Evaluation of passenger delays under different levels of information using a large-scale adaptive route choice model

Mads Paulsen[a,*], Thomas Kjær Rasmussen[a], Otto Anker Nielsen[a]

[a]*Transport Division, Department of Technology, Management and Economics, Technical University of Denmark, Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark*

Abstract

One thing about public transport is certain; it is uncertain. This study proposes a methodology that allows determining realistic door-to-door passenger delays incurred on the basis of such uncertainties. By combining Automated Vehicle Location (AVL) data for bus and train services with an adaptive route choice model, we test the influence of obtaining information about current delays by allowing four levels of adaptivity. The model is evaluated using more than 800,000 daily trips for 65 realised timetables of the large-scale multi-modal public transport network of the Metropolitan Area of Copenhagen. The model shows that adapting routes can heavily reduce incurred delays of passengers, even more so when updating routes during trips.

Keywords:

Passenger Delays, Public Transport Assignment, Agent-Based Simulation, Adaptive Route Choice, AVL Data, Passenger Information

1. Introduction

Public transport services are rarely fully punctual, and as such its passengers will often get higher or lower travel times than expected. Even when knowing the delays of all public transport services, the corresponding passenger delays are difficult to determine as they depend on the exact routes of passengers – routes that might change during the trip as passengers miss connections or discover that betters alternatives have arisen. Developing a model allowing the calculation of such passenger delays while taking into account that passengers adapt to delays of the system is the aim of this study.

Determining passenger delays are a lot easier if personal, trackable data is available, as shown by e.g. Jiang et al. (2012); Sun et al. (2016a,b); Antos and Eichler (2016) for smart card data and Carrel et al. (2015) for smart phone data. However, such data are rarely available, and obtaining permission to use such data can be tedious due to the juridical aspect of dealing with sensitive personal data.

*Corresponding author
Email address: madsp@dtu.dk (Mads Paulsen)

Models for passenger delays without additional personal information have existed for small networks since Hickman and Bernstein (1997). A medium sized railway network was handled by Nielsen et al. (2009), who also evaluated the impact of when to start looking for other alternatives.

Leng et al. (2018) used an agent-based approach to model passenger adaptation to a railway disruption, but assumed either agents with a unrealistically low level of adaptation or knowledge about future delay. The same can be said about Paulsen et al. (2018), which furthermore had two additional shortcomings. Firstly, the method was a cumbersome iterative approach which also meant that the model could not handle the entire population in a metropolitan area. Secondly, only railway delays were included in the study. The above mentioned shortcomings have been addressed in the current study.

Zhu and Goverde (2019) introduced a dynamic passenger assignment model for major railway disruptions that incorporates passenger responses to information given to passengers. However, it was only applied to a small network with 17 stations, and only considering a single disruption.

Khani (2019) proposed a model, that theoretically could be used for modelling passenger delays, as it allows modelling uncertain travel times whilst taking the probability of missing a transfer into account. However, it lacks the ability to include correlation between delays across departures, and is not suitable for evaluating several days of AVL data, as the computation time would scales with the number of included days to the power three.

Finally, some studies deal with the managing the operational side to minimise passenger delays, e.g. Dollevoet et al. (2012); Zhen and Jing (2016); Corman et al. (2017); Ghaemi et al. (2018), but it is beyond the scope of this short paper to go into a detailed review of those.

We contribute to the literature by proposing an adaptive route choice model for large-scale multi-modal public transport networks that allows en-route decisions of its agents. Dynamic changes to the routes are allowed on-the-go, and is dependent on the level of information the agent receives, ranging from no information to receiving information about new optimal paths every two and a half minutes based on the current delays of vehicles in the public transport system.

The remainder of the paper is structured as follows. Section 2 describes the proposed adaptive route choice model in brief. Section 3 introduces the data used in the study, and tests the methodology on a large-scale case study of Metropolitan Area of Copenhagen while investigating the impact of different levels of passenger information. Section 4 summarises the findings and discuss limitations and future work related to the study.

2. Model

We propose a dynamic public transport assignment model, that starts by obtaining initially planned routes of agents determined by searching for the shortest path in the planned timetable. As time progresses, agents move through the system either by foot or in a public transport service. At some moments in time, $t = n\tau, n \in \mathbb{N}$ with $\tau = 150s$, the agents may search for the shortest path from their current location to their destination in a partially realised timetable that includes delays up to that time and constantly extrapolated delays beyond. If a better alternative exists, the agent changes its route and starts to pursue this instead of the previous.

