

Improving predictions of public transport usage during disturbances based on smart card data

Yap, M. D.; Nijënstein, S.; van Oort, N.

DOI

[10.1016/j.tranpol.2017.10.010](https://doi.org/10.1016/j.tranpol.2017.10.010)

Publication date

2018

Document Version

Accepted author manuscript

Published in

Transport Policy

Citation (APA)

Yap, M. D., Nijënstein, S., & van Oort, N. (2018). Improving predictions of public transport usage during disturbances based on smart card data. *Transport Policy*, 61, 84-95.
<https://doi.org/10.1016/j.tranpol.2017.10.010>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

© 2018 Manuscript version made available under CC-BY-NC-ND 4.0 license
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

IMPROVING PREDICTIONS OF PUBLIC TRANSPORT USAGE DURING DISTURBANCES BASED ON SMART CARD DATA

ABSTRACT

The availability of smart card data from public transport travelling the last decades allows analyzing current and predicting future public transport usage. Public transport models are commonly applied to predict ridership due to structural network changes, using a calibrated parameter set. Predicting the impact of planned disturbances, like temporary track closures, on public transport ridership is however an unexplored area. In the Netherlands, this area becomes increasingly important, given the many track closures operators are confronted with the last and upcoming years. We investigated the passenger impact of four planned disturbances on the public transport network of The Hague, the Netherlands, by comparing predicted and realized public transport ridership using smart card data. A three-step search procedure is applied to find a parameter set resulting in higher prediction accuracy. We found that in-vehicle time in rail-replacing bus services is perceived ≈ 1.1 times more negatively compared to in-vehicle time perception in the initial tram line. Waiting time for temporary rail-replacement bus services is found to be perceived ≈ 1.3 times higher, compared to waiting time perception for regular tram and bus services. Besides, passengers do not seem to perceive the theoretical benefit of the usually higher frequency of rail-replacement bus services compared to the frequency of the replaced tram line. For the different case studies, the new parameter set results in 3% up to 13% higher prediction accuracy compared to the default parameter set. It supports public transport operators to better predict the required supply of rail-replacement services and to predict the impact on their revenues.

Keywords: disturbance, passenger, prediction, public transport, smart card

Highlights:

- Prediction of public transport ridership in case of disturbances
- Use of observed smart card data from multiple case studies to infer passenger behavior
- Development of a three-step search method to determine a better fitting model parameter set
- New parameter set improves prediction accuracy during planned disturbances up to 13%

1. INTRODUCTION

The last decade, in several cities worldwide automated fare collection (AFC) systems are introduced for the public transport system by public transport operators and authorities. For these AFC systems, passengers need to use a smart card for public transport travelling. Open systems in which passengers only need to tap-in, as well as closed systems in which both a tap-in and tap-out are required, are applied in practice. Although the main purpose of the introduction of AFC systems was to enable an easier way of revenue collection, additionally large amounts of data are generated which can be used to get more insight in passengers' travel behavior. Over the last years, data from AFC systems is used for many purposes by scientists and practitioners on a strategic, tactical and operational level (Pelletier et al. 2011). Data from AFC systems is for example used for destination inference in case of open systems with tap-in only (e.g. Trépanier et al. 2007; Nunes et al. 2016), transfer inference (e.g. Hofmann and O'Mahony 2005; Jang 2010) and journey inference to estimate origin-destination (OD) matrices (e.g. Seaborn et al. 2009; Wang et al. 2011; Munigaza and Palma 2012; Zhao et al. 2012; Gordon et al. 2013). Other studies focus on fusion of smart card data of different operators (e.g. Nijënstein and Bussink 2015) or clustering public transport stops in order to identify and classify public transport activity centers based on smart card data (Cats et al. 2015).

Next to the aforementioned studies which use smart card data to describe, analyze, cluster and visualize current travel patterns, there are also studies focusing on public transport ridership prediction based on smart card inferred travel patterns. Idris et al. (2015) developed several mode choice models based on revealed preference, contrary to traditional mode choice models having the tendency to overestimate public transport ridership. Wei and Chen (2012) developed a forecasting approach for short-term ridership predictions in metros using a combination of empirical mode decomposition and neural networks, whereas Li et al. (2017) predict metro ridership under special events using a multiscale radial basis function (MSRBF) network. Ding et al. (2016) predict metro ridership using gradient boosting decision trees, thereby incorporating temporal features and bus transfer activities. In Van Oort et al. (2015a) a smart card based prediction model is developed which allows the prediction of effects of changes in public transport supply, like increasing the frequency or rerouting public transport services. This model considers the total urban public transport network and uses an elasticity approach, where parameter values are obtained based on revealed preference studies. Also effects of crowding can be incorporated in this short-term ridership prediction model (e.g. Van Oort et al. 2015b). This type of prediction model is of added value to improve prediction accuracy of the impact of structural network changes, which are usually implemented by operators on one or on a few fixed dates in the year. However, in practice many public transport operators are confronted with temporary closures of infrastructure many more times per year. These temporary infrastructure closures are for example caused by maintenance work, track renewal or redesign of public space. These closures usually result in longer travel time, more transfers, lower rider ship, lower passenger satisfaction, and less revenues. In the Netherlands, a tendency can be observed of more, larger and more long-lasting rail infrastructure closures. For example, HTM, the urban public transport operator in Den Haag, the Netherlands, was confronted with more than 20 temporary track closures in 2015. It therefore becomes more urgent for operators to predict the impact of these (planned) disturbances on their passengers, ridership and revenues. This impact of temporary track closures on demand and supply is different compared to the impact of structural network changes. Passengers might be willing to postpone a single trip, change their mode choice or route choice, or accept the use of rail-replacement bus services for temporary situations. Operators on the other hand have to accept the temporary reduction in level of service – because of rail-replacement bus services, additional travel time and transfers – and might accept the temporary additional operational costs for these bus services and communication. It can be concluded that the responses of passengers and operators differ in case of temporary network changes, compared to structural network changes. In order to predict passenger impacts of temporary network changes with sufficient accuracy, other/additional parameters and/or different parameter values in the public transport ridership prediction models are therefore required.

This study aims to improve the prediction accuracy of the impact of planned, temporary disturbances on public transport usage. To this end, in this study a new parameter set is calibrated and

validated to predict public transport ridership in case of planned disturbances. This parameter set is based on smart card data derived from AFC systems during several planned disturbances which occurred in The Hague in 2015. The study results in a new set of parameter values allowing to better predict passenger impacts of planned disturbances in urban public transportation. With this result, more insight is gained in passenger behavior during disturbances. It also supports operators to predict the impact on their revenues, and to better align supply of rail-replacement services on alternative routes to the remaining demand, in order to efficiently use their scarce resources. This paper is structured as follows. Chapter 2 describes the methodology to calibrate and validate the parameter set of the ridership prediction model. Chapter 3 describes the case study network to which the methodology is applied. Chapter 4 discusses the results of this study. At last, in chapter 5 conclusions and recommendations for further research are formulated.

2. METHODOLOGY

2.1 Origin-destination matrix estimation

When travelling in trams or busses in the Netherlands by smart card, passengers are required to tap-in and tap-out at devices which are located within the vehicle. This means that in the Netherlands the passenger fare is based on the exact distance travelled in a specific public transport vehicle. Especially for busses, this is different from many other cities in the world where often an open, entry-only system with flat fare structure is applied, for example in London (Gordon et al. 2013) and Santiago, Chile (Munigaza and Palma 2012). This means that for each individual transaction the boarding time and location, and the alighting time and location of each trip leg are known. Also, it is known in which public transport line and vehicle each passenger boarded and alighted with its unique smart card number. This closed within-vehicle system therefore eases the destination and journey inference, compared to open entry-only systems. When merging this closed within-vehicle AFC system with Automated Vehicle Location (AVL) data, also vehicle occupancies can be inferred directly from the transaction data for each line segment and vehicle.

For an urban public transportation network with tram and bus lines, journeys can be inferred by combining registered trip legs made with the same smart card ID. In this study we used a simple temporal criterion to determine whether a passenger alighting is considered as final destination or as transfer. When the boarding time to a vehicle follows within a certain time window after the alighting time of the previous trip leg made with that same card, two AFC transactions are considered as one journey. This approach is also used, for example, by Hofmann & O'Mahony (2005) and Seaborn et al. (2009). We are aware that in scientific literature more advanced transfer inference algorithms have been developed (e.g. Zhao et al. 2007; Munigaza & Palma 2012; Gordon et al. 2013; Yap et al. 2017). In Dutch practice however, operators apply only a time window threshold between the previous alighting and next boarding as transfer inference criterion. In order to compare the prediction accuracy of the new proposed parameter set with the earlier operator predictions with the default parameter set, we decided not to adjust the transfer inference algorithm in this study. In this way, we can evaluate purely the effects of our new parameter set on the prediction accuracy, while not also changing the transfer inference algorithm simultaneously. In the Netherlands, a maximum threshold transfer time of 35 minutes is applied to classify trip legs made by the same smart card ID as one journey. By aggregating all journeys, a stop-to-stop smart card based OD matrix can be inferred. In the ridership prediction model, zones are located at the stop locations. Only stop codes which belong to the same stop from a passenger perspective, are aggregated to one zone. This means that stop codes of platforms of the same stop in opposite directions, or stops located at the same intersection, are represented by one zone. This is done to prevent passenger travel patterns to be relying too strong on the exact current stop codes of boarding and alighting in the undisturbed scenario. Under assumption that the distribution of destinations j from each origin i for non-card users is similar to the distribution of smart card users, which is in line with the assumption applied by Munigaza and Palma (2012) to correct for missing tap-outs, the zone-based OD matrix can be scaled based on the small percentage of non-card users in the Netherlands. Determination of the share of non-card users is based on passenger counts.

