order to understand the impact of spatial variation, temporal variation and heterogeneity (e.g. different transport modes) on catchment estimation. Secondly, although the method used in the research is a popular method for determining the weight of the station attraction factors, it has a subjective element and may be difficult to generalise to other studies. Other methods such as discrete choice modelling can help to understand how various station choice factors contribute to the station preference by various categories of travellers. Thirdly, the accuracy of the model can be improved if more stations are included in the "choice set" and then used in linear referencing. However, this will increase the complexity of the calculation. Fourthly, the modifiable areal unit problem (MAUP), a common problem in the GIS analysis, may lead to various solutions. In this study, the suburb was used as the spatial unit of analysis due to computation complexity, although it is not the smallest available unit. Adopting a smaller unit of analysis (SA1 is the smallest spatial unit currently available in Australia) may be beneficial for catchment area identification, however that approach will require more complex computation. As a sensitivity analysis this research explored the influence of the MAUP, at eight train stations, comparing results by suburb and SA1. The catchment areas did change, but not considerably. Therefore, as a compromise between simplicity and accuracy, the suburb is considered an effective unit for analysis.

Acknowledgement

The authors would like to thank the commuters who willingly devoted their time to participate in the study. We are very grateful to PTA (The Public Transport Authority), DoT (The Department of Transport) and DoP (The Department of Planning) for their ongoing support for the study and ARC Linkage Grant LP110201150. Special thanks are due to the data collection team, Gary McCarney from DoP for his feedback on issues discussed in this paper and Jay Sandhu from Esri for his support. The views expressed in this article are those of the authors, and do not necessarily reflect the views of any organisation.

References

Australian Bureau of Statistics, 2011. Method of Travel to Work (MTWP). Australian Bureau of Statistics. Available: http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/2901.0Chapter41002011 [2014].

Australian Bureau of Statistics, 2012-2013. Regional Population Growth, Australia, 2012-13.

Australian Department of Infrastructure and Transport, 2014. Urban public transport: updated trends.

Banister, D., 1980. Transport mobility and deprivation in inter-urban areas / David Banister. Farnborough, Eng. : Saxon House, Farnborough, Eng.

Bruno, G., Improta, G., 2008. Using gravity models for the evaluation of new university site locations: A case study. Computers & Operations Research 35, 436-444.

Cambridge Systematics, Vanasse Hangen Brustlin, CGallop Corporation, Bhat, C., Shapiro Transportation Consulting, Martin/Alexiou/Bryson, 2012. Travel Demand Forecasting: Parameters and Techniques, Transportation Research Board.

Carr, L.J., Dunsiger, S.I., Marcus, B.H., 2010. Walk Score™ as a global estimate of neighborhood walkability. American Journal of Preventive Medicine 39, 460-463.

Cervero, R., 2001. Walk-and-ride: factors influencing pedestrian access to transit. Journal of Public Transportation 7.

Cervero, R., Round, A., Goldman, T., Wu, K., 1995. Rail access modes and catchment area for the BART System. Transportation Centre, The University of California, USA.

Chen, C., Xia, J.C., Smith, B., Han, R., 2014a. Development of a conceptual framework for modeling train station choice under uncertainty for Park-and-Ride users, Transportation Research Board 93rd Annual Meeting.

Chen, Z., Xia, J., Irawan, B., Caulfied, C., 2014b. Development of location-based services for recommending departure stations to Park and Ride users. Transportation Research Part C: Emerging Technologies 48, 256-268.

Curtis, C., 2008. Evolution of the transit-oriented development model for low-density cities: a case study of Perth's new railway corridor. Planning, Practice & Research 23, 285-302.

De Beule, M., Van den Poel, D., Van de Weghe, N., 2014. An extended Huff-model for robustly benchmarking and predicting retail network performance. Applied Geography 46, 80-89.

