Impacts of real-time information levels in public transport: A large-scale case study using an adaptive passenger path choice model*

Mads Paulsen*[1], Thomas Kjær Rasmussen[2], and Otto Anker Nielsen[3]

[1]Postdoc, Department of Technology, Management and Economics, Technical University of Denmark, Denmark
[2]Associate Professor, Department of Technology, Management and Economics, Technical University of Denmark, Denmark
[3]Professor, Department of Technology, Management and Economics, Technical University of Denmark, Denmark

SHORT SUMMARY

Public transport services are often uncertain, causing passengers' travel times and routes to vary from day to day. This study uses three months of historical Automatic Vehicle Location (AVL) data to calculate corresponding realised routes and passengers delays in a large-scale, multi-modal public transport network by formulating and implementing an adaptive passenger path choice model, and apply it to an agent-based scenario of Metropolitan Copenhagen with 801,719 daily trips. Five different levels of real-time information are analysed, ranging from no information at all to global real-time information being available everywhere. The results of more than 258 million inferred passenger delays show that variability of passengers' travel time is considerable and much larger than that of the public transport vehicles. Furthermore, obtaining global real-time information at the beginning of the trip reduces passengers delay dramatically, although still being inferior to receiving such along the trip.

Keywords: Agent-based simulation; Automatic vehicle location data; Passenger delays; Passenger path choice model; Public transport; Real-time information

1. INTRODUCTION

Public transport services are rarely fully punctual, resulting in passengers often witnessing 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 better alternatives have arisen.

Determining passenger delays are a lot easier if personal, trackable data is available, as shown by e.g. Jiang et al. (2012) for smart card data and Carrel et al. (2015) for smart phone data. However, such data are not always available, and obtaining permission to use such data can be tedious due to the juridical aspect of dealing with sensitive personal data. This study develops a model that can calculate such passenger delays based on realised vehicle delays while taking into account different levels of real-time information that passengers can adapt to along their way.

*The research presented in this short paper has been published in Transportation Research Part A: Policy and Practice (Paulsen, Rasmussen & Nielsen, 2021). However, it has not yet been presented at an international conference due to the cancellation of hEART2020/2021 in Lyon.

Models for calculating passenger delays supporting real-time passenger information have existed for small networks since Hickman & Bernstein (1997). Since then numerous studies have followed, however, with most of them applied to small networks (Comi et al., 2016; Nuzzolo et al., 2016; Rambha, Boyles & Waller, 2016) or urban train networks where other relevant modes of public transport were ignored (Cats & Jenelius, 2014; Landex & Nielsen, 2006; Nielsen, Landex & Frederiksen, 2009).

Additionally, real-information was only available at the stop-level in Nuzzolo, Russo & Crisalli (2001), while Estrada et al. (2015) did not allow transfers. In the works with real-time information using MILATRAS (Wahba & Shalaby, 2011) and in Yao et al. (2017), the real-time information scenarios were not compared to a scenario without such information. Leng & Corman (2020) modelled door-to-door passenger delays, but considering only real-time information pre-trip, and only concerning a single major disruption. Finally, Zargayouna et al. (2018) dealt with omnipresent real-time information, in a case study of Toulouse, France.

However, whereas Zargayouna et al. (2018) evaluates the effect of various levels of real-time information in terms of the proportion of passengers being exposed to such information, this study evaluates the effect of *when* such information is given to (all) passengers. This is done by analysing the following five levels of real-time information, and analysing the marginal effect of increasing the real-time information offered to passengers from one level to the next.

R0 *No information*: Agents pursue their intended path, i.e. access stop, egress stop, and possibly transfer stop(s), only allowing temporal adjustments in case of missed connections.

R1 *Pre-trip information*: Agents search for the shortest path at the beginning of their trip using the real-time information available at that point.

R2 *Information at stops*: In addition to potentially updating their path at the beginning of their trip (R1), agents can also adapt to real-time information while waiting at stops/stations (but not on-board a service).

R3 *Information everywhere*: No restriction on where agents can adapt, i.e.they also use real-time information to search for better alternatives while walking and on-board public transport services.

R∞ *Perfect information*: As opposed to the other real-time information levels, the passengers know all past, current, and future delays in advance, allowing always choosing the optimal path a priori without en-route adaptation.