When travelling by train or metro in the Netherlands, there is also a closed system where transactions are required during boarding and alighting. For train and metro, devices are however located at the station gates. This means that train-train or metro-metro transfers, as well as exact chosen routes cannot

be determined directly from the data, and that trip and transfer inference algorithms are necessary to obtain these insights.

2.2 Public transport ridership prediction model

For the prediction of public transport usage in case of planned disturbances, in this study the public transport ridership prediction model as described in Van Oort et al. (2015a) is used as basis. For an urban public transportation network, let the set of public transport stops and lines be denoted by S and L respectively. Each line $l \in L$ is defined by an ordered sequence of stops $l = (s_{l,1}, s_{l,2}, \dots, s_{l,|l|})$. $L^t \in L$ and $L^b \in L$ represent the subset of tram lines and bus lines of the considered network, respectively. $|L|$ expresses the number of public transport lines in the total set L . Trip schedules are imported in the model, based on which the frequency and stop-to-stop travel times are inferred for each line $l \in L$ in each distinguished time period t . Public transport demand is connected to this network by an OD matrix between all stops $s \in S$ for each distinguished time period t . The OD matrix of the undisturbed base scenario δ_0 is based on smart card data and estimated as explained in chapter 2.1, using a conversion table between the stop ID of the boarding and alighting location in the smart card transaction data and the modelled zones in the prediction model, in order to connect travel demand to the modelled urban public transportation network.

For public transport ridership predictions, this model is based on a demand elasticity. For each OD pair i, j the generalized travel costs – being the sum of costs for in-vehicle time, transfer walking time, waiting time, transfers and travel fares with their corresponding weights – are calculated for the base scenario δ_0 and for each scenario δ . Equation 1 shows the calculation of the generalized costs, expressed in monetary terms. Applying a demand elasticity parameter to the relative change in generalized travel costs between δ_0 and δ for each OD pair allows the calculation and assignment of a new public transport OD matrix for each scenario δ . Equation 2 shows the calculation of new public transport demand.

The default parameter values for a_1, a_2, a_3, a_4, a_5 used in this prediction model for structural network changes are obtained based on a combination of model calibration and literature review (Balcombe et al. 2014). In this calibration process, model assignment results (number of passengers and passenger-distance on the network, per line $l \in L$ and per link) for the undisturbed base scenario δ_0 were compared with the raw smart card transaction data. The parameter set resulting in the highest fit between assignment results and raw smart card data, with parameter values within bounds found in literature, is applied to this model. The weight of in-vehicle time a_1 equals 1.0, whereas one minute walking time a_2 or waiting time a_3 are valued 1.5 times more negatively compared to one minute in-vehicle time. This is also in line with values found in literature (e.g. Balcombe et al. 2004; Arentze and Molin 2013). Given the focus on an urban public transport network with usually relatively short trips, a relatively small transfer penalty of 3 minutes is applied for a_4 . In this prediction model we only consider the marginal travel costs per travelled kilometer, without incorporating the base fare of €0.88 which applies for all passengers and all trips in urban public transport in the Netherlands. This is justified since this fixed cost component, which is the same for each public transport route, does not add explanatory power to passenger route choice in the model. The marginal travel costs per travelled kilometer in the model are reflected by a_5 and equal €0.05/km. Compared to the marginal travel costs of €0.15/km currently in the Netherlands (MRDH 2016), this value shows a limited price sensitivity. This can be explained due to the fact that also passengers which are price-inelastic are incorporated in the data. These passengers do not have to pay for their tickets themselves (e.g. business trips paid by the company, or student trips paid by the Dutch government), have monthly or yearly travel passes (where the marginal travel costs are usually lower), or travel with discount (e.g. elderly, children). The Value-of-Time for the Dutch situation is determined based on Significance et al. (2013).

In the model used in this study no segmentation in parameter values is used for passengers with different trip purposes or for different time periods. Differences in sensitivity to generalized cost components can for example be related to trip purpose or socio-economic status of passengers. In the model used for this study, no explicit distinction is made between the home-end and activity-end of a journey. Therefore, it is not possible to determine the home-end of each journey explicitly, and to relate this for example to the socio-economic status of a specific area of the city. Also regarding trip purpose an explicit

segmentation cannot directly be inferred from the AFC data. For example, in The Hague there are 58 different product types available for travelling by smart card (e.g. business card, student cards, monthly cards). For some card types it is relatively easy to match card type (e.g. student card) and trip purpose (e.g. education). However, for many other card types this match is less trivial. Besides, for this study no information was available regarding the product type used for each AFC transaction. Also, our AFC dataset did not contain the ID of the smart card used for travelling. Therefore, it was not possible to infer trip purpose from long-term travel patterns. Different sensitivities during different periods of the day are mostly related to different mixtures of passenger segments (e.g. a high percentage commuters during peak periods, or a high percentage leisure / shopping passengers during off-peak periods), as for example already shown by Nazem et al. (2011). Instead of aiming to infer these different travel segments from available travel patterns, we decided to come up with a robust parameter set which is – on average – able to improve prediction accuracy over these different passenger segments. For further research, there is definitely potential to further improve the parameter set specific for different passenger segments and areas of the city.

$$C_{ij} = (\alpha_1 IVT_{ij} + \alpha_2 WKT_{ij} + \alpha_3 WTT_{ij} + \alpha_4 NT_{ij}) * VoT + \alpha_5 d_{ij} \quad (1)$$

With:

C_{ij}	Generalized costs on OD pair i,j
$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$	Weight coefficients in generalized costs calculation
IVT_{ij}	In-vehicle travel time on OD pair i,j
WKT_{ij}	Walking time on OD pair i,j
WTT_{ij}	Waiting time on OD pair i,j
NT_{ij}	Number of transfers on OD pair i,j
VoT	Value-of-Time (€/hour)
d_{ij}	Distance travelled in public transport on OD pair i,j

$$D_{ij}^{\delta} = \left(E \left(\frac{C_{ij}^{\delta}}{C_{ij}^{\delta_0}} - 1 \right) + 1 \right) * D_{ij}^{\delta_0} \quad (2)$$

With:

D_{ij}^{δ}	Demand on OD pair i,j in scenario δ
E	Elasticity
C_{ij}^{δ}	Generalized costs in scenario δ
$C_{ij}^{\delta_0}$	Generalized costs in base scenario δ_0
$D_{ij}^{\delta_0}$	Demand on OD pair i,j in base scenario δ_0

We can thus conclude that there is already a calibrated parameter set which is used to predict public transport ridership for undisturbed situations. In this study, we specifically investigate to what extent this parameter set needs to be adjusted to perform accurate passenger predictions in case of planned disturbances.

2.3 Evaluation framework

An evaluation framework is developed to evaluate the accuracy of different parameter sets for ridership predictions in case of (planned) disturbances. First, the subset $L^a \in L$ is determined, where L^a contains all public transport lines which are directly or indirectly affected by a certain disturbance δ . Public transport lines of which the route, mode and/or frequency is adjusted due to a disturbance, are considered directly affected lines. Public transport lines parallel to the directly affected lines, which can function as alternative for passengers to reach their destination, are considered indirectly affected lines. Although route and frequency of these lines are not changed, the number of passengers and passenger-kilometers travelled over

these lines can be affected due to the directly disturbed public transport lines. For all lines $l \in L^a$ we consider the impact of a disturbance for the different distinguished time periods t . Since we do not estimate different parameter sets for different time periods, it is important to explicitly evaluate and compare the prediction accuracy of different parameter sets over the different time periods in order to develop a robust parameter set.

We used two indicators to express the impact of a disturbance: the relative impact of a disturbance on the number of passengers P and on the number of passenger-kilometers PK . This means we consider the relative increase or decrease in P and PK for all affected public transport lines in all distinguished time periods during a specific disturbance δ , compared to the undisturbed base scenario δ_0 . For each disturbance, this leads to a total number of $|L^a| * |T| * 2$ elements for which the empirical and predicted relative effect of a disturbance are compared. Equations (3) and (4) formalize the calculation of the relative impact of a disturbance on the number of passengers P and passenger-kilometers PK , respectively.

$$\Delta P = \left(\frac{P_{\delta,lt} - P_{\delta_0,lt}}{P_{\delta_0,lt}} \right) * 100 \quad \forall l \in L^a \quad \forall t \quad (3)$$

$$\Delta PK = \left(\frac{PK_{\delta,lt} - PK_{\delta_0,lt}}{PK_{\delta_0,lt}} \right) * 100 \quad \forall l \in L^a \quad \forall t \quad (4)$$

With:

$P_{\delta,lt}$	Number of passengers in disturbed scenario δ
$P_{\delta_0,lt}$	Number of passengers in undisturbed base scenario δ_0
$PK_{\delta,lt}$	Number of passenger-kilometers in disturbed scenario δ
$PK_{\delta_0,lt}$	Number of passenger-kilometers in undisturbed base scenario δ_0

