6th European Transport Research Conference
MOVING FORWARD:
Innovative Solutions for Tomorrow's Mobility
National Stadium, Warsaw, Poland - 18-21 April 2016
www.traconference.eu

Proceedings of 6th Transport Research Arena, April 18-21, 2016, Warsaw, Poland

Data driven enhancement of public transport planning and operations: service reliability improvements and ridership predictions

Niels van Oort[a,b]

[a]*Delft University of Technology*
[b]*Goudappel Coffeng, mobilty consultants*

Abstract

Automatic Vehicle Location (AVL) and smartcard data are of great value in planning, design and operations of public transport. We developed a transport demand model, which utilizes smartcard data for overall and what-if analyses, by converting these data into passengers per line and OD-matrixes and allowing network changes on top of a base scenario. This new generation model serves in addition to the existing range of transport demand models and approaches. It proved itself in practice during a case study in The Hague, where it helped the operator gain valuable insights into the effect of small network changes, such as a higher frequency.
Data also supports measures to improve service reliability. We introduced a new network design dilemma, namely the length of a transit line vs. its reliability. Long lines offer many direct connections, thereby saving transfers. However, the variability in operation is often negatively related to the length of a line, leading to poorer schedule adherence and additional waiting time for passengers. A data driven case study shows that in the case of long lines with large variability, enhanced reliability resulting from splitting the line could result in less additional travel time. This advantage compensates for the additional time of transferring if the transfer point is well chosen.

Keywords: public transport, data, ridership prediction, service reliability

---

1 Introduction

Tighter public transport budgets in the Netherlands mean, similar to other countries, funding is being reduced, which makes it even more challenging to improve the quality and capacity, as required by the demanding (potential) customers. Simultaneously, maximizing efficiency is challenging, but not impossible though. With the right data and tools operators are able to address the key factors to enhanced and more cost efficient public transport, namely shorter travel times and enhanced service reliability (Lee et al. 2014, Van Oort et al. 2015a). More insights into the mechanisms of these aspects are becoming available by accessing several new data sources. Since the last decade, the public transport industry have started to consider how data retrieved from passenger smart cards (see e.g. Neema et al., 2015 and Pelletier et al., 2011), board computers (AVL (automatic vehicle location), see e.g. Furth et al., 2006 and Hickman, 2004) and mobile phones (see e.g. Calabrese 2011) could improve the design of public transport networks and timetables to achieve efficient and high quality operations. In addition, social media data, such as user data of Facebook, Twitter and Flickr, may yield new knowledge on public transport usage (see e.g. Bregman, 2012). Video, Bluetooth and Wi-Fi-trackers may provide new insights into pedestrian flows in stations, at platforms and in vehicles (e.g. Van den Heuvel, 2015).

The objective of this paper is to illustrate the potential benefits of using data, focusing on Dutch AVL and smartcard data. We illustrate this value by two example cases.

The outline of this paper is as follows: First, a short description of Dutch AVL and smartcard data is given in Section 2. Section 3 presents our first case of applying data in real life: a prediction model supporting frequency impacts on passengers. The second case deals with service reliability. We demonstrate the impact of line length in Section 4. The paper ends with conclusions, acknowledgements and references.

2 Dutch public transport data

This paper presents two examples from the Netherlands, where new data sources became available recently. The challenge is to transform that data into information that can be used for service improvements and thus to make optimal use of the value it possesses.

2.1 The Dutch Smart Card System: OV-Chipkaart

In recent years The Dutch smartcard, the OV-Chipkaart, has been introduced (Cheung 2006). The Dutch smartcard uses nfc-chip technology and passengers have to check in and out on all public transport in the Netherlands (bus, tram, metro and train) to pay for their trips. Therefore, valuable information is obtained continually about both origins and destinations of all public transport users (on station/stop level). The check-in and check-out-devices are either located on the platform or station entrances (for trains and metros) or located inside the vehicle (for buses and trams). The most detailed information is available in the latter case, where each trip in a journey is tracked (a journey may consist of multiple trips, with a transfer in between). The route through the public transport network is therefore completely traceable. In case the smartcard devices are located on platforms or station entrances, an additional route search in the network is often necessary to match passengers to particular lines, as the only data available is from smartcards entering the system at point A and leaving the system at point B. More insight into the Dutch smartcard is provided by Van Oort et al. (2015c).

