

## Data-driven public transport ridership prediction approach including comfort aspects

Niels van Oort · Marc Drost · Ties Brands · Menno Yap

**Abstract** The most important aspects on which passengers base their choice whether to travel by public transport are the perceived travel time, costs, reliability and comfort. Despite its importance, comfort is often not explicitly considered when predicting demand for public transport. In this paper, we include comfort level in a modelling framework by incorporating capacity in the public transport assignment. This modelling framework is applied in the public transport model of HTM, the urban public transport operator of The Hague. The current transportation demand is directly derived from smart card data and future demand is estimated using an elasticity based approach. The case study results indicate that not considering capacity and comfort effects can lead to a substantial underestimation of effects of certain measures aiming to improve public transport (up to 30%). We also illustrate that this extended modelling framework can be applied in practice: it has a short computation time and leads to better predictions of public transport demand.

**Keywords:** Comfort · Crowding · Demand prediction · Public transport modelling · Smart card data

---

Niels van Oort  
Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology,  
P.O. Box 5048, 2600 GA Delft, The Netherlands  
Goudappel Coffeng, Mobility consultants, Deventer, The Netherlands  
Email: [N.vanOort@TUDelft.nl](mailto:N.vanOort@TUDelft.nl)

Marc Drost  
HTM, urban public transport operator of The Hague, The Netherlands  
Email: [m.drost@htm.net](mailto:m.drost@htm.net)

Ties Brands  
Goudappel Coffeng, Mobility consultants, Deventer, The Netherlands  
Centre for Transport Studies, Faculty of Engineering Technology, University of Twente,  
P.O. Box 217, 7500AE Enschede, The Netherlands  
Email: [tbrands@goudappel.nl](mailto:tbrands@goudappel.nl)

Menno Yap  
Goudappel Coffeng, Mobility consultants, Deventer, The Netherlands  
Email: [myap@goudappel.nl](mailto:myap@goudappel.nl)

---

## 1 Introduction

To design and to optimise public transport networks, timetables and operations (as for instance demonstrated by Furth et al. 2006, Cats et al. 2012 and Van Oort et al. 2012), insights into future passenger flows are necessary. Although several methods exist to assign demand for public transport to a network (for example Brands et al. 2014), several aspects may be improved, because some important quality aspects are not considered explicitly. Service reliability and robustness for instance, is limitedly accounted for in demand modelling and analysis, as reported by Van Oort et al. (2015b) and Cats et al. (2015). Comfort is another important quality aspect that is often not taken into account when predicting future public transport demand. In this paper we focus on this aspect by incorporating capacity into a data-driven public transport model.

In literature several approaches exist to assign traffic to a public transport network, incorporating capacity. Cepeda et al. (2006) put a hard capacity constraint to public transport links, so that a traveller has to make a detour when capacity is reached on his preferred route. Florian (2002) uses a crowding function where link costs depend on the flow. Just like in static car assignment this practically means that the capacity of links remains unlimited, but the costs will sharply increase if flow is near or even over capacity. Pel et al. (2014) go one step further by introducing a crowding function both upon boarding (using an additive trip penalty) and in transit line sections (using a time multiplier). Schmöcker et al. (2011) use a similar approach by introducing the ‘fail-to-sit’ probability. This requires a Markov type network definition with two states: sit and stand, using priority rules when passengers change state. Pel et al. (2014), Schmöcker et al. (2011) and Cepeda et al. (2006) apply their methods to real world case studies, but only as an assignment method in itself (not as a part of a larger modelling framework).

In this paper we introduce a method that uses a crowding function using both seat capacity and crush capacity values and apply this method to the entire bus and tram network of The Hague in the Netherlands. We incorporate this assignment method in a modelling framework and demonstrate its effect on demand prediction in a case study.

The objective of our research is to enable quick, short-term predictions of future public transport use, incorporating the comfort level, which depends both on seat capacity and on crush capacity. In this paper we present our approach and illustrate it using the case study in The Hague. This paper is organised as follows. In Chapter 2 we present our short-term prediction model, based on smart card data. In Chapter 3 we present our methodology of incorporating comfort aspects. In Chapter 4 the method is demonstrated by applying two example measures to an actual case study. Finally, Chapter 5 reflects on our approach and presents the main conclusions.