R3 can be thought of as an app continuously performing searches and providing notifications when better alternatives emerge (Estrada et al., 2015; Zargayouna et al., 2018). R∞ is unrealistic and solely included in order to establish a hypothetical lower bound for the passenger delay of each trip.

Apart from modelling the impact of specific levels of real-time information that previous models could not isolate, the study also contributes to the literature by being the first of its kind to model real-time passenger information and door-to-door passenger delays for an entire metropolitan area with everyday vehicle delays across many days. Whereas this has previously mostly been done for major disruptions (Leng & Corman, 2020) or in uni-modal networks (Landex & Nielsen, 2006; Nielsen, Landex & Frederiksen, 2009; Cats & Jenelius, 2014), this study considers realistic everyday vehicle delays based on actual days of observed AVL data of trains and buses.

The remainder of the paper is structured as follows. Section 2 describes the overall framework and the proposed adaptive passenger path choice mode in brief. Section 3 introduces the data used in the study, and investigates the impact of different levels of real-time passenger information when applying the methodology to a large-scale case study of Metropolitan Copenhagen. Section 4 summarises the findings and discuss future work and applications.

2. METHODOLOGY

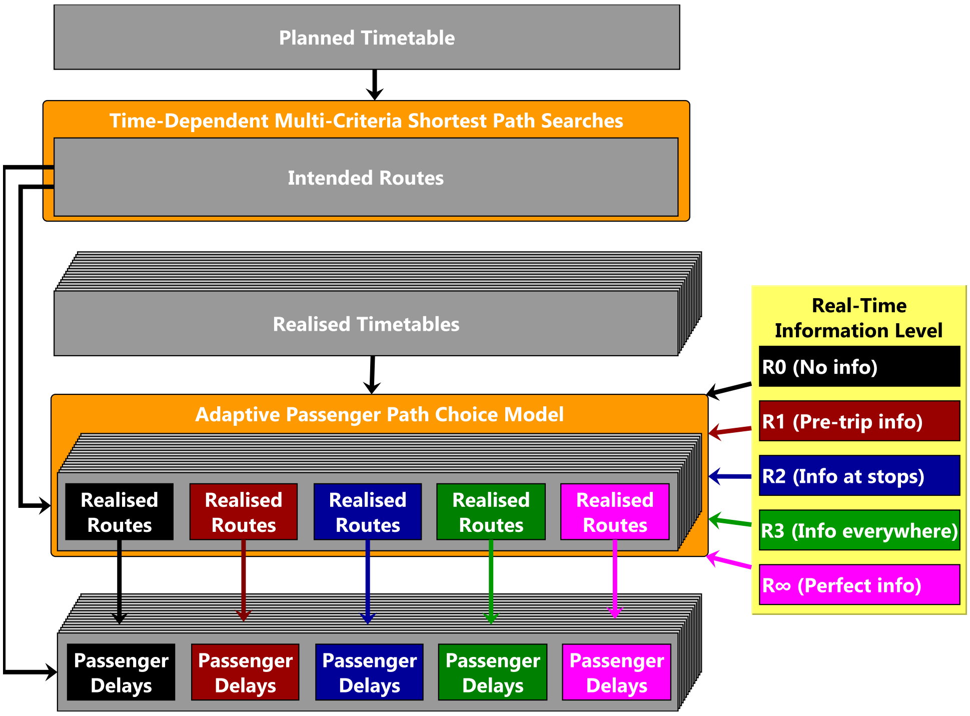


Figure 1: The overall framework.

For each of the five introduced real-time information levels, realised timetables constructed from 65 weekdays of actual AVL data for trains and buses in Metropolitan Copenhagen will be used and combined with an adaptive passenger path choice model in order to determine *realised* routes and travel times for 801,719 daily trips in the area. By comparing these to the corresponding *intended* routes and travel times based on the planned timetable, the door-to-door *passenger delays* can be determined, see Figure 1.

This study takes a model-based approach to model the necessary *realised* routes, and proposes an agent-based adaptive passenger path choice model that allows en-route decisions of its agents based on global real-time information. Figure 2 provides a graphical representation of the model, and shows how our adaptive passenger assignment model takes agents adaptively through the public transport system.