2.2 The Dutch AVL System: GOVI

In the Netherlands, all public transport operators cooperate with Borderless Public Transport Information ("Grenzeloze Openbaar Vervoer Informatie"), in short: GOVI. This initiative intends to make a wide range of public transport information, from timetables and routes to fares, vehicle location and punctuality, publicly available and processable. The data exchange interfaces are defined by a set of standards called BISON, in order to ensure the transferability of tools, apps and analyses.

GOVI was designed to facilitate data communication between vehicles and the land side to enable dynamic passenger information on board and at stops. As an additional benefit, the actual and scheduled vehicle positions and times are logged in a database. Although this database was not the main objective of the GOVI system, it is extremely helpful to monitor and analyze public transport performance. In Van Oort et al. (2015b) more insights into the GOVI system and its data are provided.

In the next sections we illustrate two applications of smartcard and GOVI data, showing the value of data in the planning process of public transport.

3 Case 1: Forecasting passenger numbers by smartcard data

3.1 Ridership prediction model

Several studies exist where origin destination information (OD-matrices) is derived from the smartcard data, either in aggregated form (e.g. Wang et al. 2011) or disaggregated form (e.g. Munizaga and Palma 2012 and Bouman et al. 2013). Once a matrix is created that reproduces the passenger trips, it becomes possible to perform what-if analyses. In our approach we apply an elasticity approach in combination with the smartcard data. These kind of models are relatively simple to make (thus saving time and budget) and can make good use of the available data. The accuracy level is lower than multimodal models, but still enough for several research objectives.

Equation 1 shows the calculation of generalized costs for OD pair i,j. Note that the coefficient of fare α_4 is equal to 1, because costs are expressed in monetary values.

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

With:

C_{ij} Generalized costs on OD pair i,j
α_1,α_2,α_3,α_4 Weight coefficients in generalized costs calculation
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

Figure 2 shows the steps in our elastic demand calculation. First, using a public transport route choice algorithm (e.g. Brands et al. 2014), the generalized cost matrices are calculated for the base situation and the situation that includes a network scenario. Note that this requires successful calibration of the route choice parameters. Comparing the cost matrices results in relative cost changes per OD pair. Using the OD matrix for the base situation (from smartcard data) and an elasticity value (e.g. Wardman 2012, TRB 2004, Balcombe et al. 2004), the relative changes in OD flows are calculated, resulting in an OD matrix for the network scenario. The final step is to assign this OD demand to the public transport network, again using the public transport route choice algorithm. More details of our (deterministic) approach are provided in Van Oort et al. (2015c).

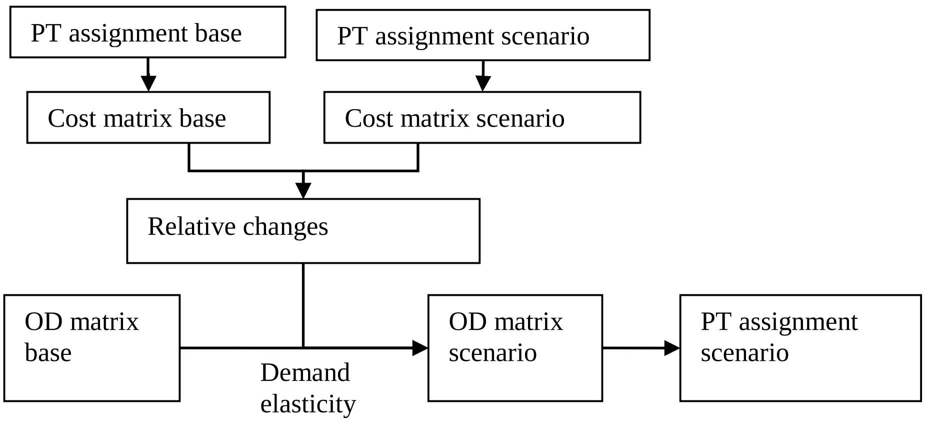

Fig. 2: Schematic representation of the demand prediction model.