---

## 2 Smart card driven public transport predictions

Smart card systems have the potential of providing more and better insights of revealed passenger behaviour (Bagchi and White 2005). Recently, many cities and regions introduced a smart card system for their public transport systems (Pelletier et al 2011). In addition to historical insights, smart card data is also a sound basis for short-term predictions. Based on smart card data and an elasticity method, we developed a short-term prediction approach (Van Oort et al. 2015a). This approach is shortly explained in this section and extended with comfort aspects in the next section.

When an OD matrix is available from observations in smart card data, the step to short and medium time demand prediction can be made. Demand changes may for example occur due to changing the frequency of lines, changing routes of lines, introducing new routes or increasing the speed of a line. When also route choice is introduced in the model, it becomes possible to assess the network effects of this kind of measures. The measures may be temporary or permanent.

In this paper we present a method which is based on demand elasticity: the relative change in costs per OD pair has an effect on transportation demand on that OD pair. Litman (2013) provides a good overview of this type of elasticity models. The costs of a trip are the generalised costs, comprising in-vehicle time, waiting time, number of transfers (penalties) and fare. All attributes of the trip are expressed in monetary values by the coefficients  $\alpha_1$  to  $\alpha_4$ . A Value of Time for the Dutch situation of 6 Euros per hour is used for in-vehicle time and access and egress time ( $\alpha_1$ ). For waiting time, a factor is used which equals one and a half times the factor for in-vehicle time (i.e.  $\alpha_2 = 9$  Euros per hour). The transfer penalty  $\alpha_3$  is set to 0 minutes, because this appeared to better fit the smart card data and this makes detours (because of capacity problems) more likely to occur. Equation 1 shows the composition of the generalised costs for an OD pair  $i,j$ . Note that the coefficient of fare ( $\alpha_4$ ) is equal to 1, because costs are expressed in monetary values.

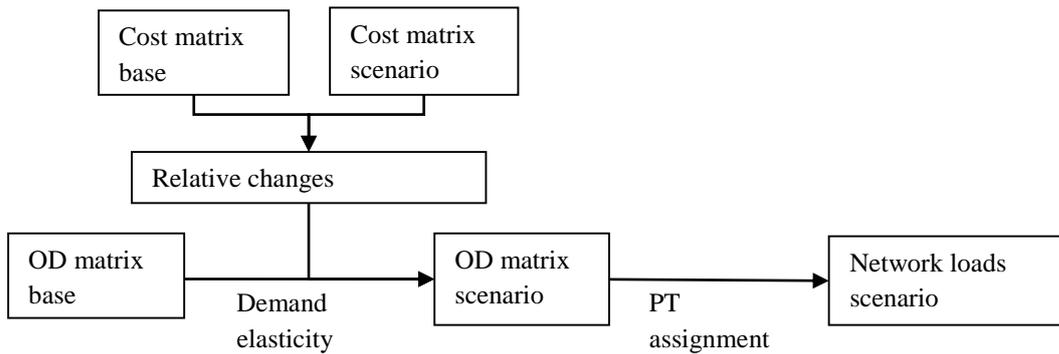
$$C_{ij} = \alpha_1 T_{ij} + \alpha_2 WT_{ij} + \alpha_3 NT_{ij} + \alpha_4 F_{ij} \quad (1)$$

With:

$C_{ij}$	Generalised costs on OD pair $i,j$
$\alpha_1, \alpha_2, \alpha_3, \alpha_4$	Weight coefficients
$T_{ij}$	In-vehicle travel time on OD pair $i,j$
$WT_{ij}$	Waiting time on OD pair $i,j$
$NT_{ij}$	Number of transfers on OD pair $i,j$
$F_{ij}$	Fare to be paid by the traveler on OD pair $i,j$

Fig. 1 shows the steps in our elastic demand calculation. First, generalised cost matrices are calculated for the base situation and the situation that includes a network scenario. For that, a public transport route choice algorithm is used (see Brands et al., 2014). Note that this requires successful calibration of the route choice

parameters: we here assume that the route choice algorithm is able to reproduce the line loads in the base situation. Comparing the cost matrices results in relative cost changes per OD pair. Using the OD matrix for the base situation (from smart card data) and an elasticity value (see for instance Wardman, 2012), the relative changes in OD flows are calculated, resulting in an OD matrix for the network scenario. This process is also captured in Equation 2. The new OD demand (in the situation that includes the network scenario) is calculated from the base demand, the costs in both situations and the elasticity values. The final step is to assign this OD demand to the public transport network, again using a public transport route choice algorithm. The result is a loaded network for the situation with network scenario. From this loaded network, performance indicators can be derived, for example total number of passenger, number of passenger kilometres or occupancy rates.