In this study we use the well-known R^2 measure to quantify the prediction accuracy of different parameter sets for all $|L^a| * |T| * 2$ elements, which compares empirically derived values $\widehat{\Delta P}_{lt}$ and $\widehat{\Delta PK}_{lt}$ with predicted values $\Delta \bar{P}_{lt}$ and $\Delta \bar{PK}_{lt}$. The used prediction model consists of an undisturbed base scenario δ_0 , of which the number of passengers and passenger-kilometers are calibrated based on imported smart card data corresponding to this undisturbed base network (chapter 2.2). The passenger impact of disturbed scenarios $\delta \in \Delta$ is predicted using the described elasticity approach after modelling the network corresponding to each scenario δ . Based on the model output for $P(K)_{\delta_0}$ and $P(K)_{\delta}$, predicted values $\Delta \bar{P}_{lt}$ and $\Delta \bar{PK}_{lt}$ are calculated. The values $\widehat{\Delta P}_{lt}$ and $\widehat{\Delta PK}_{lt}$ can be inferred from the AFC data. For both the undisturbed base scenario δ_0 and each disturbed scenario $\delta \in \Delta$, the raw smart card data is scaled for non-card users, thereby applying the same scaling factor as applied in the prediction model. Since the period of the year which is used to infer P_{δ_0} and PK_{δ_0} differs from the period which is used to infer P_{δ} and PK_{δ} , a correction for possible ridership variation which cannot be attributed to the adjusted public transport network needs to be applied. There can be systematic differences in travel patterns between the days used for δ_0 and δ , for example due to holiday periods. Also seasonal differences can occur (e.g. different public transport ridership in January compared to June), as well as daily differences due to weather (rain or sunny weather), day of the week and random fluctuations. To prevent a biased comparison between $\Delta \bar{P(K)}_{lt}$ and $\widehat{\Delta P(K)}_{lt}$, we made sure that the generic travel patterns are similar for δ_0 and δ . This means that if δ occurs within the summer holiday, we also inferred data from the (undisturbed part of this) summer holiday for δ_0 . Similarly, in case δ occurs during regular working periods, δ_0 is also based on days during a working period. We also made sure that the share of the different days of the week was similar for δ_0 and δ – by using AFC data from an equal number of Mondays, Tuesdays etc. for δ_0 and δ – to prevent bias due to differences in ridership over the different days of the week. This however does not exclude seasonal and daily random variation from the comparison yet. This is especially relevant, given the large number of disturbances occurring on the urban tram network of operators in the Netherlands. As illustration: for the tram network of The Hague, only 6 months out of the 36 months between July 2014 and July 2017 were

totally free from planned disturbances. For example, if δ lasted during the whole month of November, there is no other option than using AFC data from one of these undisturbed months (e.g. March) for δ_0 . In order to correct for possible seasonal and random variation between these periods, we applied a correction to the inferred values P_{δ_0} and PK_{δ_0} based on the ridership differences found on all public transport lines $l \notin L^a$ between δ_0 and δ . We calculated the average difference for P_t and PK_t over all public transport lines which are not directly, nor indirectly, affected by a disturbance δ , for each distinguished time period. Under the assumption that the seasonal and random effect for all lines $l \in L^a$ is similar to the average effect found for all lines $l \notin L^a$, we corrected the originally directly inferred values $\widetilde{P}_{lt, \delta_0}'$ and $\widetilde{PK}_{lt, \delta_0}'$ using equation (5). This allowed us to compute $\Delta \widetilde{P}_{lt}$ and $\Delta \widetilde{PK}_{lt}$ without bias using equation (3) and equation (4), based on which the R^2 measure could be computed using equation (6).

$$\widetilde{P(K)}_{lt, \delta_0} = \widetilde{P(K)}_{lt, \delta_0}' * \left(\frac{1}{|L| \notin L^a} \right) \sum_1^{L \notin L^a} \frac{\widetilde{P(K)}_{lt \notin L^a, \delta}}{\widetilde{P(K)}_{lt \notin L^a, \delta_0}} \quad \forall l \in L^a \quad \forall t \quad (5)$$

$$R^2 = 1 - \frac{\sum_t^T \sum_l^{L \in L^a} (\Delta \widetilde{P(K)}_{lt} - \Delta \widetilde{P(K)}_{lt})^2}{\sum_t^T \sum_l^{L \in L^a} (\Delta \widetilde{P(K)}_{lt} - \Delta \widetilde{P(K)})^2} \quad \forall l \in L^a \quad \forall t \quad (6)$$

2.4 Experimental design

In this study four planned disturbances which occurred on the HTM network in 2015 are considered, denoted by the total set Δ . Subsets $\Delta_A \in \Delta \{\delta_1, \delta_2\}$ and $\Delta_B \in \Delta \{\delta_3, \delta_4\}$ are defined, which both contain 50% of the investigated disturbances for calibration and validation purposes, respectively. For the calibration phase, a rule-based three-step procedure is used to determine an improved parameter set which leads to a better fit between empirical AFC data and predicted public transport ridership during planned temporary disturbances.

1. Selection of parameters with potentially different values during disturbances

In order to predict public transport usage in case of planned disturbances, it is important to determine which parameter values could be different, compared to the values used to predict regular passenger route choice and ridership for undisturbed scenarios as described in chapter 2.2. In theory, values can be different for all 6 parameters of equation (1) - $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, VoT$ - and the elasticity parameter E of equation (2). In a first step narrowing down the solution space, we determine which parameters can have potentially different values during disturbances based on theory. It is confirmed using a sensitivity analysis that parameters excluded from the solution space based on theory, indeed do not or hardly affect prediction accuracy.

First, the value of the elasticity parameter E_δ in case of disturbances is of relevance. As mentioned in chapter 1, passengers react differently to temporary network changes compared to structural network changes. On the one hand, passengers might accept a longer travel time for a certain amount of time (indicating a less negative value of E_δ). On the other hand, passengers might decide to change their mode choice or destination choice, or to postpone their trip in case of temporary track closures, until regular operations are restored (indicating a more negative value of E_δ). Second, the modelling of rail-replacement services is of relevance. Let $L^R \in L$ be the subset of rail-replacing bus services. In many cases, operators will supply rail-replacing bus services in case of track closures. These rail-replacing services differ from regular bus lines in several ways. For example, the existence, route and stop locations of such services are often less well known by passengers. Given the temporary existence of these lines, passengers are less familiar with aspects as departure time, travel time and reliability. When these busses replace rail services, these services have to use temporary stop locations nearby the closed rail stop, which often have less visibility and equipment like dynamic arrival information or shelters. It is therefore possible that passengers experience waiting time for a rail-replacement services more negatively compared to waiting time for regular tram or bus services (indicating a higher value of parameter a_3 , related to waiting time WTT^R specific for rail-replacement services). Besides, these services transport passengers who are familiar with rail-bound services. From literature it is known that when a bus service is transformed into a tram line, travel time is perceived less negatively compared to bus travelling (Bunschoten et al. 2013). Therefore, it

can be hypothesized that the replacement of a tram line by busses will be perceived more negatively by passengers familiar with rail-bound travelling. Therefore, the value of parameter a_1 related to in-vehicle time perception in rail-replacement busses IVT^R might be more negative compared to regular trams or busses. Rail-replacement busses usually operate with higher frequencies than the original tram line, to compensate for the lower capacity of a bus compared to a tram. However, it is unclear to what extent passengers really perceive and incorporate this theoretical benefit in their route and mode choice. It is therefore questionable whether modelling the realized frequencies of the rail-replacement services f^R , or the original frequencies of the tram line which is being replaced f^T , leads to more accurate predictions.

The remaining parameters from Equation (1) - a_2 (multiplier for walking time perception), a_4 (fixed transfer penalty), a_5 (marginal travel costs) and VoT (value of time) - were expected to have no or a limited effect on the prediction accuracy. There is no rationale to assume that the marginal travel costs or Value of Time differ during disturbances, since the fare system applied during disturbances for rail-replacement bus services is exactly equal to the fare system used for regular tram and bus lines. Although it could be hypothesized that the walking time or a transfer to rail-replacement bus services can be perceived more negatively compared to regular tram or bus lines, a first sensitivity analysis showed only a very limited impact on the prediction accuracy using equation (6) when varying parameter values for a_2 and a_4 : increasing / decreasing parameter values of a_2 and a_4 with 33% for $\Delta_A \in \Delta \{\delta_1, \delta_2\}$, did not increase/decrease the prediction accuracy with more than 3% (Figure 1). Table 1 shows the remaining parameter values which are included in the search procedure, together with the expected direction of the parameter values compared to the default parameter value used for predictions for undisturbed network changes.