Comfort effects are incorporated in our model by making the in-vehicle travel time component of the generalized 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 2. The result of this formula is that VC = 1 when the passenger load L equals the seat capacity C_{seated} of a certain vehicle. VC = 2 when the load L is equal to the crush capacity (seated plus standing passengers) C_{crush}. The seat capacity and crush capacity can be specified for each public transport line and each modelling period (morning peak, evening peak, off-peak hours) separately, in order to distinguish between different vehicle types and lengths used on different lines during different times of the day.

$$VC = \begin{cases} \dfrac{L}{C_{seated}} \\ 1 + \dfrac{L - C_{seated}}{C_{crush} - C_{seated}} \end{cases} \tag{2}$$

Based on the VC ratio, a piecewise linear function is used to determine the factor for perceived travel time F. Starting from 80% seat occupation the comfort level starts to decline (following 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 public transport vehicles. The crush capacity as specified by the manufacturer, assuming 4.5 persons/m^2, appears not to be realized 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 with 3.5 persons/m^2. Using the crowding multipliers from MVA Consultancy (2008) – where seated and standing multipliers are expressed as function of the number of standing passengers per m^2 – we determined that the multiplication factor increases with 0.64 from C_{seated} 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 Figure 2.

Using Equation 3 this factor is applied over the real link travel times to calculate the perceived travel time, which replaces real travel time in the generalized costs function (Equation 1).

$$T_{ij}^{per} = T_{ij} * F \tag{3}$$

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

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 included. 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.

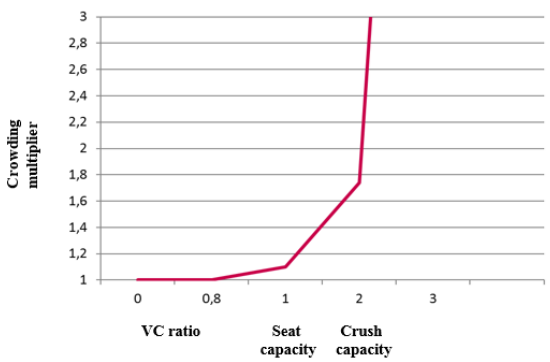

Fig. 2: Crowding function.

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%.

3.2 Predicting impacts of frequency increase

We applied our elasticity approach and performed predictions (with incorporation of comfort effects) in a case study. We connected smartcard data of HTM, the tram operator in The Hague (about 500,000 inhabitants, 3rd largest city of the Netherlands) to a transport model built in OmniTRANS. The city of The Hague has 12 tram/light rail lines with a total length of about 335 km, which transport about 70 million passengers per year.

The case consists of increasing the frequency of tram line 15 (line length: 9.4 km) from 6 to 8 trams per hour during the morning and evening peak. As this line has a high peak demand, the effect of an increase in frequency on public transport demand is investigated for the situation with and without incorporating comfort effects of this measure.

From this case we can conclude that 165 new passengers are expected in both the morning and evening peak, when only considering the benefits of the reduced average waiting time. 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. The higher number in the morning results from public transport demand being more clustered within a small period, which achieves greater comfort benefits then in the evening, when demand is more uniformly distributed. We can conclude that the traditional approach, which does not consider comfort benefits, tends to underestimate the additional public transport demand of the aforementioned higher frequency with 30% in the morning peak, and 20% in the evening peak. This means that a substantial part of the benefits of this measure can be attributed just to improved comfort levels, which would not have been detected otherwise. Figure 3 visualizes the modelled relative effect of this measure with and without considering comfort effects. It shows that the higher frequency of line 15 attracts some passengers from the parallel tram line 1 (shown in red in Figure 3).

Fig. 3: Relative network effects [% ridership change] of frequency increase on link loads a) without considering comfort (left) and b) considering comfort effects (right) during a morning peak.

4. Case 2: enhanced service reliability by optimized line length

While case 1 illustrated the value of smart card data, this section will demonstrate the potential value of AVL data.

4.1 Service reliability

Service reliability is a key quality indicator for public transport and has become increasingly important over the last years. It is defined as the certainty of service aspects compared to the schedule as perceived by the user and is one of the most important quality aspects of public transport (Lee et al. 2014). Actual vehicle trip time variability (i.e. service variability) affects service reliability and passenger travel time. The impacts of unreliable services on passengers are:

- average travel time extension;

- increased travel time variability;
- a lower probability of finding a seat in the vehicle.

Several traditional quantifications of service reliability, such as punctuality and regularity, lack attention for the impact on passengers. Traditional indicators focus too much on the supply side of public transport, which does not allow for a proper analysis of passenger effects. To deal with the shortcomings of traditional indicators, Van Oort (2014) developed a new indicator, being the average additional travel time per passenger. This indicator translates the supply-side indicators, such as punctuality, into the average additional travel time that a passenger needs to travel from an origin to a destination due to service variability. The average additional travel time may be calculated per stop or per line and it enables explicit consideration of service reliability in cost-benefit calculations, since the level of service reliability may be translated into regular travel time.