**Fig. 1** Schematic representation of the demand prediction model

$$D_{ij}^1 = \left( E \left( \frac{C_{ij}^1}{C_{ij}^0} - 1 \right) + 1 \right) * D_{ij}^0 \quad (2)$$

With:

- $D_{ij}^1$  Demand on OD pair i,j in the scenario
- E Elasticity
- $C_{ij}^1$  Generalised costs in the scenario
- $C_{ij}^0$  Generalised costs in the base situation
- $D_{ij}^0$  Demand on OD pair i,j in the base situation

---

## 3 Incorporating comfort effects in predictions

### 3.1 Effects of comfort and crowding

In urban public transport systems, the problems related to the comfort and capacity that is supplied to passengers tend to increase. However, comfort is hardly incorporated in current public transport demand models, while comfort and crowding levels do influence passengers' route choice and mode choice. Furthermore, the capacity available on public transport lines is not unlimited. For a realistic modelling of passenger demand and route choice in crowded public transport systems, it is important that the number of passengers assigned to a public transport service does not exceed the capacity of this service. Therefore, as an extension of this elastic demand model, comfort effects are incorporated in the predictions. Such a model extension is useful when a heavily used public transport network is studied.

An example in a temporary context is a transit line that is not available any more due to engineering works. In order to anticipate on this accordingly, the PT operator likes to check whether the capacity of transit lines on alternative routes is reached. Also in a permanent situation this may be relevant: when the frequency on a crowded, high-frequency line is increased, the current models only predict a very limited passenger growth due to slightly reduced waiting times. However, in reality passengers may decide to start using the service due to an increased comfort level as well.

In general, crowding in public transportation can have three different effects:

- The in-vehicle time perception of travellers may change, because a crowded vehicle is perceived as less attractive than a quiet vehicle.
- Passenger demand for a certain PT service exceeds the supplied capacity. On the short run this denied boarding leads to passengers having to wait another interval time for the next vehicle. On the longer run, for a permanent situation and for a situation with published maintenance works, an equilibrium situation will occur, where passengers adjust route and mode choice such that the supplied capacity is not exceeded. For unplanned disturbances, no equilibrium situation is expected.
- Dwell time of PT vehicles is higher if the crowding level is higher, because the boarding and alighting process will take more time.

In this study, we focus on incorporating the first two comfort effects in PT demand models.

### 3.2 Method

The two mentioned comfort effects are incorporated in the demand model by making the in-vehicle travel time component of the generalised costs function dependent on the passenger load. For this, a crowding function is used. The perceived in-vehicle

---

travel time is calculated as a multiplication factor over the real, objective travel time, which depends on the passenger load in relation to the number of seats and to the capacity for standing passengers. First, the transformed volume / capacity ( $VC$ ) ratio is determined using Equation 3. The result of this formula is that  $VC$  is equal to 1 if the passenger load  $L$  equals the seat capacity  $C_{seats}$  of a certain vehicle.  $VC$  is equal to 2 if the load  $L$  is equal to the crush capacity  $C_{crush}$  (seat capacity plus standing passengers). The seat capacity and crush capacity can be specified for each PT line and each modelling period (morning peak, evening peak, off-peak hours) separately. This enables distinguishing between different vehicle types and lengths used on different lines during different times of the day.

$$VC = \begin{cases} \frac{L}{C_{seats}} \\ 1 + \frac{L - C_{seats}}{C_{crush} - C_{seats}} \end{cases} \quad (3)$$

Most studies on valuation of crowding only use the load factor (the passenger load  $L$  divided by the seat capacity  $C_{seats}$ ) to express crowding effects (Wardman and Whelan 2011). In our study, we explicitly distinguish between the seat capacity  $C_{seats}$  and the crush capacity  $C_{crush}$  of PT vehicles. Taking both the seat capacity and crush capacity into account has the advantage that different types of vehicles with different configurations (with relatively few or many seats with respect to the total capacity) are correctly compared. In a PT vehicle with a relatively high number of seats relative to the total crush capacity (e.g. an intercity train service), crowding will be perceived differently compared to a vehicle with a relatively low number of seats in relation to the crush capacity (e.g. a light rail or metro service). This means that in reality the load factor only makes sense, when it is related to the total crush capacity of a vehicle.