By restricting every agent to search at the same time (multiples of τ), their partial realised timetables are equal, why we only need to create realised timetables and corresponding RAPTOR (Delling et al., 2015) for these moments in time. Searching in a RAPTOR graph is very fast, and since its construction time is also manageable, this approach causes the entire model to be rather fast. We have used the version of RAPTOR proposed by and implemented in JAVA by Rieser et al. (2018), which is both fast and easily configurable. For agents onboard public transport services, a slightly modified search is performed. Firstly, the search is done from the next stop of that service, H^*, at the time the service reaches that stop, t^* – however based on the delay information up until time t like all other agents. Secondly, no boarding penalty is incurred if the new search causes the agent to board the service he/she is actually already in.

Fully realised timetables are used to secure that boarding and alightings only takes place when vehicles actually arrived/departed at/from stops. By traversing the day chronologically and keeping track of the locations of agents and locations and delays of public transport vehicles, a one-shot assignment can be performed, dramatically reducing the computation time compared to iterative approaches. This allows the model to be applied to larger networks and population than iterative approaches within the same computation time.

We propose four different configurations of the model, characterised by different situations in which the agents are assumed to search for new paths. As better alternatives can only arise if new information is available, the setups can also be interpreted as levels of information provided to the passenger. The *rigid* setup does not allow any alterations to the list of stops that needs to be visited, whereas the *non-adaptive* setup allows agents to search for a route at the beginning of their trip (rigid after that). In both cases, if a route becomes infeasible because they are at a stop that is no longer connected to their next intended stop, they simple give up and the trip is *incomplete*. In the *semi-adaptive* setup agents furthermore search for a new route when they are waiting at stops/stations. In the *full-adaptive* setup a shortest path search can also be performed whilst walking or onboard public transport services. Behaviourally, this may seem very tedious, but is achievable if thought of as an app doing searches and providing notifications to passengers when better alternatives emerge.

3. Case Study

3.1. Data

3.1.1. Automated Vehicle Location Data

The AVL data used in this study are provided by the national railway manager in Denmark, Rail Net Denmark, and the regional bus agency, Movia, for every weekday (Monday-Friday) of the fall (September, October, and November) 2014 – a total of 65 days.

Delay data was not available for local train and metro lines, but are otherwise predominantly available for all departures on the remaining train network, see Figure 2. Luckily, the effect of lack of data from local train and metro lines will most likely be limited as the local trains have a low demand (only 4.3% of trips use them, see Figure 3) and the metro has a high frequency down to two minutes headway in peak hours. A total of 4,414,922 observations of a train arriving or departing at a station were available in the data, corresponding to approximately 67,922 observations per day. The average delay was 10.6 seconds, see Table 1 and Figure 4 for additional descriptive statistics.

For buses, delay information is spread across the entire model area but is less complete than the train data, as the bus company only collected AVL data for a sample of their vehicles. A total of 60,400,126 observations were present in the dataset, corresponding to roughly 929,232 observations per day.

	Trains	Buses
N	2,207,329	30,200,063
Mean [min.]	0.18	1.97
Mode [min.]	-0.67	0.70
2.5% quantile [min.]	-1.65	-1.78
Median [min.]	-0.28	1.33
97.5% quantile [min.]	5.13	9.83
Std. dev. of mean [min.]	0.98	6.35
Coverage	93.99%[1]	48.57%

Table 1: Key statistical measures vehicle arrivals for trains and buses.

The data was initially cleaned in order to secure consistency in the AVL data. The criteria used for consistency were that realised times for each service are non-decreasing from stop to stop, and that the time difference between two consecutive elements (i.e. a stop arrival or a stop departure) cannot exceed two hours. For each departure, if at least one inconsistency is found, this observation as well as the previous and next observation is updated using linear interpolation of the delays based on their corresponding planned times. If extrapolation is needed, constant extrapolation of delays is used. This eventually causes no inconsistencies or all of the delays to be undefined. In the latter case, the observations were deemed too noisy to fix, and in lack of further information, the departure was assumed to follow the published timetable for all stops. The numbers reported in Table 1 are based on the cleaned dataset. Departures assumed to operate according to the schedule (either due to lack of data or too noisy data) are those causing the coverage to drop below 100%.

[1]56.61 % if including local trains (8.734 daily arrivals) and metro (15,128 daily arrivals) in the calculation of coverage.