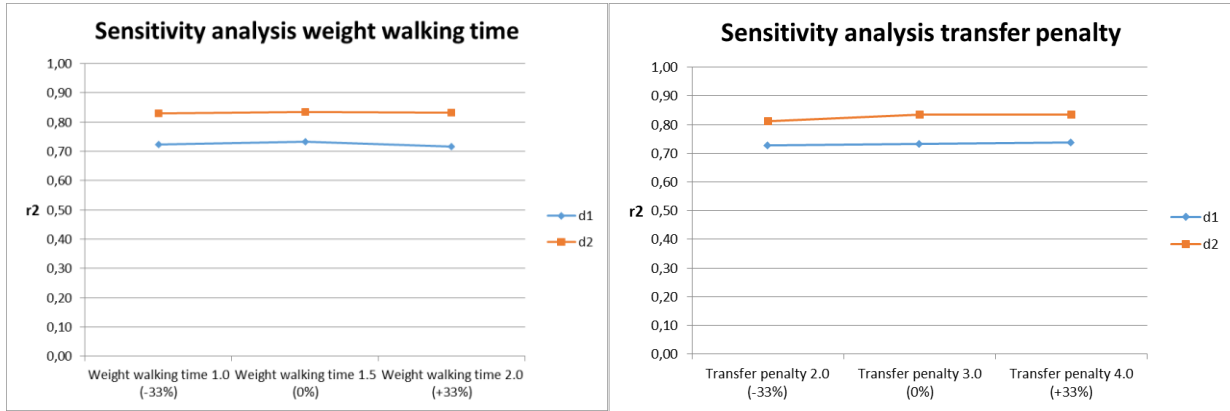


FIGURE 1 Results sensitivity analysis to walking time perception (left) and transfer penalty (right) to rail-replacement bus services

TABLE 1 Parameter Selection For Systematic Grid-Search

Parameters	Values default parameter set	Search direction parameter values
Elasticity E_δ	-1.1	More or less negative
Waiting time WTT^R	1.5	Higher
In-vehicle time IVT^R	1.0	Higher
Frequency $\{f^R, f^T\}$	f^R	f^T

2. Systematic grid-search

In the second step, a systematic grid-search is performed to scan for the best fitting parameter set(s) from predefined scenarios, using the four model parameters of Table 1. For all four parameters, plausible values are *a priori* determined. The calibrated parameter values used for passenger assignment for the undisturbed base scenario δ_0 are used as starting point ($E_\delta = -1.1$, $WTT^R = 1.5$, $IVT^R = 1.0$, $f^R = f^T$). These values are considered as reasonable starting point, since these values are calibrated and within bounds found in

literature (e.g. Balcombe et al. 2004). The direction in which each parameter value can change when predicting ridership during disturbances, compared to regular ridership predictions, is explained in the first part of this chapter 2.4 and shown in Table 1. The upper and lower bound values for E_δ , WTT^R and IVT^R are selected in such way, that they remain within literature bounds on one hand, but show sufficient variation to explore the solution space on the other hand. The modelling of the frequency of rail-replacement bus services is a binary variable, which can be equal to f^R or f^T . The second row of Table 2 shows the resulting parameter values. All combinations between these predefined parameter values are systematically evaluated using the R^2 measure for prediction accuracy from equation (6). Using the Cartesian product of the chosen plausible values, this results in 24 scenario combinations of parameter values (Table 2). The first row of Table 2 summarizes the four parameters which are hypothesized to have different values when modelling passenger behavior during disturbances specifically.

TABLE 2 Experimental Design

Parameters	Elasticity E_δ	Waiting time WTT^R	In-vehicle time IVT^R	Frequency
Parameter values	$\{-0.7, -1.1, -1.5\}$	$\{1.5, 2.0\}$	$\{1.0, 1.25\}$	$\{f^R, f^T\}$
Scenario 1 (default)	-1.1	1.5	1.0	f^R
Scenario 2	-1.1	1.5	1.0	$\text{MIN}(f^R, f^T)$
...	...	...	...	...

Based on the results of this second step the solution space can be narrowed down further, so that a further in-depth search can be performed in the third step of the search procedure. In this second step we did not only look at the exact R^2 value resulting from the 24 different scenarios, but we also investigated which patterns and scenarios were robust over the two evaluated disturbances $\Delta_A \in \Delta \{\delta_1, \delta_2\}$. Since we aim to come up with one improved parameter set to predict impacts of different (types of) disturbances, it is important to have a robust, good performing parameter set rather than optimizing too much on these specific two disturbances. We determine the most promising parameter sets out of these pre-defined scenarios as result of this second step of the procedure.

3. Specific in-depth search

In the third step of the calibration the parameter set is further improved based on the boundaries set in the second step of the procedure, using a specific in-depth search procedure. In this step, parameter values are not bound to the predefined values and scenarios anymore. Within the boundaries set in step 2, we evaluated all combinations of parameter values (with a certain minimum step size applied between the candidate parameter values). Since all these candidate parameter sets are within bounds of the promising, robust parameter sets found in step 2, we investigated which parameter set resulted in the highest prediction accuracy measured by the R^2 measure and selected this parameter set as new, preferred set. Once this parameter set is determined, this set is validated by applying it to the investigated disturbances $\delta \in \Delta_B$. For this subset Δ_B it is tested whether the prediction accuracy improved compared to the default parameter set. Note that we did not apply a full optimization in this process of improving the parameter set. We used a rule-based approach to narrow down the solution space in a three-step procedure, meaning that not all possible combinations of parameter values are evaluated. Optimality can thus not be proved in this method.

3. CASE STUDY

The methodology as described in chapter 2 is applied in a case study. The urban public transport network of The Hague, the Netherlands, is used in this study. Public transport services on this network are operated by HTM. The network consists of 12 tram lines and 8 bus lines. No metro services are operated in the city of The Hague. Two of the tram lines function as light rail connection between The Hague and the nearby suburb of Zoetermeer. On an average working day, more than 250,000 trips are made on the HTM network. 93% of the passengers use a smart card for travelling (HTM 2015). The remaining 7% buys a ticket from the driver or at the vending machine, or uses a special ticket. When modelling the HTM network, 4 different

time periods are distinguished in the frequency-based assignment and prediction model: morning peak (7am-9am), evening peak (4pm-6pm), off-peak (9am-4pm) and the evening and early morning (6pm-7am).

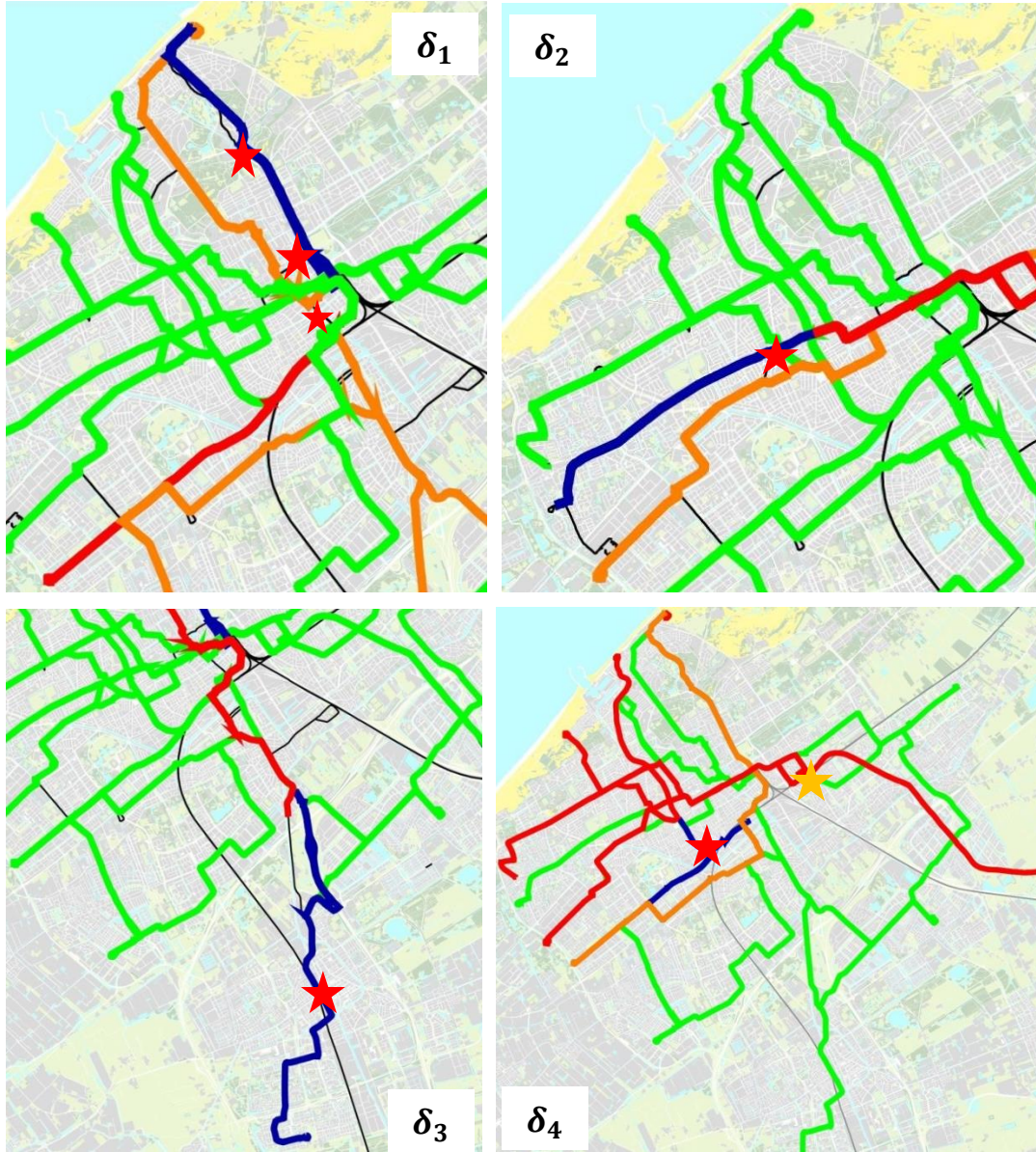


FIGURE 2 Public transport network during planned disturbances δ_1 ‘Koninginnegracht’ (upper left), δ_2 ‘Loosduinseweg’ (upper right), δ_3 ‘Westvest’ (lower left), δ_4 ‘Parallelweg +Ternoot’ (lower right) (star: work location / orange: line rerouted / red: line shortened / blue: rail-replacement bus)