In their meta study to crowding valuation in public transport, Wardman and Whelan (2011) indicate that the in-vehicle time multiplier should be expressed as function of the load factor, up to a load factor of 100% (seat capacity  $C_{seats}$ ). From load factors higher than 100%, the vehicle configuration needs to be considered as well. From a load factor higher than 100% we determine the in-vehicle time multiplier as function of both the seats and crush capacity.

Based on the  $VC$ -ratio, a piecewise linear function is used to determine the factor for perceived travel time  $F$ , based on the values in Table 1. Starting from 80% seat occupation the comfort level starts to decline (Douglas Economics, 2006). According to Douglas Economics (2006), the multiplication factor equals 1.1 when a 100% seat occupation rate is reached. Revealed occupation rates using smartcard data are used to determine the crush capacity of different types of PT vehicles. The crush capacity as specified by the manufacturer, assuming 4.5 persons/m<sup>2</sup>, appears not to be realised in practice in the Netherlands. Based on vehicle configuration and the maximum number of passengers per vehicle found in smartcard data, we determined that the crush capacity  $C_{crush}$  in the vehicles in our study is reached at the level of 3.5 persons/m<sup>2</sup>. In MVA Consultancy (2008) seats and standing multipliers are expressed as function of the number of standing passengers per m<sup>2</sup>. Using these crowding

multipliers we determined that the factor for perceived travel time increases with 0.64 when moving from  $C_{seats}$  to  $C_{crush}$ . Wardman and Whelan (2011) conclude that the in-vehicle time perception increases linear with increasing crowding levels: non-linearity's could not be justified empirically. This leads to a piecewise linear function with crowding multipliers as shown in Table 1. Using Equation 4 this factor is applied over the real link travel times to calculate the perceived travel time, which replaces real travel time in the generalised costs function, see Equation 5.

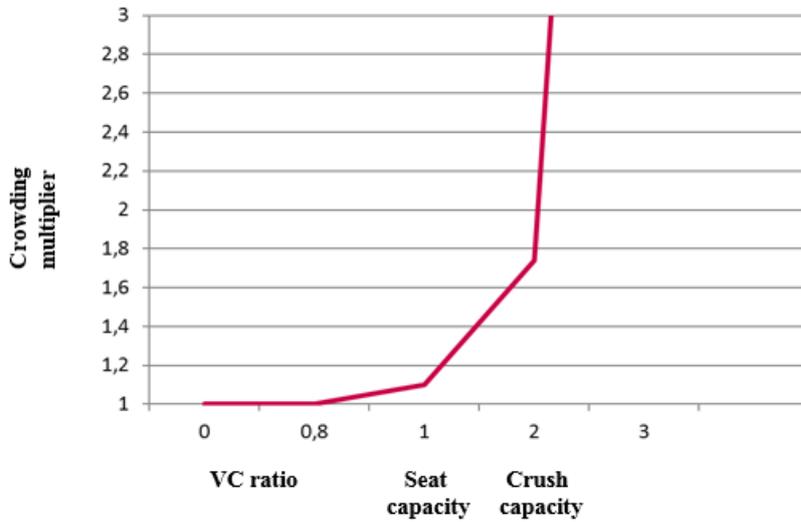
To prevent the assignment of passengers to a vehicle where  $C_{crush}$  has already been reached, the function sharply increases for VC values larger than 2.0. In this way, the attractiveness of a route with a completely crowded vehicle decreases in such a way that passengers will change their route or mode choice. This leads to the crowding function as plotted in Fig. 2.

