AN ORIGIN-DESTINATION BASED TRAIN STATION CHOICE MODEL

Mats Verschuren
NS (Netherlands Railways)

1. INTRODUCTION

For every train trip, train travellers choose a departure and arrival station. They might for example take the bus to the central station, or walk to a local station while having to change trains somewhere. These station choices might change in the following situations:

- When stations are opened or closed
- Railway timetable changes
- Changes in the access and egress system, like the opening of a new metro line to a station.

These changes, that occur frequently in urban settings with lots of stations, might influence the forecast of the number of passengers using each individual station. For NS, the Dutch national railway company, it is needed to precisely forecast this, as it may give cause to (logistical) changes in the operation of the railway system, like adding or removing train connections or train sets at certain stations. The increasing need for more detailed predictions during the timetable design process thus calls for a model improvement.

The origin and destination of a trip are obviously very important factors that influence station choices. The destination is the most important factor that is absent in the current station choice model at NS. To be able to make predictions for the effects of changes in the access and egress system, a distinction should be made among various access and egress mode to travel to or from a station, like the bicycle and bus. The goal of this research is therefore to design an origin-destination based station choice model which takes into account various access and egress modes that represents best a real train traveller's journey from e.g. its home to its destination. Walking, cycling, car (both passenger and driver) and bus/tram/metro have been chosen as access and egress modes in the model.

This paper describes briefly how the model works and why it works this way, and to what extent the model results match the real results. Then, two model applications are shown: a new metro line in Amsterdam, The Netherlands, and several timetable changes in Utrecht, The Netherlands.

2. HOW DOES THE MODEL WORK?

The objective was to create a model that represents best a real train traveller's journey from e.g. its home to its destination. The model therefore consists of three following steps. From each origin postal code to every other destination postal code in the country, it calculates:

- Which travel options do passengers have?
- Which of these travel options will they choose?
- How many passengers does it concern?

These three steps are elaborated in the next paragraphs.

2.1. Which travel options do these passengers have?

Passengers have several options to travel by train from their origin (postal code) to their destination (postal code):

- They can choose various means to travel to the station. In the model, the means walking, cycling, car and bus/tram/metro are distinguished.
- They can choose among various stations, especially in large urban areas.

In reality, passengers take several travel options into consideration before choosing one of them. This set of options, called a choice set, includes options that are possible, logical (e.g. without detours), and feasible (e.g. not walking 10 kilometres to a station).

The model should be able to create choice sets that are similar to the choice set of a real passenger. To do so, research has been done to define rules to transform an objective choice set (all possible options) to a subjective choice set (the perceived choice set by the passenger). These include:

- Maximum walking and cycling times. This filters unfeasible travel options with too long access and egress travel times.
- A maximum difference between a certain travel option and the fastest possible travel option with the same access or egress mode. This prevents non logical detours.

An example of such a choice set of travel options from a postal code in Amsterdam to a postal code in Utrecht is shown in Figure 1. It shows routes with different access modes to various stations in Amsterdam. Routes that would not be considered by real passengers because they are not logical or feasible are excluded by the rules.

Figure 1: A choice set from a postal code in Amsterdam to a postal code in Utrecht

2.2. Which of these travel option will they choose?

The question then is which of these travel options in the choice set passengers would choose. By means of a so called nested logit model, the model calculates the chance that passengers would choose each individual travel option. Research has been conducted to define which aspects influence the station choice of train passengers and to what extent. The aspects that appeared to have a significant influence are included as variables in the model. These include the following:

- The access and egress travel time (by walking, cycling, car or bus/tram/metro).
- Preference factors for access/egress modes. This indicates the base preference of the access/egress mode. For example, to travel from home to a station, the bicycle is generally the most preferred mode, and to travel from a station to work, walking and bus/tram/metro are more popular.

- The generalized train travel time, including the train travel time, adaptation time penalty (frequency) and transfer penalty.
- Variables that indicate the size of the station, as passengers tend to prefer large stations to smaller ones, e.g. because of the facilities at a station.

Other variables that were researched but not appeared significant (using the available data) were the weather, trip purpose, age, comfort, reliability, bicycle and car availability, parking costs, parking spot availability, travel costs, chance to have a seat, and the number of transfers and frequency by bus/tram/metro.