In 2015 there were several track closures on the public transport network operated by HTM. Given the closed AFC system, in combination with relatively many case studies available, the HTM network is an interesting case study area to investigate the impact of planned disturbances on public transport ridership. As explained, in total 4 different disturbances δ which occurred in 2015 on the HTM network are investigated, which are divided into two subsets $\Delta_A \in \Delta \{\delta_1, \delta_2\}$ and $\Delta_B \in \Delta \{\delta_3, \delta_4\}$ used for calibration and validation purposes, respectively. Table 3 describes the impact of each disturbance on the public transport network, as well as the period of the year the AFC data is coming from for δ_0 and δ . Figure 2 shows the adjusted public transport network for all four disturbances. Closure δ_1 ‘Koninginnegracht’ resulted in

detours for several tram lines in the city center. Besides, one of the two important connections between Central Station and Scheveningen (tram line 9) was replaced by bus services of line 69 (whole day) and 79 (only peak hours). Closure δ_2 ‘Loosduinseweg’ resulted in the shortening of two busy tram lines 2 and 4. The shortened part of the route of tram line 2 was replaced by busses (line 52). Most stops of the shortened tram line 4 were covered by tram line 6, which route was temporarily extended over the unaffected part of the shortened route of tram line 4. During closure δ_3 ‘Westvest’, the route of tram line 1 – connecting the city of Den Haag with the city of Delft – was shortened. A rail-replacement bus line 71 was provided, although it could not stop near all original tram stops due to infrastructure constraints. Closure δ_4 ‘Parallelweg+Ternoot’ reflects two different disturbances which occurred simultaneously on the network. The Parallelweg closure affected the route of tram lines 9, 11 and 12 through an area with a relatively high public transport dependency. Lines 11 and 12 – which share the route through this area – were shortened, whereas tram line 9 was rerouted. The closed track of line 9 was replaced by temporary bus line 58, whereas bus line 82 replaced the closed part of the route of tram lines 11 and 12. The closure Ternoot resulted in the cutting of both light rail services 3 and 4 in two separate parts. Passengers could walk between the short-turning terminals of both parts of these lines.

It can be concluded that the set of disturbances Δ can roughly be divided in closures in which tram lines are rerouted (δ_1, δ_4), closures in which tram lines are shortened and replaced by bus services ($\delta_1, \delta_2, \delta_3, \delta_4$), and closures in which tram lines are divided in two separate parts (δ_4). To investigate and test that the selected parameter set is robust to perform accurate predictions for different types of closures, we aimed to have a mixture of different types of closures in both the subset used for calibration $\Delta_A \in \Delta \{\delta_1, \delta_2\}$, as well as in the subset used for validation $\Delta_B \in \Delta \{\delta_3, \delta_4\}$.

For the reference network δ_0 used for closures δ_1 and δ_3 , as well as for disturbed network δ_1 , 20 working days of smart card data is used in this study. Given the ≈ 250.000 journeys (≈ 300.000 AFC transactions) at the HTM network per average working day, this roughly means that about 6 million smart card transactions are used as basis for these analyses. For the shorter lasting disturbances $\delta_2, \delta_3, \delta_4$ and the reference network δ_0 used for closures δ_2 and δ_4 , within the summer holiday, about 1.5 million smart card transactions (5 working days) are used (Table 3). All raw transactions are anonymized by removing personal information and by aggregating the data, to guarantee confidentiality and to obey Dutch privacy regulations.

TABLE 3 Overview Of Network Changes During Planned Disturbances In 2015

Disturbance δ	Period δ (# days)	Period δ_0 (# days)	Directly affected lines $l \in L^a$	Rail-replacement line (replaced tram)
δ_1 Closure ‘Koninginnegracht’	November (20)	March (20)	Tram 1/15/16/17: rerouted Tram 9: shortened + bus-replacement	Bus lines 69+79 (9)
δ_2 Closure ‘Loosduinseweg’	August (5)	August (5)	Tram 2: shortened + bus-replacement Tram 4: shortened Tram 6: extended (to replace tram 4)	Bus line 52 (2)
δ_3 Closure ‘Westvest’	October (5)	March (20)	Tram 1: shortened + bus-replacement	Bus line 71 (1)
δ_4 Closure ‘Parallelweg+Ternoot’	July (5)	August (5)	Tram 3/4: cut into two separate parts Tram 9: rerouted + bus-replacement Tram 11/12: shortened + bus-replacement	Bus line 58 (9) Bus line 82 (11+12)

4. RESULTS

4.1 Results rule-based search procedure

After the first selection of parameters with potentially different values during disturbances (step 1 of the three-step search procedure as described in chapter 2.4), this chapter presents results of the second step (systematic grid-search) and the third step (specific in-depth search) of the method. Table 4 and Figure 3 show the resulting prediction accuracy for the two disturbances δ_1 and δ_2 used for calibration for all 24

predefined scenarios (as explained in chapter 2.4 and Table 2). In this step we specifically search for patterns and robust parameter values. Based on Table 4 and Figure 3, we conclude the following:

- Regardless the value of E_δ and the modelled frequency f , scenarios with $WTT^R = 1.5$ and $IVT^R = 1.0$ are outperformed by the other scenarios. From the resulting prediction accuracy for δ_1 we see that scenario 1+2, 9+10 and 17+18 clearly have a lower R^2 than scenarios 3-8, 11-16 and 19-24 respectively with the same value for E_δ . This indicates that this combination of values for waiting time and in-vehicle time does not put sufficient perceived disutility on usage of rail-replacement busses. This means that scenarios 1, 2, 9, 10, 17 and 18 are removed from the candidate list.
- Scenarios with $E_\delta = -1.5$ are outperformed by scenarios using $E_\delta = -1.1$ or $E_\delta = -0.7$. From the resulting prediction accuracy for δ_1 we can see that for each given combination of WTT^R , IVT^R and f , scenarios 19-24 have a lower R^2 than their corresponding scenarios 3-8 and 11-16 with a different value for E_δ . This indicates that ridership loss due to temporal planned disturbances in general is not larger compared to the assumed ridership loss for structural network changes. This means that scenarios 19, 20, 21, 22, 23 and 24 are removed from the candidate list.
- Scenarios in which the realized frequency of rail-replacement busses f^R is modelled are outperformed by scenarios in which the frequency of the original tram lines which is being replaced f^T is modelled. From the resulting prediction accuracy for δ_2 we can see that for each given combination of E_δ , WTT^R and IVT^R the scenarios 4, 6, 8, 12, 14, 16 using f^T clearly outperform their corresponding scenario 3, 5, 7, 11, 13, 15 respectively using f^R . This indicates that using the frequency of the original tram line is a better ridership predictor than using the realized frequency of rail-replacement busses, and that scenarios 3, 5, 7, 11, 13 and 15 are removed from the candidate list.
- For the remaining scenarios 4, 6, 8, 12, 14 and 16, the prediction accuracy is ranked separately for the prediction performance for δ_1 and δ_2 . Scenarios 4 and 6 perform very well for δ_2 , but perform mediocre for δ_1 . Scenario 8 performs medium-high for both δ_1 and δ_2 . Scenarios 12, 14 and 16 perform medium-high to high for δ_2 , and perform medium-high to very high for δ_1 . Since our aim is to use a robust parameter set, scenarios 12, 14 and 16 are used as result from this second step of the search procedure, to apply the specific in-depth search to in the third step of the method.