VC	Perceived travel time factor F
0 – 0.8	1
0.8 – 1.0	1 – 1.1
1.0 – 2.0	1.1 – 1.74
2.0 – 3.0	1.74 - 10

**Table 1** Relation between VC and the perceived travel time factor. A factor of 1 implies no additional perceived travel time

$$T_{ij}^{per} = T_{ij} * F \quad (4)$$

$$C_{ij} = \alpha_1 T_{ij}^{per} + \alpha_2 WT_{ij} + \alpha_3 NT_{ij} + \alpha_4 F_{ij} \quad (5)$$



**Fig. 2** Crowding function

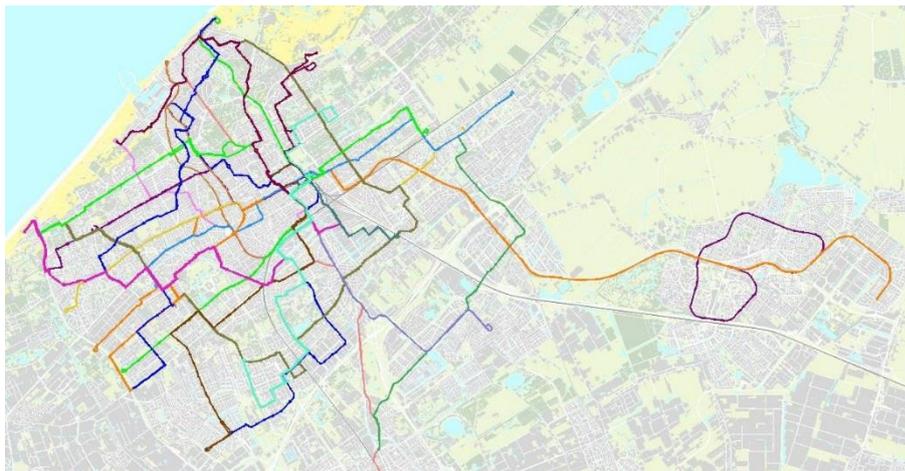
---

Note that the load is needed for a 1 hour time period, because the capacity is also given per hour (resulting from the frequency and seat / crush capacity per vehicle). When the modelled time period is longer, a correction factor is used. Depending on the evenness of the load distribution over this time period, this factor is equal to the period length in hours (in case of a perfectly uniform distribution), or is smaller than the period length. If the distribution is uneven, the busiest hour is taken as representative for the entire time period, by dividing the real number of hours by the busiest hour factor.

Since the costs of travelling now depend on the load, an iterative assignment is necessary. This assignment procedure is comparable to a user equilibrium assignment which is common in road network assignment when incorporating congestion effects. The iterative procedure is repeated until convergence is reached between iterations  $N$  and  $N+1$ . We specified a convergence criterion of 5%.

## 4 Case study

The method is applied to two different case studies based on the entire bus and tram network of The Hague, the Netherlands, operated by the public transport operator HTM (see Fig. 3). In total, the network includes 434 centroids, 64 (unidirectional) transit lines and 892 stops.



**Fig. 3** All tram and bus lines in the case study network of The Hague (each line is plotted in a separate colour)

The method is implemented in the transportation modelling software package OmniTRANS. The computation time required for the prediction of new demand for one time period (i.e. a morning peak) including a complete iterative elastic assignment is around 25 minutes on a regular core i5 laptop. Two different cases are investigated:

- Increasing the frequency of tram line 15 from 6 to 8 trams per hour during the morning and evening peak. Since this tram line has a high peak demand, the effect of an increase in frequency on PT demand is investigated for the situation with and without incorporating comfort effects of this measure.
- Transformation from bus to tram line. This is a hypothetical case, in which it is assumed that the busy bus line 25 is transformed to a tram line. The effect of this transformation on PT demand is evaluated for the situation with and without considering comfort effects. Because of the larger capacity of a tram, the frequency during the between-peak hours is reduced from 8 busses per hour to 6 trams per hour. For all other hours, no change in frequency is assumed.

Table 2 below shows the length and the number of passengers of these two case study lines.