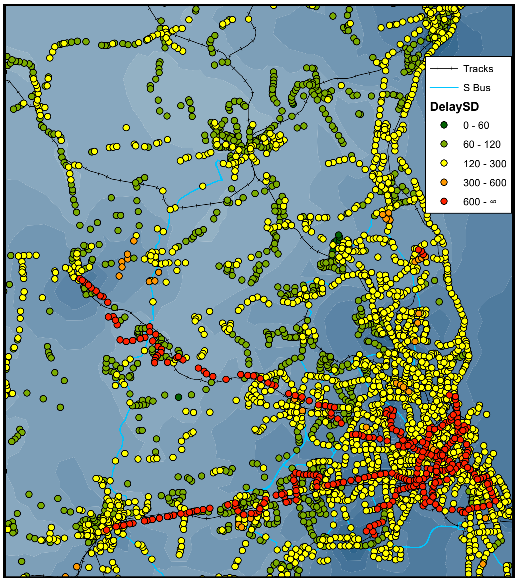

Figure 1: Standard deviation of arrival delays at bus stops.

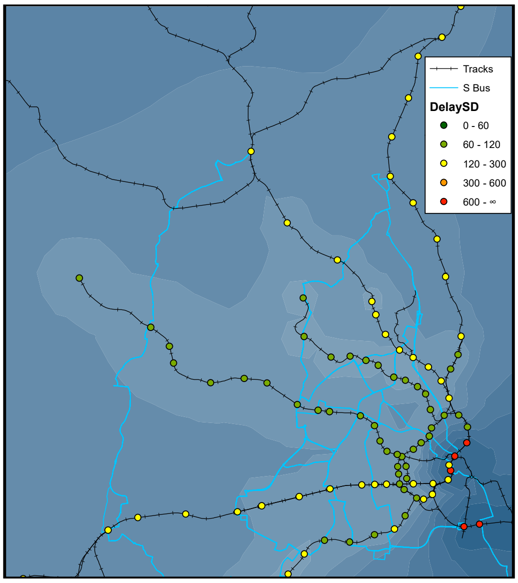

Figure 2: Standard deviation of arrival delays at train stations.

3.1.2. Demand

The demand used for the study is based on the Copenhagen Model for Person Activity Scheduling (COMPAS) (Prato et al., 2013), and consists of 387,511 agents having a total of 812,359 public transport trips. Trips not using public transport were ignored in the study.

The time use of walking, waiting, and for each of the submodes as well as number of boardings are visualised in Figure 3. It can be seen that about 10% find it most suitable to walk the entire trip in the base scenario, roughly two thirds use bus, 45% use S-trains, barely 20% use metro, regional and InterCity-trains account for roughly 12.5%, whereas less than 5% use local trains. The walking time distribution is fairly wide, but with about 60% of trips having between five and 20 minutes walk. Less than 5% of the trips boards more than three vehicles, whereas both one and two boardings account for roughly 35% of trips each.