Simulations begin at $t = t_s$ (3 am), and continues in timesteps of size τ = 150s until the end-time of the simulation t_e (3am the following day) is reached. In each timestep we import the real-time timetable for time t – a timetable enriched with current delays at time t and estimations of future delays – and construct the corresponding RAPTOR graph (Delling, Pajor & Werneck, 2015). These RAPTOR graphs are used for the shortest path searches that the agents perform during the simulation , but are always overruled by the fully realised timetables (actual operation) in terms of when vehicles arrives and departs at/from stops.

When the simulation starts, all agents have their current location set to the location of their initial activity, their status set to ACTIVITY, and their current time being t_s. These values, as well as a destination, a location for the next stop/activity in their itinerary, and their next intended public transport departure are updated as they move through the system and eventually end their trip.

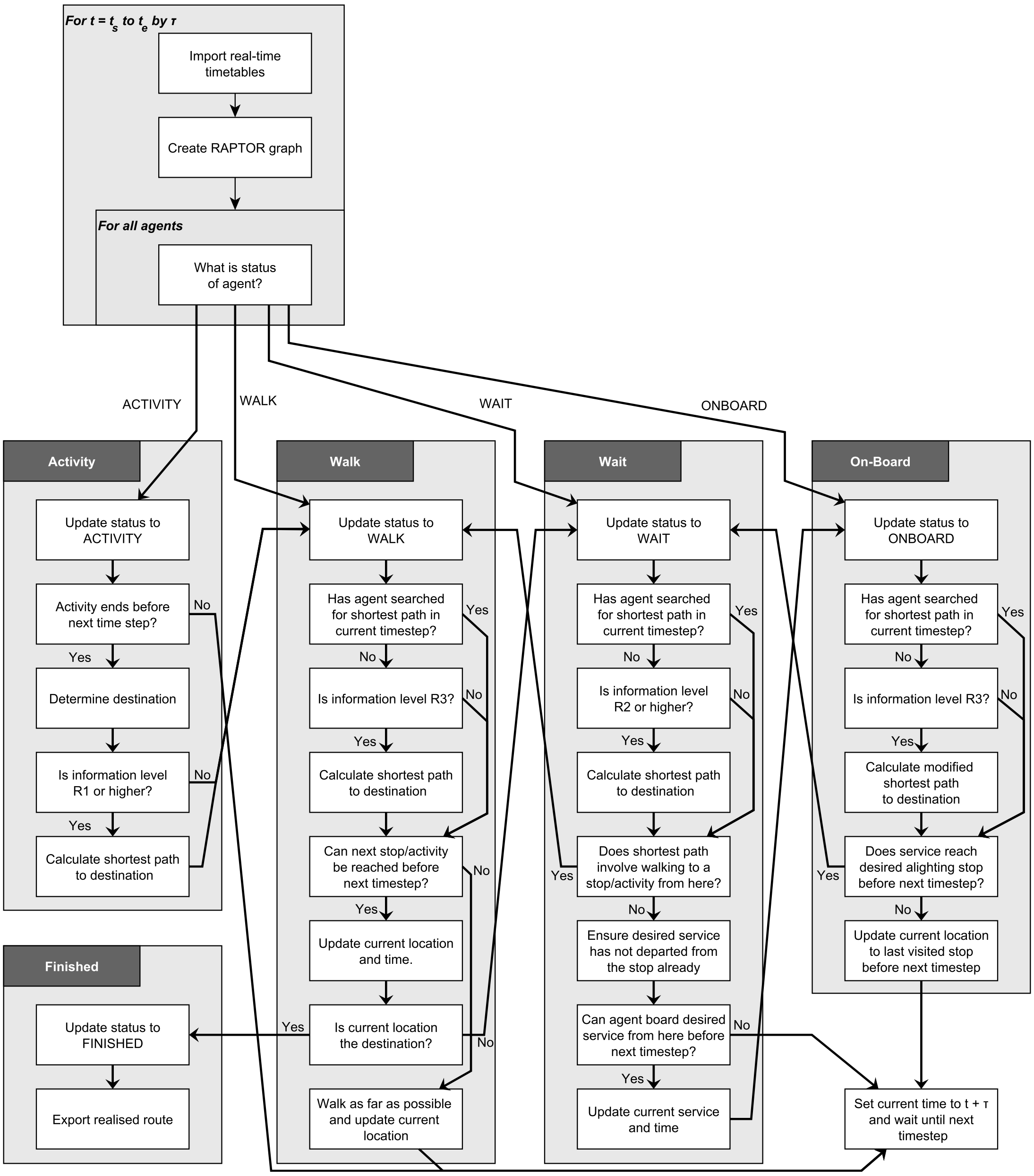


Figure 2: A graphical representation of the proposed adaptive passenger assignment algorithm.