The result of the choice model is, from each postal code to every other postal code, the percentage of passengers that would choose each individual travel option in the choice set.

2.3. How many passengers does it concern?

The last step is to define the number of passengers that travel between the two postal code zones. This is complicated, as NS only knows how many passengers travel from station to station. To convert this to postal code to postal code, the station to station data is combined with survey data. Using these data, per station it can be derived which percentage of passengers come from or go to each postal code. These percentages are then multiplied by the number of passengers between the two stations, which results in the number of passengers between the two postal codes. This is illustrated in Figure 2.

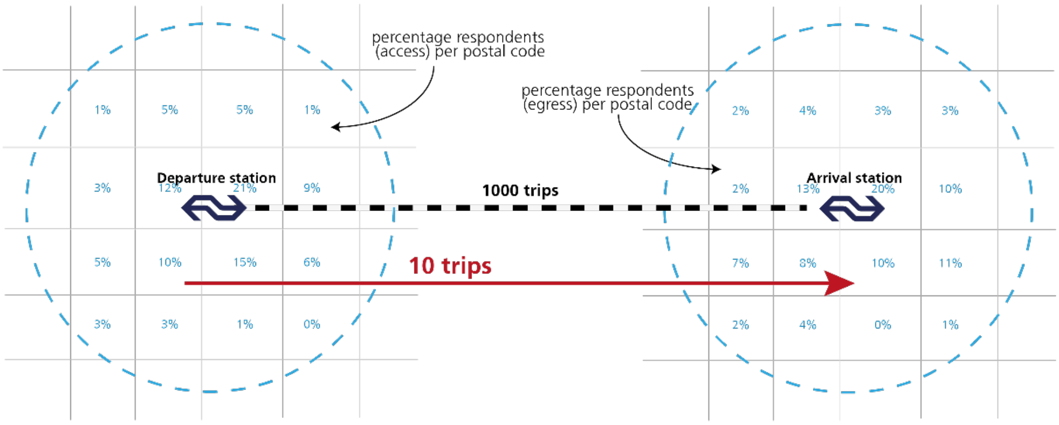

Figure 2: How to convert station-station data to postal code-postal code data

The number of passengers between the two postal codes is then multiplied by the station choice probabilities that were calculated by the nested logit model. This results in the number of passengers per travel option, and thereby also per station.

3. DOES THE MODEL WORK?

The question then is whether the model works well to forecast the effects of the three mentioned changes: opening or closure of a station, timetable changes and changes in the access and egress system. To check this, the model has been validated using several cases, which include a new metro line in Amsterdam, several national timetable changes and newly opened stations. Two of these validations are elaborated in this paragraph: a new metro line in Amsterdam and a timetable change in Amsterdam.

3.1. A new metro line in Amsterdam

In 1997, a new tangential metro line was opened in Amsterdam, connecting several suburban train stations and suburban areas. The revealed and calculated station choice effects are shown in Figure 3. The revealed results do however not solely include station choice effects, but also effects like urban development, timetable changes, and so on. This should be taken into account while comparing the two.

Figure 3: Revealed and calculated station choice effects due to the new metro line

The predictions for Amsterdam Sloterdijk, Amsterdam Zuid WTC and Duivendrecht, the main intercity stations along the new metro line at the time, correspond well to the revealed results. Also, in both situations, the central station (Amsterdam Centraal) suffers a slight decrease in the number of passengers, which is a logical result of the opening of a tangential line. For local stations on the line, like Amsterdam Lelylaan and Amsterdam De Vlugtlaan, the metro was more of a competitor than an addition, resulting in a loss of passengers. The only large difference between the revealed and calculated results is at Amsterdam RAI, which can be explained by several large timetable changes that year and some urban developments in the area around the station. In all other cases, both results are in the same direction and have a comparable extent. This validation therefore shows that the model is well able to make realistic station choice predictions for changes in the access and egress system.

3.2. A timetable change in Amsterdam

Another validation was conducted to check whether the model works to predict the effects of timetable changes. In 2007, a major timetable change took place in Amsterdam due to the addition a new railway connection between Amsterdam Bijlmer ArenA and Amsterdam RAI. Because of this, Duivendrecht lost its intercity status for trains towards the north (to Amsterdam Centraal) and south (to Utrecht) to Amsterdam Bijlmer ArenA. Also, new intercity services between Amsterdam Zuid and Utrecht were opened. The revealed and calculated results are shown in Figure 4. The revealed results show the difference between 2006 and 2008. Again, they also include other factors like urban development.