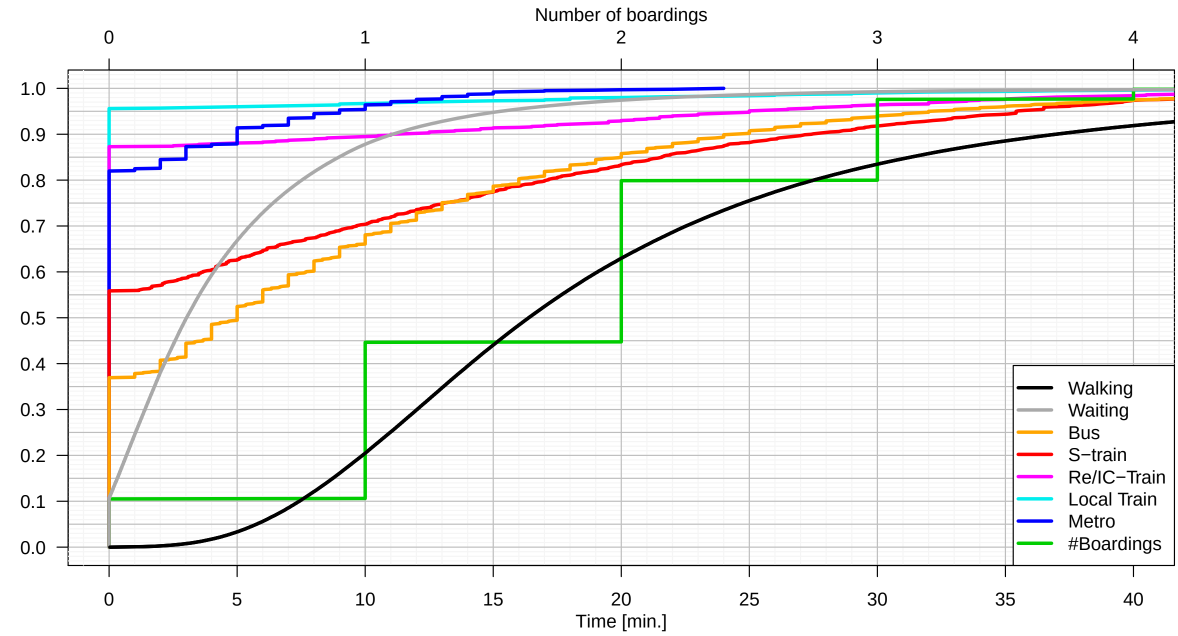

Figure 3: Empirical cumulative distribution function of number of boardings and time use in different submodes as well for waiting and walking.

3.2. Results

The model was tested using the data presented in Section 3.1. Different utility parameters for submodes of the public transport system have been used in order to allow for more realistic paths than when only using pure travel time. The parameters are based on previous studies in Copenhagen, e.g. Eltved et al. (2018), and can be found in Table 2.

	Bus	Re/IC Trains	Local/S-Trains	Metro	Wait	Walk	Boarding
Utility per minute	-1	-1.1	-0.9	-0.85	-1.3	-1.6	-
Utility per event	-	-	-	-	-	-	-4

Table 2: Utilities for various modes, waiting, walking, and transfers

A base scenario using the published timetable, and scenarios using realised timetables for the four possible setups have been run for each of the 65 days with the parameters from Table 2 and a time step of $\tau = 150s$ from 03:00am to 03:00am the following day. With all setups finishing in at most half a calculation day per simulation day, see Table 3, the model computes faster than real-time, and can therefore be deemed as large-scale operational in practice.

The cumulative distribution function for the difference in generalised cost for the four setups with respect to the base scenario are depicted in Figure 4 alongside the distributions of train and bus vehicle arrivals.

	Base	Rigid	Non-Adaptive	Semi-Adaptive	Non-Adaptive
Computation time [min.]	60.0	130.7	495.5	532.0	652.7

Table 3: Average running time per day for various setups.

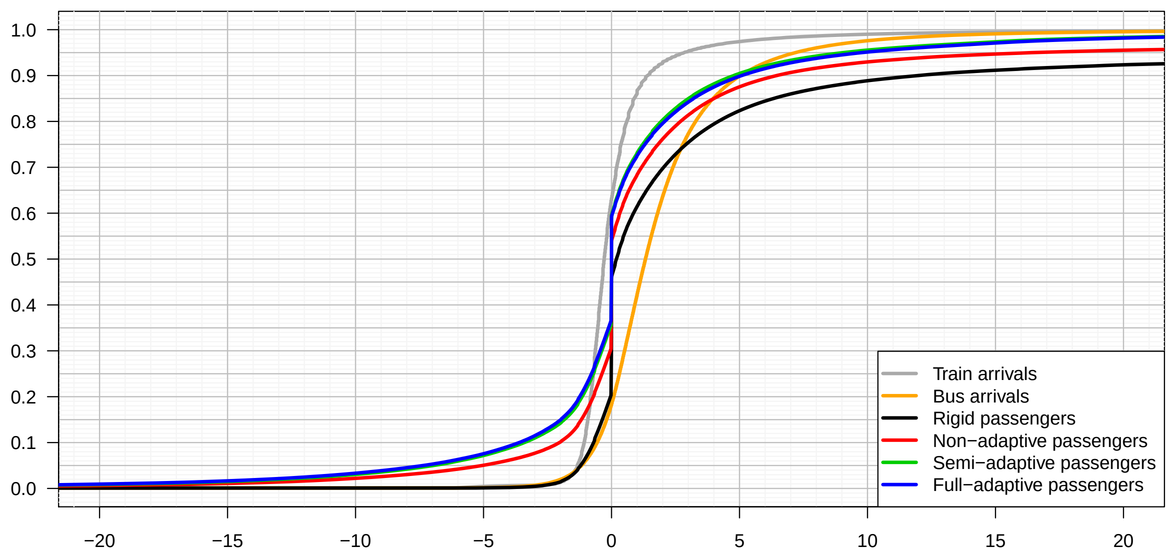

Figure 4: Empirical cumulative distribution function for bus arrivals [min.], train arrivals [min.], and passenger delays [Gen. cost] for the four different setups.