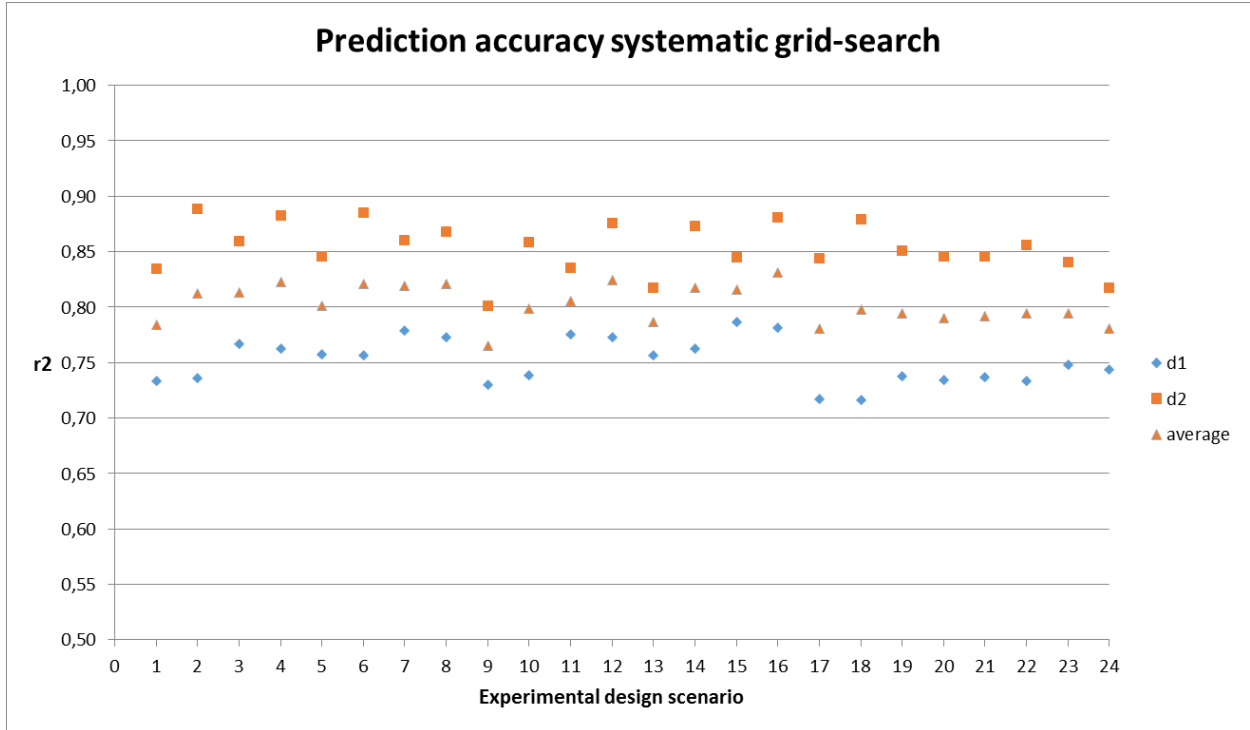


FIGURE 3 Results prediction accuracy systematic grid-search procedure**TABLE 4 Overview Of Prediction Accuracy For All 24 Scenarios**

Parameters	Elasticity E_δ	Waiting time WTT^R	In-vehicle time IVT^R	Frequency	R^2 δ_1	R^2 δ_2	R^2 average
Parameter values	$\{-0.7, -1.1, -1.5\}$	$\{1.5, 2.0\}$	$\{1.0, 1.25\}$	$\{f^R, f^T\}$			
Scenario 1.1 (default)	-1.1	1.5	1.0	f^R	0.73	0.83	0.78
Scenario 1.2	-1.1	1.5	1.0	$\text{MIN}(f^R, f^T)$	0.74	0.89	0.81
Scenario 1.3	-1.1	1.5	1.25	f^R	0.77	0.86	0.81
Scenario 1.4	-1.1	1.5	1.25	$\text{MIN}(f^R, f^T)$	0.76	0.88	0.82
Scenario 1.5	-1.1	2.0	1.0	f^R	0.76	0.85	0.80
Scenario 1.6	-1.1	2.0	1.0	$\text{MIN}(f^R, f^T)$	0.76	0.88	0.82
Scenario 1.7	-1.1	2.0	1.25	f^R	0.78	0.86	0.82
Scenario 1.8	-1.1	2.0	1.25	$\text{MIN}(f^R, f^T)$	0.77	0.87	0.82
Scenario 1.9	-0.7	1.5	1.0	f^R	0.73	0.80	0.77
Scenario 1.10	-0.7	1.5	1.0	$\text{MIN}(f^R, f^T)$	0.74	0.86	0.80
Scenario 1.11	-0.7	1.5	1.25	f^R	0.78	0.84	0.81
Scenario 1.12	-0.7	1.5	1.25	$\text{MIN}(f^R, f^T)$	0.77	0.88	0.82
Scenario 1.13	-0.7	2.0	1.0	f^R	0.76	0.82	0.79
Scenario 1.14	-0.7	2.0	1.0	$\text{MIN}(f^R, f^T)$	0.76	0.87	0.82
Scenario 1.15	-0.7	2.0	1.25	f^R	0.79	0.85	0.82
Scenario 1.16	-0.7	2.0	1.25	$\text{MIN}(f^R, f^T)$	0.78	0.88	0.83
Scenario 1.17	-1.5	1.5	1.0	f^R	0.72	0.84	0.78
Scenario 1.18	-1.5	1.5	1.0	$\text{MIN}(f^R, f^T)$	0.72	0.88	0.80
Scenario 1.19	-1.5	1.5	1.25	f^R	0.74	0.85	0.79
Scenario 1.20	-1.5	1.5	1.25	$\text{MIN}(f^R, f^T)$	0.73	0.85	0.79
Scenario 1.21	-1.5	2.0	1.0	f^R	0.74	0.85	0.79
Scenario 1.22	-1.5	2.0	1.0	$\text{MIN}(f^R, f^T)$	0.73	0.86	0.79
Scenario 1.23	-1.5	2.0	1.25	f^R	0.75	0.84	0.79
Scenario 1.24	-1.5	2.0	1.25	$\text{MIN}(f^R, f^T)$	0.74	0.82	0.78

Scenarios 12, 14 and 16 have the same value for E_δ (-0.7) and for f (f^T). This means we fix these two parameter values. These scenarios differ in the parameter values used for WTT^R and IVT^R , which range between 1.5-2.0 and 1.0-1.25, respectively. This means the waiting time perception for rail-replacement busses varies between the same value as for regular tram and bus lines, and a 33% higher value. The in-vehicle time perception of 1 minute in a rail-replacement bus varies between 0.8 and 1 minute in the original tram line. In this in-depth search we explore all different combinations of these parameter values, increasing parameter values with a step size of $\approx 10\%$. This leads to values for waiting time perception $\{1.5, 1.67, 1.84, 2.0\}$ and in-vehicle time perception in the original tram line for 1 minute travelling in the rail-replacement bus $\{0.8, 0.9, 1.0\}$. Inversed this leads to in-vehicle time perception in the rail-replacement bus for 1 minute travelling with the original tram line $\{1.0, 1.11, 1.25\}$. These 3×4 combinations are evaluated, of which is known from step 2 that the combination of using a waiting time perception of 1.5 with 1.0 minute in-vehicle time perception simultaneously (scenario 2.9) does not provide a good prediction accuracy. Table 5 shows these 12 scenarios with the resulting R^2 averaged over δ_1 and δ_2 .

From Table 5 we can conclude that the averaged prediction accuracy is highest for scenarios 2, 3, 4 and 8. To select a final parameter set, we applied these four parameter sets to the subset of disturbances used for validation $\Delta_B \in \Delta \setminus \{\delta_3, \delta_4\}$ as well. The aim was to investigate which parameter set resulted in the highest prediction accuracy for Δ_B , in order to select the most robust parameter set. This resulted in the selection of scenario 8, in which $WTT^R = 2.0$ and $IVT^R = 1.11$.

TABLE 5 Overview Of Prediction Accuracy For In-Depth Search

Parameters	Elasticity E_δ	Waiting time WTT^R	In-vehicle time IVT^R	Frequency	R^2 average
Parameter values	$\{-0.7\}$	$\{1.5, 1.67, 1.84, 2.0\}$	$\{1.0, 1.11, 1.25\}$	$\{f^T\}$	
Scenario 2.1	-0.7	1.5	1.25	$\text{MIN}(f^R, f^T)$	0.82
Parameters	Elasticity E_δ	Waiting time WTT^R	In-vehicle time IVT^R	Frequency	R^2 average
Scenario 2.2	-0.7	1.67	1.25	$\text{MIN}(f^R, f^T)$	0.83
Scenario 2.3	-0.7	1.84	1.25	$\text{MIN}(f^R, f^T)$	0.83
Scenario 2.4	-0.7	2.0	1.25	$\text{MIN}(f^R, f^T)$	0.83
Scenario 2.5	-0.7	1.5	1.11	$\text{MIN}(f^R, f^T)$	0.81
Scenario 2.6	-0.7	1.67	1.11	$\text{MIN}(f^R, f^T)$	0.82
Scenario 2.7	-0.7	1.84	1.11	$\text{MIN}(f^R, f^T)$	0.82
Scenario 2.8	-0.7	2.0	1.11	$\text{MIN}(f^R, f^T)$	0.83
Scenario 2.9	-0.7	1.5	1.0	$\text{MIN}(f^R, f^T)$	0.80
Scenario 2.10	-0.7	1.67	1.0	$\text{MIN}(f^R, f^T)$	0.81
Scenario 2.11	-0.7	1.84	1.0	$\text{MIN}(f^R, f^T)$	0.81
Scenario 2.12	-0.7	2.0	1.0	$\text{MIN}(f^R, f^T)$	0.82

4.2 Resulting parameter set

Table 6 shows the values for the proposed new parameter set, compared to the parameter values of the default set. In case of planned disturbances, an elasticity parameter value E_δ of -0.7 resulted in the best fit with the raw smart card data. The hypothesis that passengers might perceive waiting time for a rail-replacement service more negatively compared to waiting time for regular tram and bus services was confirmed, since applying a higher waiting time coefficient did improve the prediction accuracy. Therefore, the use of a specific waiting time coefficient of 2.0 for rail-replacement services is proposed. Also, applying a more negative in-vehicle time perception in bus services replacing an existing tram line did improve the prediction accuracy. In the used prediction model in this study, this is reflected by applying a certain multiplication factor for the operational speed of a rail-replacement service. Using a speed factor of 0.9 – which equals the inverse in-vehicle time coefficient of 1.11 – resulted in the best fit with the raw smart card data. The value for the in-vehicle time coefficient derived from realization data is somewhat lower than the speed factor found in Bunschoten et al. (2013) using a stated preference experiment, but points towards the same direction. Regarding the frequency of rail-replacement services, study results indicate that modelling the frequency of the original tram line f^T leads to a better fit than using f^R , if $f^R > f^T$. This indicates that passengers do not incorporate the benefit of the higher frequency of the rail-replacement services compared to the original tram line in their route and mode choice. From a theoretical perspective, this can be explained because vehicle capacity is not incorporated in the prediction model used in this study. The higher frequency f^R compared to f^T is often due to the lower capacity of a bus compared to a tram vehicle. Since the negative effect of a lower bus capacity is not incorporated in the model, only incorporating the positive effect of a higher bus frequency aimed to compensate for this non-incorporated capacity effect would overestimate the level of service of the rail-replacement bus service. In case $f^R < f^T$, one could apply f^R in the prediction model. In this case it would even underestimate the negative effect of the bus replacement service, since only the additional waiting time (and not the lower vehicle capacity) is incorporated in the ridership prediction. Therefore it is proposed to use the minimum of f^R and f^T as frequency of rail-replacement bus services in ridership predictions, when vehicle capacity is not incorporated in the used model.