Figure 4: Revealed and calculated station choice effects due to a major timetable change

At all stations, the results are in the same direction and have a similar extent. Amsterdam Zuid gains lots of passengers due to the new connection. Amsterdam Centraal therefore loses some passengers. Duivendrecht suffers huge losses to Amsterdam Zuid and Amsterdam Bijlmer ArenA. The model predicts a higher increase for Amsterdam Bijlmer ArenA. The station did increase that much, but not in one year. People do not tend to change their behaviour over a day: this takes a while. One year later, the revealed results also showed such a large increase.

This validation thus shows that the model also works well to calculate the effects of railway timetable changes.

4. APPLICATIONS OF THE MODEL

In order to show the various applications of the model, two cases are shown. The first one concerns timetable changes and the opening of a new station, the second one the opening of a new metro line.

4.1. Timetable changes and a new station in Utrecht, The Netherlands

This case shows the effects of timetable changes and the opening of a new railway station on the station choice of passengers. In 2017, NS made major changes to its timetable. This case shows the station choice effects due to these timetable changes in Utrecht, a city of 334,000 inhabitants in the centre of the country. The city has one central station, Utrecht Centraal, that serves all intercity and international trains. Several local train stations are located in other (suburban) parts of the city. The results are shown in Figure 5. The following things changed:

- The new station of Utrecht Vaartsche Rijn opened in 2017. It is located close to the city centre and several residential and commercial areas. As all local trains halt here, it is a good alternative for those who travel to destinations nearby. The model predicts that about 3,000 passengers (per day), who formerly used the stations of Utrecht Centraal and Utrecht Lunetten, will use this station instead.
- Frequencies increased at the station of Utrecht Leidsche Rijn from two to four regional trains per hour. This attracts passengers that formerly used the nearby Utrecht Terwijde and Utrecht Centraal stations, where the four times an hour frequency already existed in these directions.
- Utrecht Zuilen got a new direct regional train connection to Amsterdam. Passengers that live near this station who travel to Amsterdam will make more use of this station rather than the Utrecht Centraal station.
- Due to infrastructural restrictions, the twice an hour local train between Utrecht Centraal, Hilversum and Almere cannot stop at Utrecht Overvecht anymore. This station thereby loses its connection to Almere and loses two trains per hour to Utrecht Centraal and Hilversum. According to the model, people that used to take this train at this station will then take the bike or bus to Utrecht Centraal to catch it there, leading to a large shift of passengers from this station to Utrecht Centraal.