Although buses have a wider distribution that trains, both distributions are more narrow than any of the passenger delay distributions, verifying what is typically the case, namely that passenger delays have larger variations than vehicle delays. It is also seen that even in the rigid setup 20% of trips are better off than intended. The rigid trips are predominantly a lot worse than intended, though.

By updating the route at the beginning of trips (the non-adaptive setup), the losses can be limited – and some dramatic improvements can also be obtained. Updating routes throughout the trip either at stations (semi-adaptive) or anywhere (full-adaptive) improves this tendency, and reduces the average delay by additional 11.7 generalised cost on top of the 10.3 achieved by the non-adaptive setup, see Table 4. There is not much difference between the semi- and full-adaptive scenario (0.04 on average), and the increase in variation is also negligible (0.03).

Although the difference between the semi- and full-adaptive setup is small, the full-adaptive trip does deviate more from the intended route according to Table 5. It is also seen that the two setups where routes are adapted on-the-go essentially cause all trips to complete. Notice that the previous tables and diagrams are based solely on the complete

	Rigid	Non-Adaptive	Semi-Adaptive	Full-Adaptive
N	48,462,901	50.469.428	52.797.979	52,798,103
Mean	23.27	12.94	1.24	1.20
2.5% quantile	-1.63	-9.62	-11.63	-11.22
Median	0.24	0.00	0.00	-0.00
97.5% quantile	360.41	95.21	20.76	20.81
Std. dev. of mean	719.41	45.53	5.88	5.91

Table 4: Key statistical measures for passenger delays [normalised generalised cost].

trips, as it is difficult to put a proper cost on an incomplete trips. This does, however, favour the non-adaptive and especially the rigid setup, as such trips are then not punished at all.

	Rigid	Non-Adaptive	Semi-Adaptive	Full-Adaptive
Incomplete trips [%]	8.22	4.42	0.01	0.01
Same stops [%]	100.00	74.18	61.60	55.96
Same lines [%]	97.03	76.05	66.71	63.76
Same departures [%]	91.09	63.57	53.59	50.86

Table 5: Similarities between intended route and modelled routes under different level of information. Incomplete trips are reported as percentage of all attempted trips.

To further explore the difference between the various setup, in the following we filter the trips so that we only consider passengers intending (in the base scenario) to use the regional coast line from Elsinore to Sweden, which is the eastern-most railway with available data in Figure 2. We then calculate the intended number of activities (boardings and unboardings) at each stop in the base scenario and for all days in the four setups. The average difference in activities for each stop is shown for each of the four setups in Figures 5-8. The underlying surface maps has been derived using Inverse Squared Distance Weighting (ISDW) (Shepard, 1968).

As expected, no stops gains activities when using the rigid setup, as all passengers stick to their intended stations - if they manage to complete their trip, that is. Already in the non-adaptive setup, we begin to see some alternatives in the bus, metro, local railway, and suburban railway system being used more frequently, and that passengers especially avoid stations in the southern part of the regional coast line. With the semi- and the full-adaptive setup we see are larger spatial deviation, for instance with the suburban railway line from Hillerød (dark green spot in the north western part of the maps) being used very frequently, as well as local train lines and bus lines leading to it. All these results seems behaviourally realistic.

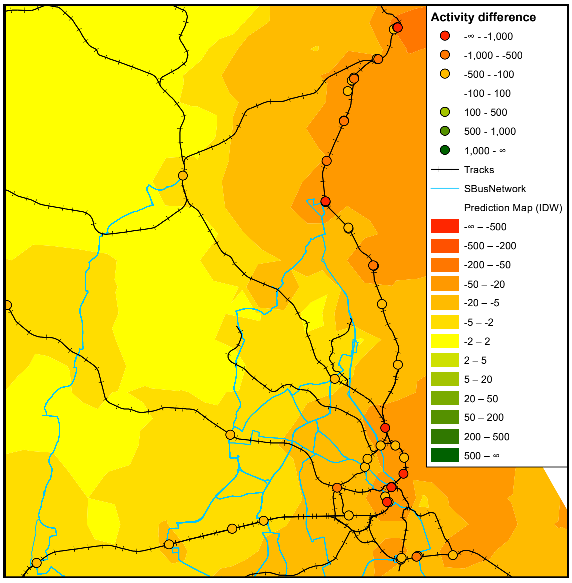

Figure 5: Difference in activities at stops in the rigid setup w.r.t. base level.