Debrezion, G., Pels, E., Rietveld, P., 2007. Choice of deaprture station by railway users. European Transport 37, 78-92.
Debrezion, G., Pels, E., Rietveld, P., 2009. Modelling the joint access mode and railway station choice. Transportation Research Part E: Logistics and Transportation Review 45, 270-283.
Dolega, L., Pavlis, M., Singleton, A., 2016. Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. Journal of Retailing and Consumer Services 28, 78-90.
DPI, 2007. Cockburn central catchment study.
Dramowicz, E., 2005. Retail Trade Area Analysis Using the Huff Model, Directions Magazine.
El-Geneidy, A., Levinson, D., 2006. Access to destinations: Development of accessibility measures.
El-Geneidy, A.M., Tetreault, P., Surprenant-Legault, J., 2010. Pedestrian access to transit: Identifying redundancies and gaps using a variable service area analysis, Transportation Research Board 89th Annual Meeting.
Haimes, Y.Y., Steuer, R.E., 2012. Research and Practice in Multiple Criteria Decision Making: Proceedings of the XIVth International Conference on Multiple Criteria Decision Making (MCDM) Charlottesville, Virginia, USA, June 8–12, 1998. Springer Science & Business Media.
Huff, D., McCallum, B.M., 2008. Calibrating the Huff Model Using ArcGIS Business Analyst. ESRI White Paper.
Huff, D.L., 1963. A probabilistic analysis of shopping center trade areas. Land Economics 39, 81-90.
Huff, D.L., 2003. Parameter Estimation in the Huff Model, ArcUser. Esri, pp. 34-36.
Joseph, L., Kuby, M., 2011. Gravity modeling and its impacts on location analysis, Foundations of location analysis. Springer, pp. 423-443.
Kastrenakes, C.R., 1988. Development of a rail station choice model for NJ transit. Transportation Research Record 1162, 16-21.
Kuai, X., 2015. Examining healthy food accessibility in baton rouge, louisiana using a huff-modified 2SFCA method, The Department of Geography and Anthropology. Louisiana State University, Baton Rouge, Louisiana.
Landex, A., Hansen, S., 2009. GIS-based approaches to catchment area analyses of mass transit, ESRI International User Conference.
Lee, G.C., Masao, N., 1988. Market share analysis: Evaluating competitive marketing effectiveness. Kluwer Academic Publishers.
Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., Hugo, G., 2007. Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. Health & place 13, 111-122.
Lin, T., Xia, J., Robinson, T.P., Goulias, K.G., Church, R.L., Olaru, D., Taplin, J., Han, R., 2014. Spatial analysis of access to and accessibility surrounding train stations: A case study of accessibility for the elderly in Perth, Western Australia. Journal of Transport Geography 39, 111-120.
Luo, J., 2014. Integrating the Huff Model and Floating Catchment Area Methods to Analyze Spatial Access to Healthcare Services. Transactions in GIS 18, 436-448.
Malczewski, J., 1999. GIS and multicriteria decision analysis New York : J. Wiley & Sons, New York.
McGowan, M., 2013. METRONET to provide biggest ever boost in car parking.
Nakanishi, M., Cooper, L.G., 1974. Parameter estimation for a multiplicative competitive interaction model: Least squares approach. Journal of Marketing Research 11, 303-311.
O'Sullivan, S., Morrall, J., 1996. Walking distances to and from light-rail transit stations. Transportation

Research Record: Journal of the Transportation Research Board 1538, 19-26.
Olaru, D., Smith, B., Xia, J., Lin, T., 2014. Travellers' attitudes towards Park-and-Ride (PnR) and choice of PnR station: Evidence from Perth, Western Australia. Procedia - Social and Behavioral Sciences 162, 101-110.
Olaru, D., Smith , B., Xia, J.C., Chen, C., Lin, T.G., Ying, H., Han, R., 2013. "Ticket to ride": factors affecting park-and-ride travel in Perth, WA, the 13th World Conference on Transportation Research, Rio de Janeiro, Brazil.
Parliament of WA, 2014. Car Park Full Times Survey 26-27 March 2014. Parliament of WA.
PTA, 2009. The Mandurah Line: Fast facts about our newest train line. Public Transport Authority, Western Australia.
PTA, 2015. Time table.
Reilly, W.J., 1931. The law of retail gravitation. WJ Reilly.
Rietveld, P., 2010. THE IMPACT OF TRAVEL TIME UNRELIABILITY ON THE CHOICE OF RAIL, ACCESS MODE AND DEPARTURE STATION, 12th WCTR, Lisbon, Portugal.
Ryan, M., Lin, T.G., Xia, J.C., Robinson, T., 2016. Comparison of perceived and measured accessibility between different age groups and travel modes at Greenwood Station, Perth, Australia. EJTIR 16, 406-423.
Sanko, N., Shoji, K., 2009. Analysis on the structural characterisitics of the station catchment area in Japan, 11th Conference on Competition and Ownership in Land Passenger Transport, Delft University of Technology, The Netherlands.
Shao, C., Xia, J.C., Lin, T., Goulias, K.G., Chen, C., Han, R., 2015. Logistic regression models for the nearest train station choice: A comparison of captive and non-captive stations. Case Studies on Transport Policy
Stevens, R.E., 2006. The marketing research guide. Routledge.
Upchurch, C., Kuby, M., Zoldak, M., Barranda, A., 2004. Using GIS to generate mutually exclusive service areas linking travel on and off a network. Journal of Transport Geography 12, 23-33.
Viera, A.J., Garrett, J.M., 2005. Understanding interobserver agreement: the Kappa statistic. Fam Med 37, 360-363.
Zhao, H., Yan, X., Gao, Z., 2013. Transportation serviceability analysis for metropolitan commuting corridors based on modal choice modeling. Transportation Research Part A: Policy and Practice 49, 270-284.