The shortest paths needed in the simulation are found by an improved version of the original RAPTOR graph approach proposed by Delling, Pajor & Werneck, 2015. The methodology was extended and implemented for MATSim (Horni, Nagel & Axhausen, 2016) in Rieser, Métrailler & Lieberherr, 2018 while allowing different utilities to different submodes of the transport network. In addition, to speed up the searches, for this study we furthermore implemented a goal-directed search with pre-processed minimum travel costs between stops as proposed by Wagner & Willhalm (2007), allowing unattractive search directions to terminate earlier.

It is beyond the scope of this short paper to go into further details with the technicalities of the model.

3. RESULTS AND DISCUSSION

Data

We tested the model for 65 realised days – every weekday of the fall of 2014. 2,207,393 train delays for intercity, regional and suburban trains were provided by Rail Net Denmark, whereas the 29,718,431 delay data points for buses were provided by Movia (the regional bus agency). 95.2% of the train departures were covered by the delay data, whereas 72.5% of the bus departures were covered, however, with popular lines having higher coverage. For each of these days – and for each of the real time information levels – passenger delays were calculated for 801,719 trips. These trips were based on the demand model of COMPAS (Prato et al., 2013).

Configuration and computation times

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, Nielsen & Rasmussen (2018), and can be found in Table 1.

	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 1: 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 1. With all real-time information levels finishing in at most six hours per simulation day, see Table 2, the model is definitely large-scale operational in practice.

	Base	R0	R1	R2	R3	R∞
Computation time [min.]	30.9	35.6	82.6	148.8	331.3	72.5

Table 2: Average computation time per day for various real-time information levels.

Results

Key figures on passenger delays are given in Table 3 for each of the five information levels. In line with the literature (Landex & Nielsen, 2006; Nielsen, Landex & Frederiksen, 2009; Paulsen, Rasmussen & Nielsen, 2018), it is seen that overall the passenger delays have higher volatility than vehicle delays, see also Figure 3. Acquiring pre-trip info reduces mean delays and their standard deviation considerably, although such agents cannot spatially update their path en-route. However, it is still inferior to also applying real-time information while waiting at stops (R2). Having the ability to apply real-time information everywhere (R3) only reduces the mean delay negligibly compared to R2. Perfect information (R∞) reduces the standard deviation by more than 50% compared to R2 and R3 and almost diminishes the average delay, showing that both R2 and R3 are quite far from this theoretical lower bound.

The cumulative distribution functions of passenger delays for each of our five real-time information levels, as well as the empirical distribution functions of train and bus arrivals can be found in Figure 3. It is clearly seen that passenger delays have much thicker tails than the vehicle delays, again indicating that passenger delays are far more volatile.

	Vehicle arrivals		Passengers delays				
	Trains	Buses	R0	R1	R2	R3	R∞
Mean	0.18	1.99	8.74	4.96	3.10	3.07	-0.05
Std. dev.	2.59	3.30	22.67	18.90	12.19	12.43	5.87
2.5% quantile	-1.65	-1.78	-1.10	-8.63	-9.61	-10.15	-13.63
Median	-0.28	1.33	2.06	0.86	0.63	0.65	0.00
97.5% quantile	5.13	9.85	53.77	47.49	29.25	29.10	10.29

Table 3: Key statistical measures of delays for vehicle arrivals [minutes] and passengers [gcu].