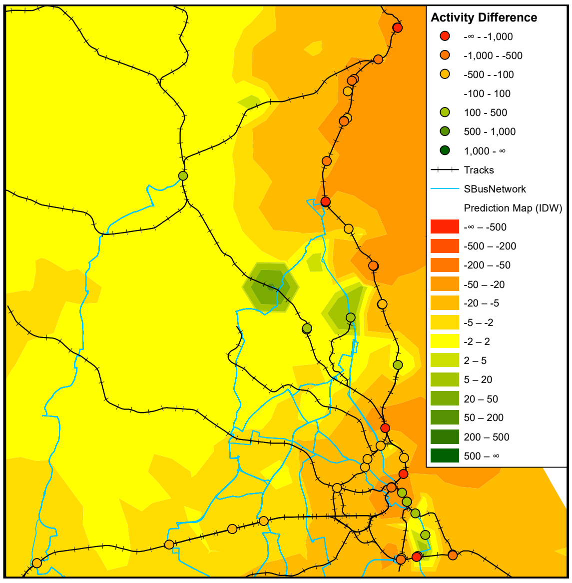

Figure 7: Difference in activities at stops in the non-adaptive setup w.r.t. base level.

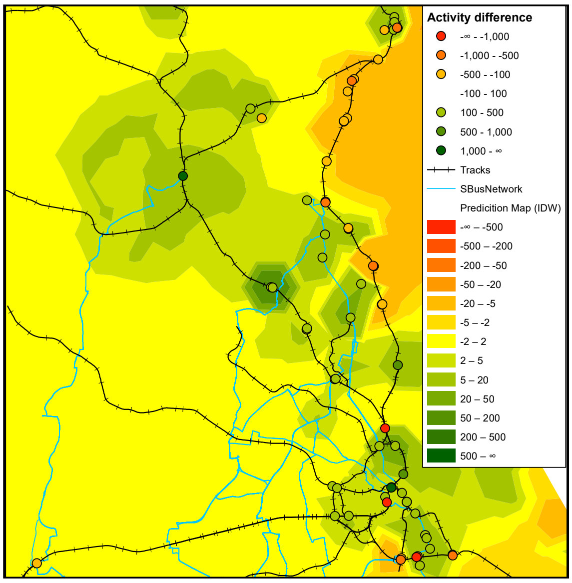

Figure 6: Difference in activities at stops in the semi-adaptive setup w.r.t. base level.

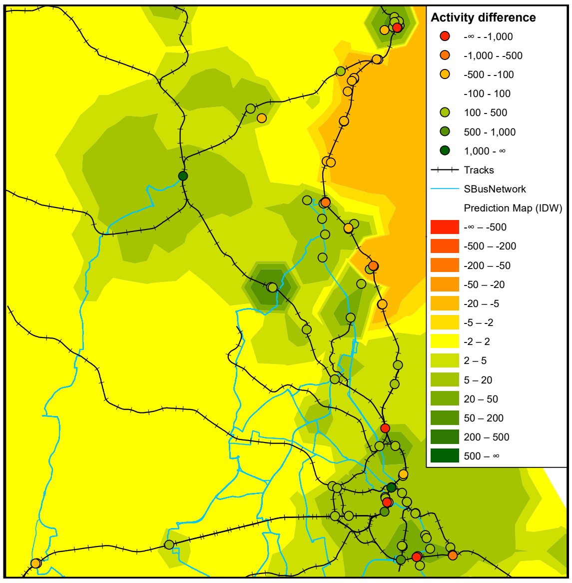

Figure 8: Difference in activities at stops in the full-adaptive setup w.r.t. base level.

We now skip the filter, and tries to investigate where the differences between the semi- and full-adaptive setup are. In in Figure 9 the locations where passengers receive notifications of better alternatives whilst travelling by train/bus/metro in the full-adaptive setup are plotted. It is seen that the notifications are often clustered where railways or major bus lines intersect. This also holds for the part of the railway system in central Copenhagen, where all except one regional and suburban railway line share stops.