4.2 Instruments improving service reliability

Literature shows that in urban public transport substantial attention is usually given to ways to improve service reliability on an operational level (Van Oort 2014). However, it is not clear how and to what extent strategic and tactical design decisions in public transport systems might affect service reliability.

We developed several strategic and tactical measures that offer good possibilities to improve service reliability. The selected instruments adjust the process of public transport thereby being a regulator in the feedforward loop, as shown by Figure 4. More insights into these instruments are provided by Van Oort (2011). In the next section we continue by addressing the instrument of line length.

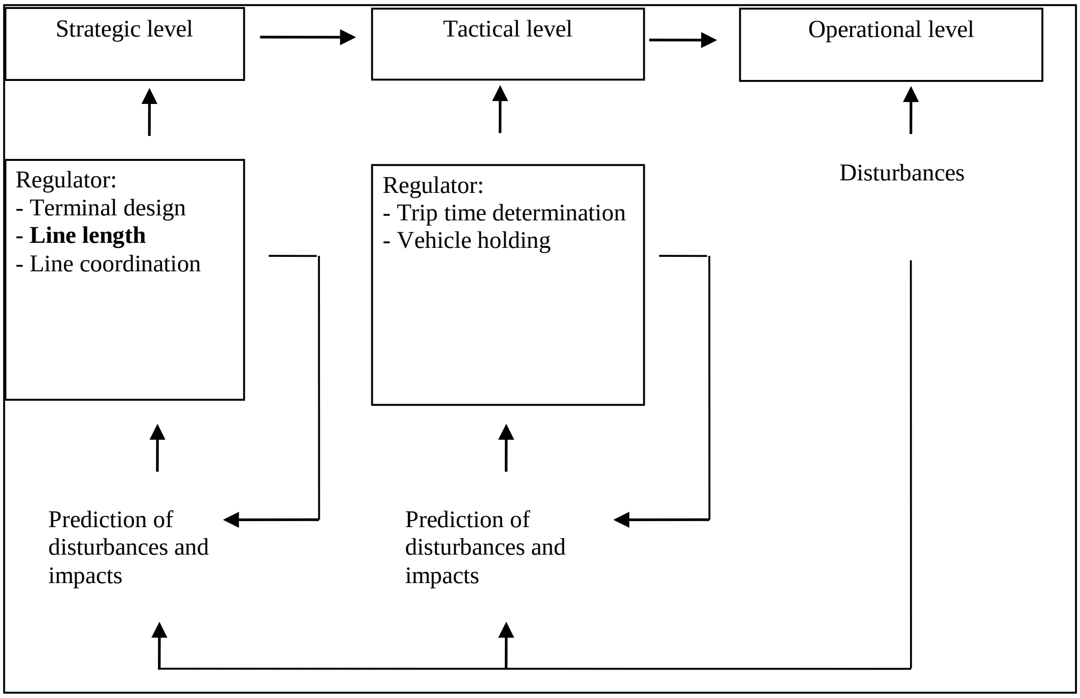

Fig. 4: Feedforward loop in public transport planning, including our investigated instruments at strategic and tactical level.

To apply the feedforward loop, shown in figure 4, it is of utter importance that the expected disturbances of decisions during the strategic and tactical design can be forecasted quickly and accurately. AVL data is a valuable source to support this. In addition, the impacts of the disturbances on passengers, the expected level of service reliability, should be predicted as well. In this way, an iterative improvement process of the regulator is possible, optimizing the impact of passengers. Van Oort (2014) elaborates on this in more detail.

4.3 The instrument of line length design

In public transport, there has been a tendency to connect or extend lines. The added value of these new links is that they allow for more direct connections from origin to destination. When a particular line for instance is extended from a city center to a suburban area, passengers may skip a transfer, which saves them travel time. However, this time saving may be impeded completely by frequent delays resulting from the greater potential variability and thus unreliability of a longer service. Figure 5 illustrates the effects of decreasing line length. Line 1 operates from A to B. The variability of driving time increases along the line, of course depending on several aspects such as signal priority and exclusive lanes (Van Oort 2015b). Van Oort (2011) presents more details on the specific relationship of line length and service reliability. Line 2 follows the same route as line 1, but now it is divided into two parts: From A to C and from C to B.