	Line length (km)	Number of passengers per average workday (x1000)
Tram line 15	9.4	12
Bus line 15	7.6	7

**Table 2:** Line length and number of passengers per average workday per case study PT line

#### 4.1 Increased frequency of tram line 15

Table 3 shows the expected increase in public transport demand for tram line 15 without and with considering comfort effects. Table 4 shows the expected absolute increase in ridership as consequence of this measure on the public transport network as a whole, considering substitution effects between lines as well. From this table we can conclude that 165 new passengers are expected in both the morning peak and evening peak, when only benefits from a reduced average waiting time are considered. When both the effects of reduced waiting time and improved comfort are incorporated, 240 and 200 new passengers are expected in the morning and evening peak respectively. Since in the morning peak public transport demand is more clustered within a small period, comfort benefits of this measure are larger during the morning peak, compared to the evening peak where demand is more uniformly distributed. We can conclude that traditional models, which do not consider comfort benefits, tend to underestimate the additional ridership because of this measure with 30% in the morning peak, and with 20% in the evening peak. This means that a substantial part of the benefits of this measure can be attributed to improved comfort levels, which would not be detected otherwise. Figure 4 visualises the modelled relative effect of this measure with and without considering comfort effects. It shows that the higher frequency of tram line 15 attracts some passengers from the parallel tram line 1 (shown in red in Figure 4).

	Model without comfort	Model including comfort
Average work day	+8%	+10%

**Table 3:** Estimated relative increase in PT demand tram line 15 after frequency

increase in morning and evening peak (without and including comfort effects)

	Model without comfort	Model including comfort
Morning peak	+ 165	+ 240
Evening peak	+165	+ 200

**Table 4:** Estimated increase in ridership on a network level after increase of frequency in morning and evening peak (without and including comfort effects).



**Fig. 4:** Relative network effects of frequency increase on link loads a) without considering comfort (left) and b) considering comfort effects (right) during a morning peak.

#### 4.2 Transformation from bus to tram line

In Table 5, we show the effects of the hypothetical transformation of bus line 25 (Vrederust – Grote Markt) to a tram line. In Table 6, the absolute effects on public transport demand are shown on a network level for the morning peak, evening peak, and remaining part of the work day, thereby again considering network substitution effects as well. This table shows that the model without comfort effects does not predict a change in public transport demand during the morning peak and evening peak, since all frequencies remain equal. The model including comfort effects however predicts 40 and 20 new passengers in the morning peak and evening peak respectively due to the higher seat probability and higher comfort level supplied to passengers, even with all frequencies remaining equal. Again, comfort benefits are larger in the morning peak because of the higher concentration of demand. For the remaining part of the work day the model without comfort predicts a decrease in ridership (-50 passengers), because the reduced frequency from 8 busses to 6 trams per hour in the between-peak interval increases average waiting times. The model including comfort shows a trade-off. On the one hand, average waiting times increase somewhat because of the lower frequency. On the other hand, the total capacity supplied per hour increases substantially because the capacity per tram overcompensates the lower frequency, which improves the public transport attractiveness regarding comfort. For this period of the day, the additional comfort benefits are larger than the costs of additional waiting time, leading to 200 new

passengers in total. Figure 5 shows the relative passenger effect for the remaining part of the work day, predicted by the model without and with comfort.

It should be noted that this effect of a transformation of a bus line into a tram line only incorporates the capacity related comfort aspects. It is known from Bunschoten et al. (2013) that in-vehicle travel time in rail vehicles is experienced more positively than in-vehicle travel time in bus vehicles. Taking this into account would further increase the benefits of a transformation from bus to tram.

	Model without comfort	Model including comfort
Average work day	-3%	+3%

**Table 5:** Estimated relative effect on ridership of line 25 after transformation to tram line (without and including comfort effects)

	Model without comfort	Model including comfort
Morning peak	0	+ 40
Evening peak	0	+ 20
Remaining part work day	-50	+ 200

**Table 6:** Estimated effect on ridership on a network level after transformation of bus line to tram line (without and including comfort effects).



**Fig. 5:** Relative network effects of transformation of bus line 25 to tram line on link loads a) without considering comfort (left) and b) considering comfort effects (right) for a work day (excluding morning and evening peak).

## 5 Conclusions and further research

This paper explores the potential for quick, short-term public transport predictions using smart card data. The intention is to provide relatively simple (easy to build) models to perform what-if analyses. We explicitly take comfort into account, since it is a relevant quality indicator, which is often neglected.