TABLE 6 Comparison Between Default And New Proposed Parameter Set

Parameter	Default parameter values	New parameter values
Elasticity E_δ	-1.1	-0.7
Waiting time coefficient a_3 for WTT^R	1.5	2.0
In-vehicle time coefficient a_1 for IVT^R	1.0	1.11
Frequency f^R	f^R	$\text{MIN}(f^R, f^T)$

Figure 4 (left) and Figure 4 (right) show for all four disturbances investigated in this study the effect of the default and new proposed parameter set on the predicted ridership reduction and on the average generalized travel costs per passenger, respectively. It can be seen that for all disturbances the new parameter set predicts a lower reduction in public transport ridership compared to the default parameter set. For δ_2 this effect is limited. However, for $\delta_1, \delta_3, \delta_4$ the predicted ridership is 20-25% lower compared to the default parameter set. This can mainly be explained by the less negative elasticity value E_δ (-0.7 instead of -1.1). From Figure 4 (right) it can also be concluded that the average generalized travel costs decrease for $\delta_1, \delta_3, \delta_4$ with 5-20% when applying the new parameter set, while these increase with 20% for δ_2 compared to the default parameter set. In the new parameter set, the use of rail-replacement bus services results in more perceived disutility compared to the default set, due to the use of a higher value for WTT^R , IVT^R and the longer average waiting time due to using $\min(f^R; f^T)$ instead of f^R . This would apply an increase in average generalized costs. However, the lower value for E_δ leads to a substantial lower predicted ridership reduction for $\delta_1, \delta_3, \delta_4$, of which the net effect is a decrease in average generalized costs. Since the predicted ridership reduction in the new parameter set hardly decreases for δ_2 , for this disturbance this effect does not fully compensate the increased generalized costs, therefore leading to a net increase in average generalized costs.

Figure 5 shows that the new parameter set results in an increased predicted accuracy for all four disturbances considered in this study. This shows that the new proposed parameter set is robust to better predict ridership effects of several (types of) planned disturbances. The relative improvement in prediction accuracy ranges between 3 and 13%. In the calculation of the R^2 , for $\delta_1, \delta_2, \delta_4$ in total $|L^a| * |T| * 2$ elements are computed. However, for δ_3 only $|L^a| * 2$ elements are used for the R^2 computation. This is because δ_3 occurred during the full duration of the Autumn holiday. During this holiday, travel patterns differ from travel patterns during regular working days or during Summer holiday, especially regarding the distribution over the day. Since there is no period of the year for which δ_0 occurred with a similar travel pattern, the effects per time period cannot be compared between δ_0 and δ_3 . Therefore, we only compared these scenarios aggregated over all time periods per working day per line $l \in L^a$.

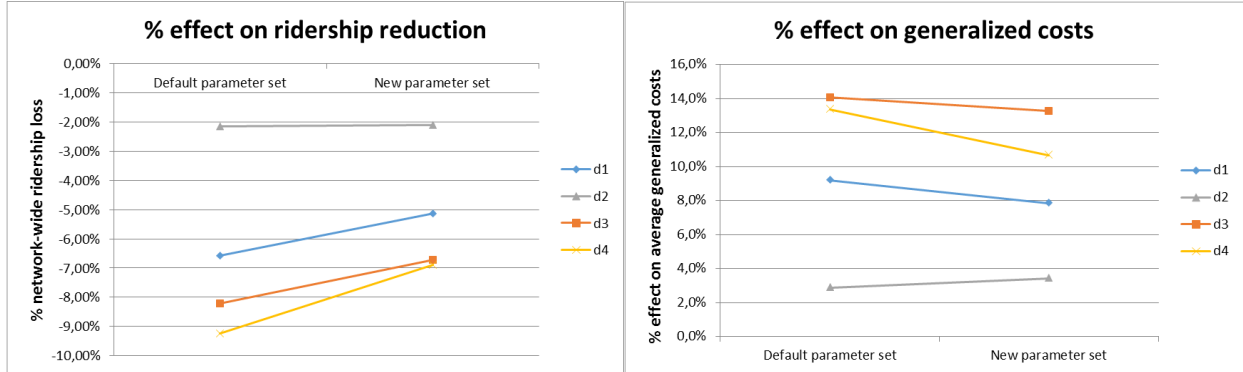


FIGURE 4 Effect of new parameter set on predicted ridership reduction (left) and on the average generalized travel costs per passenger (right)

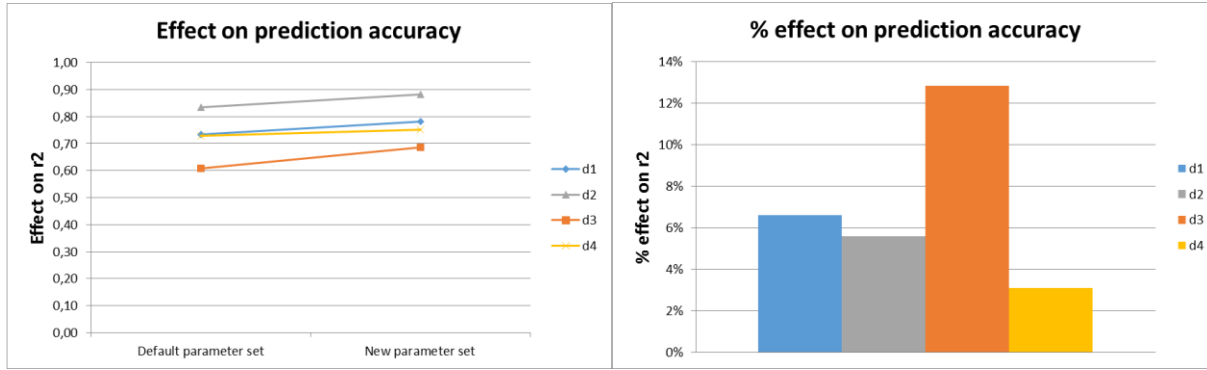


FIGURE 5 Absolute (left) and relative (right) effect of new parameter set on prediction accuracy

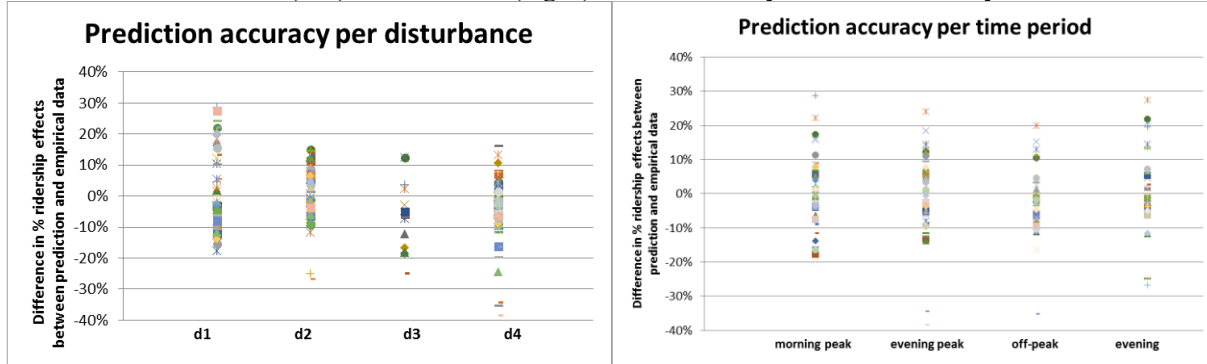


FIGURE 6 Resulting prediction accuracy for all $|L^a| * |T| * 2$ elements clustered per disturbance (left) and per time period (right)

Figure 6 shows the differences between the predicted and empirical relative effect of disturbances on public transport ridership for all elements, clustered per disturbance (left) and per time period (right). A positive value indicates that the predicted relative effect is less negative / more positive compared to the empirical data: ridership loss is thus *underestimated*, whereas ridership increases are *overestimated*. A negative value indicates the opposite, meaning that ridership losses are *overestimated* and ridership increases are *underestimated*. As can be seen, most predictions deviate in an acceptable range of $\pm 10\%$ from the empirical data. For δ_2, δ_4 (disturbances both occurring during Summer holiday) it can be observed that predictions tend to overestimate ridership losses, whereas for δ_1 (a disturbance occurring during working period on a line with a high share of commuting and business passengers) predictions tend to underestimate ridership losses (Figure 6 left). We can hypothesize that frequent, commuting passengers have a higher sensitivity for network changes than average, whereas passenger segments with a relatively high share of leisure / shopping trip purposes have a lower sensitivity for network changes than average. When considering prediction accuracy for different time periods, we see similar patterns. Most predictions in the different time periods do not deviate more than $\pm 15\%$ from the empirical data. For the morning peak, a slight tendency can be observed that predictions underestimate ridership losses, whereas for the evening period an opposite tendency can slightly be observed. Based on this we might hypothesize that passenger segments dominant during morning peak (mostly business and commuting passengers) have a relatively high sensitivity to network changes compared to passenger segments dominant during the evening period (mostly leisure passengers).