Two main differences exist between lines 1 and 2:

- The variability of line 2 is smaller;
- Introduction of a transfer at C for line 2.

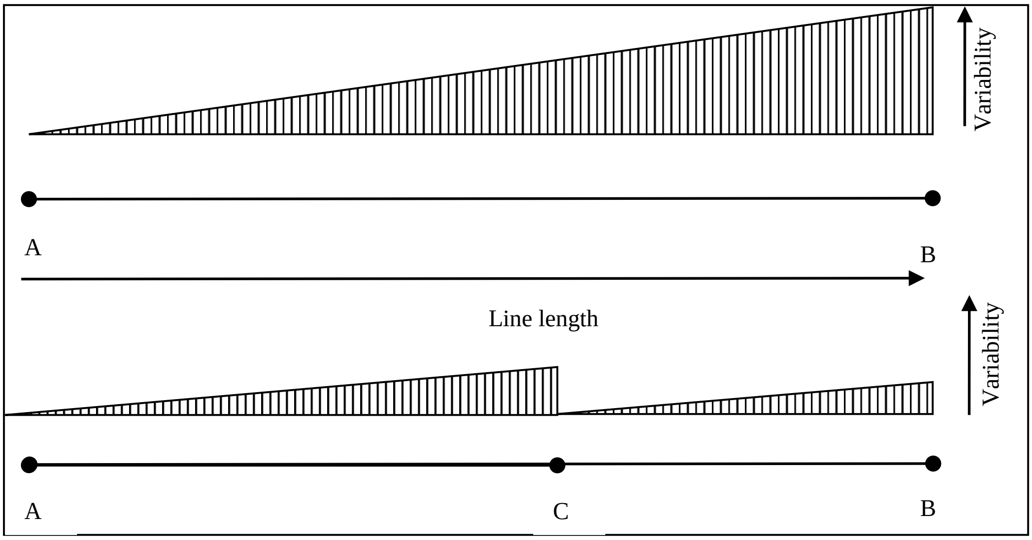

Fig. 5: The effects of splitting a line (at C) on vehicle trip time variability.

The additional transfer will lead to additional travel time for passengers passing point C and the decrease in variability will lead to better schedule adherence and shorter additional waiting times for passenger travelling between C and B.

During the design process, it is important to take both the potential time saving due to skipped transfers and the possible additional travel time due to unreliability into account while planning to extend or (dis)connect certain lines. We have analyzed this design dilemma in a quantitative way, using our framework calculating average additional travel time per passenger due to unreliability and transfers (Van Oort and Van Nes 2009). We investigated all long tram and bus lines in The Hague concerning the impacts of splitting them into two parts (at a logical stop in the city center/at railway stations). In this section, we present the general results and a cost benefit analysis.

The results of the analysis are shown in Figure 6. Per line the average additional travel time per passenger is illustrated as a result of average additional waiting time and transfer time. Both the current ("long") and the expected ("short") additional travel time when the line is divided into two short lines are shown. The effect of splitting the lines differs per line due to differences in travel patterns and punctuality characteristics. The additional waiting time resulting from transferring on line 15/16 is very small, because this line actually consists of two lines, which are connected to each other only because of operational efficiency, not to offer more direct connections. It can be seen that this efficiency measure increases the additional travel time by 50%. This effect was probably not considered when the decision was made to integrate both lines. The effect of splitting is positive for line 1 and line 17 as well; the effect is about 30% less additional waiting time. Although it was expected that splitting bus line 23 (by far the longest and least punctual in The Hague's public transport network) would decrease the additional waiting time as well, the effect of splitting is actually negative. The main reason for this is that there is no ideal transfer point on this line: the number of through passengers on the main part of the line is never below 18%.