We showed that the ridership effect of a frequency increase in a congested transit line and a transformation from bus to tram becomes larger when comfort effects are included. The results indicate that not considering capacity and comfort effects can

---

lead to a substantial underestimation of effects of certain measures aiming to improve public transport (up to 30%). From a policy perspective this also indicates that benefits of such measures can be underestimated when comfort is not incorporated in the demand modelling framework. We also illustrated the potential of these models to be applied in practice, given the limited computation times required.

We also showed network wide effects and will provide results for different types of measures in the network. Interesting future applications include validation of the model by revealed effects after a measure has been implemented (using smart card data). Also, incorporating effects of denied boarding and extended dwell times at stops because of crowding are interesting future extensions of the modelling framework. The quantified comfort effect can also be used in cost-benefit analyses, because decreasing perceived travel time can be considered as a societal benefit.

**Acknowledgements:** This research was supported by HTM and Goudappel Coffeng. The authors are grateful that they shared their data and tooling.

## References

- Bagchi, M., White, P. (2005) The Potential of Public Transport Smart Card Data. *Transport Policy*, 12(5), 464-474.
- Brands, T., de Romph, E., Veitch, T., Cook, J. (2014) Modelling public transport route choice, with multiple access and egress modes. *Transportation Research Procedia*, 1, 12-23.
- Bunschoten T., E. Molin, R. v. Nes (2013) Tram or bus; does the tram bonus exist? European Transport Conference. Frankfurt, Germany, AET
- Cats O., Larijani A.N., Ólafsdóttir A., Burghout W., Andreasson I. and Koutsopoulos H.N. (2012). Holding Control Strategies: A Simulation-Based Evaluation and Guidelines for Implementation. *Transportation Research Record*, 2274, pp.100-108.
- Cats, O., M.D. Yap, N. van Oort (2015) Exposing the role of exposure: identifying and evaluating critical links in public transport networks, INSTR conference, Nara, Japan, 2015.
- Cepeda, M., Cominetti, R., Florian, M. (2006) A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. *Transportation Research Part B: Methodological*, 40, 437–459.
- Douglas Economics (2006) Value and Demand Effect of Rail Service Attributes. Report to RailCorp. Wellington, New Zealand.
- Florian, M. (2002). Frequency based transit route choice models. Chapter 6 in: *Advanced Modeling for Transit Operations and Service Planning*, ed. William H.K. Lam, Michael G. H. Bell.
- Furth, P. G., B. Hemily, T.H.J. Muller, J.G. Strathman (2006) TCRP Report 113: Using Archived AVL-APC Data to Improve Transit Performance and Management. Washington, D.C.

- 
- Litman, T. (2013) Transport Elasticities: Impacts on Travel Behaviour. Sustainable Urban Transport Technical Document, GIZ.
- MVA Consultancy. (2008) Valuation of Overcrowding on Rail Services. Prepared for Department for Transport.
- Pelletier, M., Trepanier, M. and Morency, C. (2011) Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557-568.
- Pel, A. J., Bel, N. H., Pieters, M. (2014) Including passengers' response to crowding in the Dutch national train passenger assignment model. *Transportation Research Part A: Policy and Practice*, 66, 111-126.
- Schmöcker, J. D., Fonzone, A., Shimamoto, H., Kurau chi, F., Bell, M. G. (2011) Frequency-based transit assignment considering seat capacities. *Transportation Research Part B: Methodological*, 45(2), 392-408.
- Van Oort, N., Boterman, JW & Nes, R van (2012). The impact of scheduling on service reliability: trip-time determination and holding points in long-headway services. *Public Transport*, 4(1), pp. 39-56.
- Van Oort, N., T. Brands, E. de Romph (2015a), Short term ridership prediction in public transport by processing smart card data, *Transportation Research Record*, (in press)
- Van Oort, N., T. Brands, E. de Romph, J.A. Flores (2015b), Unreliability effects in public transport modelling, *International Journal of Transportation Vol.3, No.1*, pp.113-130.
- Wardman, M. (2012) Review and meta-analysis of U.K. time elasticities of travel demand. *Transportation*, 39(3), 465-490.