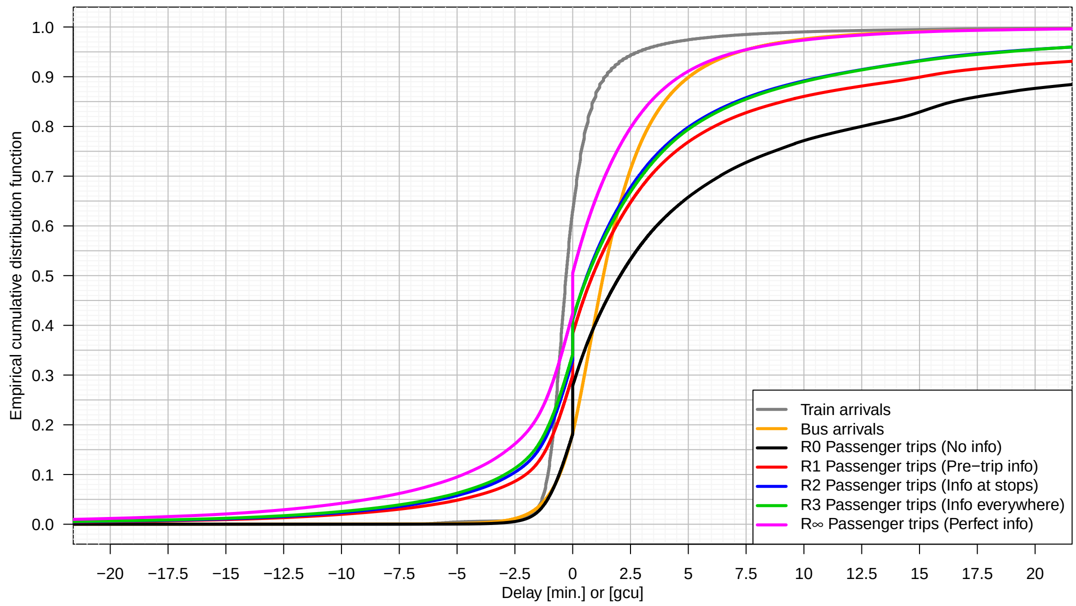
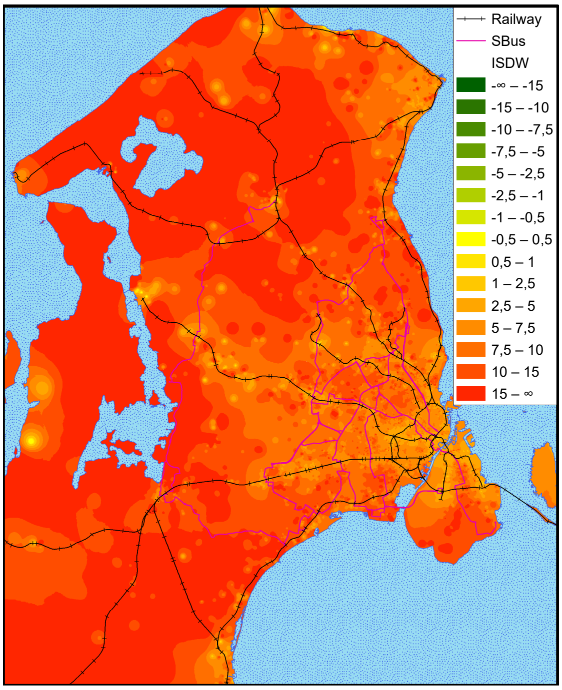


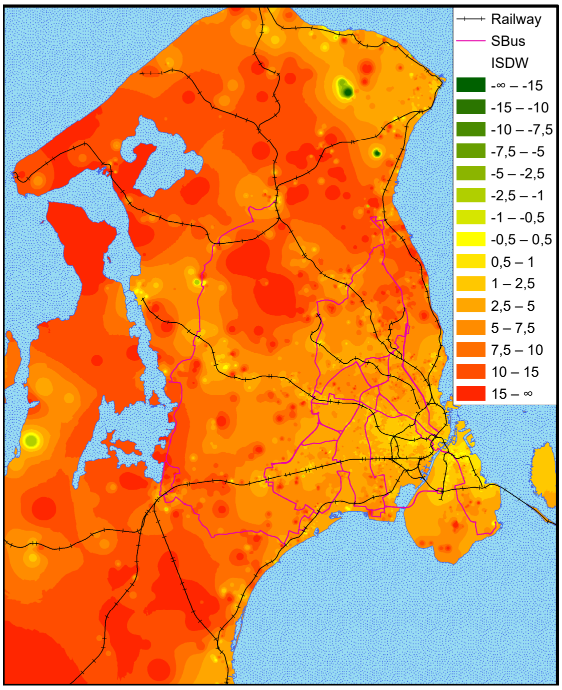
Figure 3: Empirical cumulative distribution function for bus and train arrivals [min.] and passenger delays [gcu] for the four information levels. Negative values imply having lower generalised cost than predicted by the planned timetable for passengers, and arriving early for trains/buses.