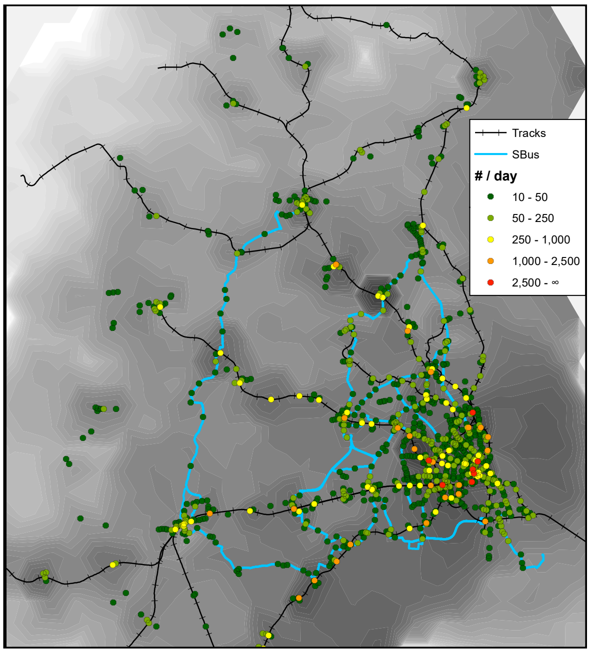

Figure 9: Spatial distribution of locations where notifications of better alternatives are received in the full-adaptive setup.

4. Conclusion and Future Work

This study proposed a methodology for evaluating passenger delays by combining AVL data with an adaptive route choice model using partially and fully realised timetables. As the first passenger delay model that does not use personal data, it has modelled passenger delays in a truly large-scale multi-modal scenario including delays of both trains and buses. The large-scale case study of more than 800,000 daily trips for 65 days, showed that updating routes (i.e. providing information to passengers) in the public transport system is crucial for minimising passenger delays. Delays can on average be halved by updating at the beginning of the trip, and can be reduced to about one minute if updating frequently during the trip, e.g. by targeting information to individual travellers waiting at stops. The additional gain of providing personalised information to onboard passengers was found to be very limited.

One of the major limitations of this study is that bus delay data was only available for barely half of the departures, and no data was available for local trains nor the metro. It would be interesting to perform a similar study with such data available, as this would most likely reduce the number of zero entries of passenger delays found in this study.

While being already highly useful for learning from and understanding issues of past timetables, an additional potential utilisation of the model could be to apply it to synthetic realised timetables of proposed future timetables. This will require an additional model that can create artificial delays, e.g. similar to what Jensen et al. (2017) proposes for railway delays. With such a methodology available, our proposed model could be a valuable support tool for public transport planners, as it would allow passenger delays of future timetables to be estimated.

References

Antos, J., Eichler, M.D., 2016. Tapping into Delay: Assessing Rail Transit Passenger Delay With Data from a Tap-In, Tap-Out Fare System. Transportation Research Record: Journal of the Transportation Research Board 2540, 76–83. URL: http://trrjournalonline.trb.org/doi/10.3141/2540-09, doi:10.3141/2540-09.

Carrel, A., Lau, P.S., Mishalani, R.G., Sengupta, R., Walker, J.L., 2015. Quantifying transit travel experiences from the users' perspective with high-resolution smartphone and vehicle location data: Methodologies, validation, and example analyses. Transportation Research Part C: Emerging Technologies 58, 224–239. URL: http://linkinghub.elsevier.com/retrieve/pii/S0968090X15001060, doi:10.1016/j.trc.2015.03.021.

Corman, F., D'Ariano, A., Marra, A.D., Pacciarelli, D., Samà, M., 2017. Integrating train scheduling and delay management in real-time railway traffic control. Transportation Research Part E: Logistics and Transportation Review 105, 213–239. URL: http://dx.doi.org/10.1016/j.tre.2016.04.007, doi:10.1016/j.tre.2016.04.007.

Delling, D., Pajor, T., Werneck, R.F., 2015. Round-Based Public Transit Routing. Transportation Science 49, 591–604. URL: http://pubsonline.informs.org/doi/10.1287/trsc.2014.0534, doi:10.1287/trsc.2014.0534.

Dollevoet, T., Huisman, D., Schmidt, M., Schöbel, A., 2012. Delay Management with Rerouting of Passengers. Transportation Science 46. URL: https://doi.org/10.1287/trsc.1110.0375, doi:10.1287/trsc.1110.0375.

Eltved, M., Nielsen, O.A., Rasmussen, T.K., 2018. The influence of frequency on route choice in mixed schedule and frequency-based public transport systems – The case of the Greater Copenhagen Area. Proceedings of the Conference for Advanced Systems in Public Transport - CASPT .

Ghaemi, N., Zilko, A.A., Yan, F., Cats, O., Kurowicka, D., Goverde, R.M., 2018. Impact of railway disruption predictions and rescheduling on passenger delays. Journal of Rail Transport Planning \& Management URL: http://linkinghub.elsevier.com/retrieve/pii/S2210970617300872, doi:10.1016/j.jrtpm.2018.02.002.