4.3 Reflection

In this study the accuracy of predicting the impact of temporary track closures on the number of passengers and passenger-distance is substantially improved by using the new proposed parameter set. Future research following this study can focus on four different aspects. First, in this study a rule-based three-step search procedure is applied to evaluate the prediction quality of several parameter sets. After first scanning the

solution space using a systematic grid-search over different combinations of parameter values, an in-depth search around promising parameter values is performed. In order to further optimize the parameter set, it is recommended to perform a full optimization evaluating all combinatorial possibilities, and/or to estimate a discrete choice model based on the revealed preference smart card data found for several disturbances to infer parameter values. This can lead to a set of parameter values which fit the empirical data in an optimal way and thus improving the prediction accuracy. Second, another interesting approach for future research is the application of machine learning approaches, training the prediction model based on passenger behavior during several historical disturbances which occurred on the public transport network. Prediction accuracy results from machine learning approaches can then be compared with our developed methodology. Third, it is recommended to investigate whether location-specific, and/or purpose-specific parameter values increase the prediction accuracy. In this study only generic and mode-specific parameter values are applied. Distinguishing between areas where a disturbance occurs based on socio-economic characteristics (e.g. age, income, percentage of public transport captives) might lead to better predictions. Also using different parameter values for different passenger segments based on their trip purpose, or based on different time periods (peak/off-peak, weekday/weekend) might improve the prediction accuracy, since sensitivities for different parameters might be different for these segments. Fourth, a path for future research could focus on the development of supply-side indicators to identify and categorize disturbances. Our study mainly focuses on demand predictions resulting from disturbances, without making a distinction between different types of disturbances. However, developing indicators to describe disturbances and categorize the type of disturbance can be valuable, since it allows for the development of different demand prediction parameter sets for different types of disturbances. This can result in a further improvement of the prediction accuracy.

5. CONCLUSIONS

The introduction of automated fare collection (AFC) systems in public transport travelling the last decades leads to the availability of large quantities of smart card data, which can be used to analyze current travel patterns and to predict the effect of structural network changes on future public transport usage. Predicting the impact of planned disturbances, like temporary track closures, on public transport ridership is however still an unexplored area. In the Netherlands, this area becomes increasingly important, given the many track closures public transport operators are being confronted with the last and upcoming years. Better predicting the impact of these disturbances gives more insights in passenger behavior during disturbances, and supports operators in aligning the supply of rail-replacement bus services with remaining demand and in predicting the impact on their revenues.

In this study we investigated the passenger impact of planned disturbances by comparing predicted and realized public transport ridership using smart card data. Four different disturbances which occurred in 2015 on the public transport network operated by HTM in The Hague, the Netherlands, are analyzed. A rule-based three-step search procedure is applied to find a parameter set resulting in higher prediction accuracy. The parameter set is calibrated using two of the four investigated disturbances, whereas the remaining two disturbances are used to validate the model.

Based on the study results we found a more negative in-vehicle time perception in rail-replacing bus services compared to in-vehicle time perception in the initial tram line. One minute tram travelling shows to be perceived as about 1.11 minute travelling in a rail-replacement bus service. Besides, when modelling rail-replacement services, it is recommended to use the frequency of the initial tram line instead of the usually higher frequency of the rail-replacement services. Passengers do not seem to perceive this theoretical benefit of higher frequencies of the rail-replacement bus, since this compensates for the lower vehicle capacity of a bus compared to a tram. If vehicle capacity is not incorporated in the prediction model, only incorporating the positive effect of a higher bus frequency aimed to compensate for this non-incorporated capacity effect would overestimate the level of service of the rail-replacement bus service. At last, also a 1.3 times higher waiting time perception for temporary rail-replacement services could be found, compared to waiting time perception for regular tram and bus lines. The new parameter set leads up to 13% higher prediction accuracy compared to the default parameter set, and shows to be a robust and valuable tool for public transport operators. The prediction model is used by HTM in practice, in which

the parameter set as recommended in this study is applied. Monitoring and further improving the prediction accuracy of the model will remain an important focus in the future.

ACKNOWLEDGEMENTS

This research is performed in cooperation with HTM, Delft University of Technology, Department of Transport & Planning and Goudappel Coffeng.

REFERENCES

- Arentze, T.A., and Molin, E.J.E. (2013). Travelers' preferences in multimodal networks: Design and results of a comprehensive series of choice experiments. *Transportation Research Part A*, 58, pp. 15-28.
- Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M. and White, P. (2004). The demand for public transport: a practical guide.
- Bunschoten, T., Molin, E.J.E. and Nes, R. van (2013). Tram or bus: does the tram bonus exist? *European Transport Conference*, Germany.
- Cats, O., Wang, Q. and Zhao, Y. (2015). Identification and classification of public transport activity centers in Stockholm using passenger flow data. *Journal of Transport Geography*, 48, pp. 10-22.
- Ding, C., Wang, D., Ma, X. and Li, H. (2016). Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. *Sustainability*, 8, pp. 1-16.
- Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H.M. and Attanucci, J.P. (2013). Automated inference of linked transit journeys in London using fare-transaction and vehicle location data. *Transportation Research Record: Journal of the Transportation Research Board*, 2343, pp. 17-24.
- Hofmann, M., and O'Mahony, M.. (2005). Transfer journey identification and analyses from electronic fare collection data. *Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems*, pp. 825-830.
- HTM Personenvervoer N.V. (2015). *Annual report 2015* (in Dutch). <https://www.htm.nl/media-/364324/jaarverslag-2015.pdf>. Accessed July 21, 2016.
- Idris, A., Habib, K. and Shalaby, A. (2015). An investigation on the performances of mode shift models in transit ridership forecasting. *Transportation Research Part A*, 78, pp. 551-565.
- Jang, W. (2010). Travel time and transfer analysis using transit smart card data. *Transportation Research Record: Journal of the Transportation Research Board*, 2144, pp. 142-149.
- Li, Y., Wang, X., Sun, S. and Ma, X. (2017). Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. *Transportation Research Part C*, 77, pp. 306-328.
- Metropoolregio Rotterdam Den Haag (2016). *Openbaar vervoer*. <http://mrdh.nl/project/openbaar-vervoer>. Accessed July 12, 2016.
- Munizaga, M.A., and Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C*, 24, pp. 9-18.
- Nazem, M., Trépanier, M. and Morency, C. (2011). Demographic analysis of route choice for public transit. *Transportation Research Record: Journal of the Transportation Research Board*, 2214, pp. 71-78.
- Nijënstein, S. and Bussink, B. (2015). Combining multimodal smart card data: exploring quality improvements between multiple public transport systems. *European Transport Conference*, Germany.
- Nunes, A.A., Dias T.G., and eCunha, J.F. (2016). Passenger journey destination estimation from automated fare collection system data using spatial validation. *IEEE Transactions on Intelligent Transportation Systems*, 17, pp. 133-142.
- Pelletier, M.P., Trépanier, M. and Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19, pp.

- 557-568.
- Seaborn, C., Attanucci, J. and Wilson, N.H.M. (2009). Analyzing multimodal public transport journeys in London with smart card fare payment data. *Transportation Research Record: Journal of the Transportation Research Board*, 2121, pp. 55-62.
- Significance, VU University John Bates Services TNO, NEA, TNS NIPO and PanelClix. Values of time and reliability in passenger and freight transport in the Netherlands. *Report for the Ministry of Infrastructure and Environment*, Significance, Den Haag, 2013.
- Trépanier, M., Tranchant, N. and Chapleau, R. (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems*, 11, pp. 1-14.
- Van Oort, N., Brands, T. and Romph, E. de. (2015a). Short-term prediction of ridership on public transport with smart card data. *Transportation Research Record: Journal of the Transportation Research Board*, 2535, pp. 105-111.
- Van Oort, N., Drost, M., Brands, T. and Yap, M. (2015b). Data-driven public transport ridership prediction approach including comfort aspects. *Proceedings of the Conference on Advanced Systems in Public Transport*, pp. 1-13.
- Wang, W., Attanucci, J.P. and Wilson, N.H.M. (2011). Bus passenger origin-destination estimation and related analyses using automated data collection systems. *Journal of Public Transportation*, 14, pp. 131-150.
- Wei, Y. and Chen, M. (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C*, 21, pp. 148-162.
- Yap, M.D., Cats, O., Van Oort, N. and Hoogendoorn, S.P. (2017). Robust transfer inference: a transfer inference algorithm for public transport journeys during disruptions. *Prepared for the 20th EURO Working Group on Transportation Meeting 2017, Budapest, Hungary*.
- Zhao, J., Rahbee, A. and Wilson, N.H.M. (2007). Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering*, 24, pp. 376-387.