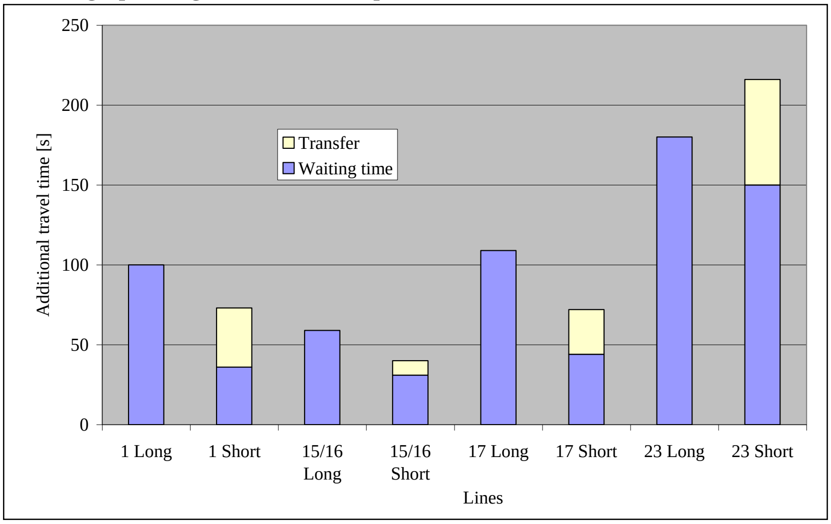

Fig. 6: Effects of splitting lines on average additional waiting time per passenger.

There are more elements than service reliability that influence the decision if a line should be cut. The (availability of) infrastructure is an important factor as well, since the introduction of shorter lines may require additional investment in for instance terminal facilities. This especially counts for rail bound public transport, as a turning loop (for one-directional-vehicles) or a cross over (for bi-directional vehicles) is required in order for trams and metros to be turned. In the latter case this could impose capacity constraints on the transport network. The turning facilities, if not already present, need to be constructed and maintained. The annual costs of this are estimated to be around €150,000. For buses the number is far less, as generally existing roads near the terminal can be used. However, these must be sufficiently equipped to be able to accommodate the vehicles. In addition to this all, the terminal could (by collective agreements) require the installation of crew facilities, such as waiting areas and toilets.

Considering the operational costs, two short lines are typically more expensive than one long one, since the total amount of layover time will be longer. This additional time resulting from the operation of two shorter lines may also require the deployment of an additional vehicle. We estimated these costs to be € 500,000 (bus) to € 1 million (tram) per year, including both capacity and operating costs (i.e. longer layover times). On the other hand, costs may be reduced as frequencies on both shorter lines can be adjusted to the demand on both parts separately instead of equal frequencies on the entire longer line.

Van Oort and Van Nes (2009) presented the trade-off between the positive and negative impacts for passengers of reducing line lengths. Shorter lines imply more reliable services and less passenger travel time on the one hand and additional transfers on the other hand. Depending on the passenger pattern and the departure punctuality characteristics, shorter lines may result in overall shorter passenger travel times and less unreliability. Above, we demonstrated cases in The Hague where shorter lines lead to overall shorter passenger travel times. The welfare gains for existing and new passengers are roughly estimated to be € 2.5 million per year. Considering the general characteristics of a passenger journey and the associated demand elasticity the decrease of total travel time of about 10% may result in about 5% more passengers and revenues per line.

5. Conclusions

Public transport operators are exposed to massive data collection from their smartcard and AVL systems. In this paper we illustrated the potential value of these data by two cases.

We developed an approach in which the smartcard matrix is assigned to the network to reproduce the measured passenger flows. Once the assignment can reproduce the passenger flows simple what-if analysis becomes possible. With the introduction of an elasticity method on the demand matrix, simple modal-split calculations are possible. This method was applied in a case study, being the tram network in The Hague. The tool turned out to be very valuable for the operator to gain insights into small changes, such as a frequency increase.

We also introduced an additional design dilemma, namely the length of line vs. reliability. Long lines offer many direct connections, thereby saving transfers. However, the variability is often negatively related to the length of a line, leading to poorer schedule adherence and additional waiting time for passengers. In this section we suggest taking into account both the positive and negative effects of extending or connecting lines. We calculated the average additional travel time per passenger due to variability and transfers based on actual trip and passenger data. A case study conducted in The Hague shows that in the case of long lines with large variability, splitting the line may result in less additional travel time due to improved service reliability. This advantage compensates for the additional time of transferring if the transfer point is well chosen. Splitting a long line into two lines with an overlap in the central part may even result in more time saving. In that case, fewer passengers will need to transfer.

6. Acknowledgements

This research is performed in cooperation with HTM, Delft University of Technology and Goudappel Coffeng mobility consultants. The authors thank their partners for their support.

7. References

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M., White, P(2004). The demand for public transport: a practical guide.