Hickman, M.D., Bernstein, D.H., 1997. Transit Service and Path Choice Models in Stochastic and Time-Dependent Networks. Transportation Science 31, 129–146. URL: http://pubsonline.informs.org/doi/abs/10.1287/trsc.31.2.129, doi:10.1287/trsc.31.2.129.

Jensen, L.W., Landex, A., Nielsen, O.A., Kroon, L.G., Schmidt, M., 2017. Strategic assessment of capacity consumption in railway networks: Framework and model. Transportation Research Part C: Emerging Technologies 74, 126–149. URL: https://linkinghub.elsevier.com/retrieve/pii/S0968090X16302066, doi:10.1016/j.trc.2016.10.013.

Jiang, Z.B., Li, F., Xu, R.H., Gao, P., 2012. A simulation model for estimating train and passenger delays in large-scale rail transit networks. Journal of Central South University 19, 3603–3613. URL: http://link.springer.com/10.1007/s11771-012-1448-9, doi:10.1007/s11771-012-1448-9.

Khani, A., 2019. An online shortest path algorithm for reliable routing in schedule-based transit networks considering transfer failure probability. Transportation Research Part B: Methodological 126, 549–564. URL: https://linkinghub.elsevier.com/retrieve/pii/S0191261517302771, doi:10.1016/j.trb.2019.04.009.

Leng, N., De Martinis, V., Corman, F., 2018. Agent-based simulation approach for disruption management in rail schedule. Proceedings for the 14th Conference on Advanced Systems in Public Transport and Transit Data 2018 URL: https://www.research-collection.ethz.ch:443/handle/20.500.11850/281348.

Nielsen, O.A., Landex, O., Frederiksen, R.D., 2009. Passenger delay models for rail networks, in: Schedule-Based Modeling of Transportation Networks. Springer US, Boston, MA. chapter 2, pp. 27–49. URL: http://link.springer.com/10.1007/978-0-387-84812-9_2, doi:10.1007/978-0-387-84812-9{_}2.

Paulsen, M., Rasmussen, T.K., Nielsen, O.A., 2018. Modelling Railway-Induced Passenger Delays in Multi-Modal Public Transport Networks, in: Hickman, M.D. (Ed.), Proceedings for the 14th Conference on Advanced Systems in Public Transport and Transit Data 2018, Brisbane, Australia.

Prato, C.G., Rasmussen, T.K., Nielsen, O.A., Watling, D.P., 2013. A disaggregate pseudo-dynamic assignment for the activity-based model of the

Greater Copenhagen Area, in: Joao, V. (Ed.), 13th World Conference on Transport Research (WCTR), Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. pp. 1–19.

Rieser, M., Métrailler, D., Lieberherr, J., 2018. Adding Realism and Efficiency to Public Transportation in MATSim, in: 18th Swiss Transport Research Conference (STRC), Monte Verità / Ascona, Switzerland.

Shepard, D., 1968. Two- dimensional interpolation function for irregularly- spaced data, in: Proceedings of the 1968 23rd Acm National Conference, ACM Press. pp. 517–524. doi:10.1145/800186.810616.

Sun, H., Wu, J., Wu, L., Yan, X., Gao, Z., 2016a. Estimating the influence of common disruptions on urban rail transit networks. Transportation Research Part A: Policy and Practice 94, 62–75. URL: http://linkinghub.elsevier.com/retrieve/pii/S0965856416303913, doi:10.1016/j.tra.2016.09.006.

Sun, Y., Shi, J., Schonfeld, P.M., 2016b. Identifying passenger flow characteristics and evaluating travel time reliability by visualizing AFC data: a case study of Shanghai Metro. Public Transport 8, 341–363. URL: http://link.springer.com/10.1007/s12469-016-0137-8, doi:10.1007/s12469-016-0137-8.

Zhen, Q., Jing, S., 2016. Train rescheduling model with train delay and passenger impatience time in urban subway network. Journal of Advanced Transportation 50, 1990–2014. URL: http://doi.wiley.com/10.1002/atr.1441, doi:10.1002/atr.1441.

Zhu, Y., Goverde, R.M.P., 2019. Dynamic Passenger Assignment for Major Railway Disruptions Considering Information Interventions. Networks and Spatial Economics 19, 1249–1279. URL: http://link.springer.com/10.1007/s11067-019-09467-4, doi:10.1007/s11067-019-09467-4.