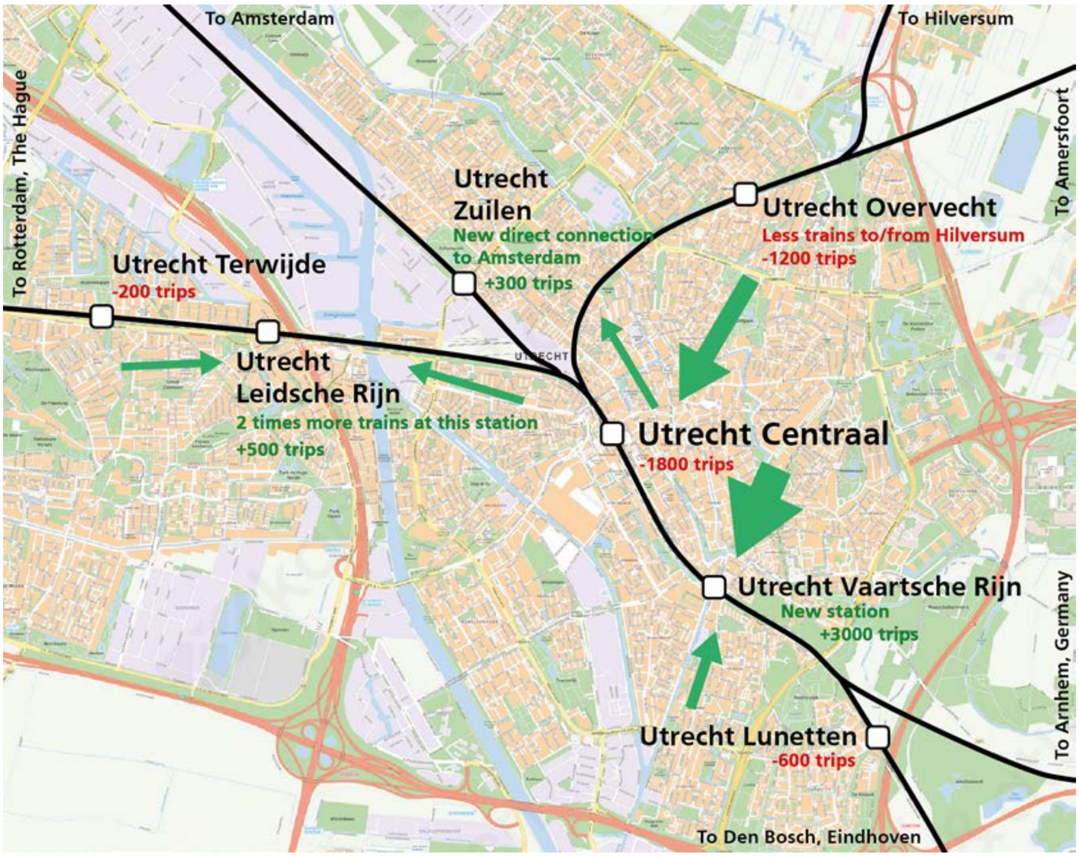

Figure 5: Calculated station choice effects in Utrecht, 2017

The model gives interesting insights in the effects of timetable changes and new stations on the distribution of train travellers among the various station in the city, that would be unknown without a station choice model. It gives the opportunity to foresee these effects in an early phase of the timetable development, thereby leading to a timetable that better reflects the behaviour and wishes of passengers.

4.2. A new metro line in Amsterdam, The Netherlands

Another possible application of the model is to calculate the effects of changes in the access and egress network, like the opening of a new metro line. In Amsterdam, a new metro line will be opened in 2018. It connects the northern part of the city with the central station (Amsterdam Centraal), the city centre and the Amsterdam Zuid station. The Amsterdam Zuid station is located in the central business district of Amsterdam. It serves two metro lines and lots of intercity services. The metro line will dramatically improve its accessibility from the northern part of the city. A shift of train travellers from Amsterdam Centraal to Amsterdam Zuid was therefore expected. The results are shown in Figure 6.

As expected, Amsterdam Zuid has the largest increase. The number of travellers at the central station however does not decrease. A large number of passengers does shift from Amsterdam Centraal to Amsterdam Zuid, but due to its improved accessibility, it will also gain passengers from other nearby stations.

Interestingly, the train stations along the city centre, like Amsterdam Amstel and Amsterdam Lelylaan, suffer the largest changes. They lose a large share of passengers that travel to or from the central areas of the city. Instead of taking the tram or bike to these stations, they will then take the metro to Amsterdam Centraal or Amsterdam Zuid, which provides a faster connection to their destination.