Bouman, P., E. van der Hurk, L. Kroon, T. Li, and P. Vervest (2013). Detecting activity patterns from smart card data. In 25th Benelux Conference on Artificial Intelligence (BNAIC 2013).

Brands, T., E. de Romph, T. Veitch, J. Cook (2014). Modelling public transport route choice with multiple access and egress modes. *Transportation Research Procedia*, Vol 1, pp. 12-23.

Bregman, S. (2012), Uses of Social Media in Public Transportation. Transit Cooperative Research Program (TCRP) Synthesis 99. Transportation Research Board, Washington.

Calabrese, F. (2011), Estimating origin-destination flows using mobile phone location data. In: Pervasive Computing, IEEE

Cheung, F. Implementation of Nationwide Public Transport Smart Card in the Netherlands: Cost-Benefit Analysis. Transportation Research Record, *Journal of the Transportation Research Board*, No. 1971, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 127-132.

Douglas Economics (2006). Value and Demand Effect of Rail Service Attributes. Report to RailCorp. Wellington, New Zealand.

Furth, P. G., Hemily, B., Muller, T. H. J., & Strathman, J. G. (2006), TCRP Report 113: Using Archived AVL-APC Data to Improve Transit Performance and Management. Washington, D.C.

Heuvel, J. van den, J., A. Voskamp, W. Daamen, S.P. Hoogendoorn (2015), Using Bluetooth to estimate the impact of congestion on pedestrian route choice at train stations, Traffic and Granular flow '13, M.Charaibi et. al (eds), Switzerland.

Hickman, M. (2004), Evaluating the Benefits of Bus Automatic Vehicle Location (AVL) Systems, in: D. Levinson and D. Gillen (eds.), Assessing the Benefits and Costs of Intelligent Transportation Systems, Chapter 5, Kluwer, Boston.

Neema, N., Hickman M. and Z-L. Ma (2015), Activity detection and transfer identification for public transit fare card data, *Transportation*, Vol. 42, No. 4, pp. 683-705.

Pelletier M., Trepanier M., Morency C. (2011), Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies* 19(4):557-568.

Lee, A. N. van Oort, R. van Nes (2014), Service reliability in a network context, Transportation Research Record, No. 2417, pp. 18-26.

Munizaga, M., C. Palma. Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smart card data from Santiago, Chile. *Transportation Research C*, Vol. 24, 2012, pp. 9-18.

MVA Consultancy. (2008). Valuation of Overcrowding on Rail Services. Prepared for Department for Transport.

TRB (2004). TCRP REPORT 95: Transit Scheduling and Frequency Traveler Response to Transportation System Changes. Chapter 9.

Van Oort, N. and R. van Nes (2009), Line length versus operational reliability: network design dilemma in urban public transportation, *Transportation research record*, No. 2112, pp.104-110.

Van Oort, N. (2011), Service Reliability and Urban Public Transport Design, T2011/2, TRAIL PhD Thesis Series, Delft (http://www.goudappel.nl/media/files/uploads/2011_Proefschrift_Niels_van_Oort.pdf)

Van Oort, N. (2014) Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations, Research in Transportation Economics, Volume 48, December 2014, Pages 92-100

Van Oort, N., T. Brands, E. de Romph, J.A. Flores (2015a), Unreliability effects in public transport modelling, *International Journal of Transportation*, Vol.3, No.1, pp.113-130

Van Oort, N., D. Sparing, T. Brands, R.M.P. Goverde (2015b), Data driven improvements in public transport: the Dutch example, *Public Transport*, Vol 7(3), pp.369-389.

Van Oort, N., T. Brands, E. de Romph (2015c), Short term ridership prediction in public transport by processing smart card data, *Transportation Research Record* (in press).

Wang, W., Attanucci, J.P., N.H.M. Wilson. Bus Passenger Origin-Destination Estimation and Related Analyses. *Journal of Public Transportation*, Vol. 14, No. 4, 2011, pp. 131-150.

Wardman, M. and Whelan, G. (2011). Twenty Years of Rail Crowding Valuation Studies: Evidence and Lessons from British Experience. *Transport Reviews*, 31(3), 379-398.

Wardman, M (2012). Review and meta-analysis of U.K. time elasticities of travel demand. Transportation, Vol. 39, No. 3, pp. 465-490. DOI: 10.1007/s11116-011-9369-2