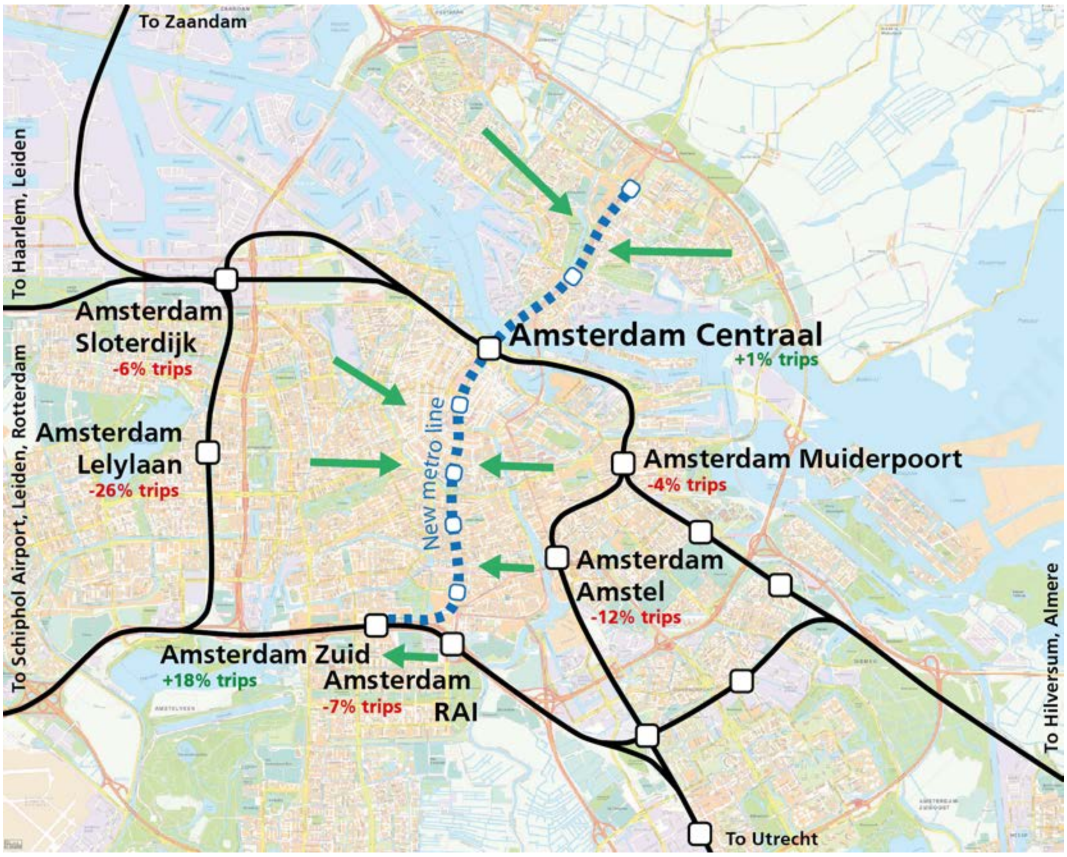

Figure 6: Calculated station choice effects in Amsterdam due to the opening of the new metro line

The results show that station choice is indeed a complex phenomenon, especially in urban settings with lots of stations: after opening the metro line, a lot of passengers will, as expected, choose the Amsterdam Zuid station at the expense of the Central station. However, the effects at some surrounding stations that are not connected by the new metro line will be even bigger. An effect that is difficult to predict without a station choice model.

5. CONCLUSION

Station choices change when stations are opened or closed, due timetable changes and changes in the access and egress system. These changes have a huge effect on the distribution of train travellers among stations and the railway network. It is important for railway operators to predict these effects at an early stage of the timetable development, so that choices can be made that correspond to passengers' behaviour and needs with regard to the timetable.

The proposed model calculate station choice effects in three steps. First, it generates choice sets of travel options from a certain postal code to another postal code. Rules were defined to exclude impossible, unfeasible and illogical travel options. Second, using a logit model, the choice probabilities of each individual travel option are defined by taking into account various aspects that influence passengers' station choice. Finally, it is defined how many passengers travel by train between the two postal codes. Multiplying this by the choice probabilities of the travel options results in the number of passengers per travel option, and thus per station.

The model has been validated by several cases. In this paper, two validations in the Amsterdam area were shown that demonstrate that the model works well to predict the station choice effects for changes in the railway timetable and the access and egress system. The revealed results were similar to the model outcomes. Also, two applications were shown of a new metro line and several timetable changes and new stations that demonstrate the complexity of station choice effects in large urban areas. The model thus gives interesting and detailed insights that otherwise would be unknown, and is able to do so for all three mentioned situations.

The information derived from the model can be a very helpful tool for making decisions in the timetable design process, in order to design railway timetables that better match passengers' needs and wishes. The methodology is generic, so the philosophy used in the model is also suitable for other railway operating companies.

6. BIBLIOGRAPHY

Debrezion, G., Pels, E., & Rietveld, P. (2007). Choice of departure station by railway users. *European Transport*, 78-92.

Guis, N., & Nijënstein, S. (2015). Modelleren van klantvoorkeuren in dienstregelingsstudies. *Colloquium Vervoersplanologisch Speurwerk.* Antwerpen: NS.

Keizer, B. d., Bruyn, M. d., & Vries, B. d. (2009). Nieuw prognosemodel “De Kast” als beleidsinstrument. *Colloquium Vervoersplanologisch Speurwerk.* Antwerpen.

Verschuren, M. (2016). *An origin-destination based train station choice model.* Utrecht: TU Delft, NS.