From Figure 3 it is seen, that without acquiring any information agents are very unlikely to save lots of time, wheres they are often considerably delayed. The proportion of large delays can be reduced by acquiring pre-trip information, which also facilitates decent probabilities of achieving considerable travel time savings, with 7.5% saving more than 3 generalised cost units (gcu). Such savings occur even more often for R2 and R3, which, interestingly, R2 and R3 are almost indistinguishable.

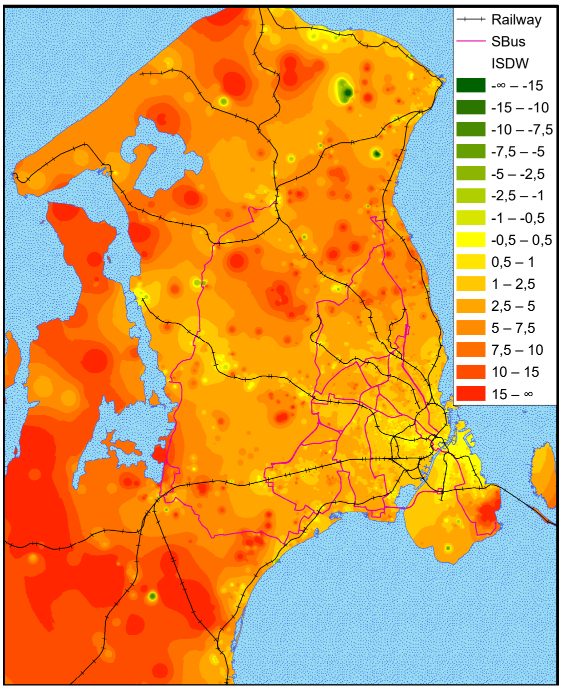
The proposed door to door passenger assignment model allows analysing in great detail variation in passenger delays across origins and destinations. In Figure 4 an inverse squared distance weighting (ISDW) of the average trip delay (across 65 days) of morning peak trips towards Central Copenhagen can be found for each of the four information levels. A dramatic decrease in mean delay is seen when applying information when waiting or everywhere. The entire area seems to improve by receiving real-time information, showing that real-time information is not only an advantage for the urban core, although most effective inside Greater Copenhagen.



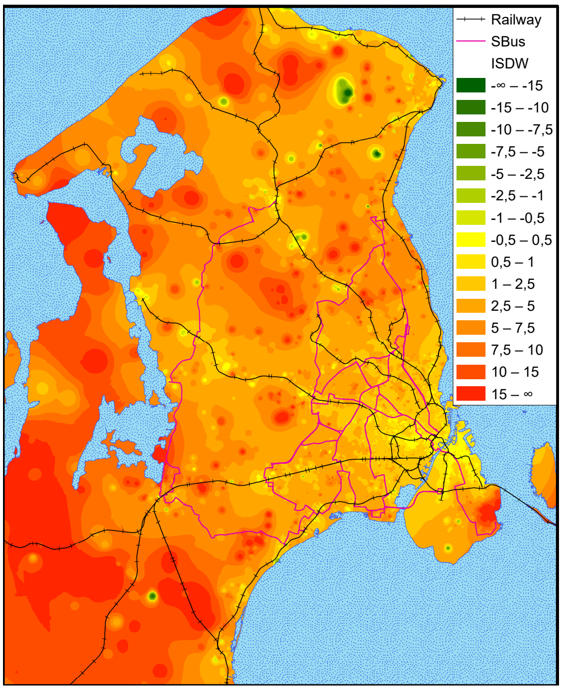
(a) R0 (No info)



(b) R1 (Pre-trip info)



(c) R2 (Info at stops)



(d) R3 (Info everywhere)

Figure 4: Inverse squared distance weighting of the mean of delays of trips departing between 6am and 9am towards Central Copenhagen.

4. CONCLUSIONS

This paper has proposed an adaptive passenger path choice model and a framework that allows large-scale evaluation of passenger delays while considering different levels of real-time information availability. The model was applied to 65 realised days in an agent-based scenario of Metropolitan Copenhagen considering the entire public transport system and real-life AVL data of trains and buses. For each of the days, the five real-time information levels were modelled for 801,719 daily public transport trips with average computation times ranging between one and five and a half hours depending on the real-time information level. The results show that variability of passengers' travel time is considerable and much larger than that of the public transport vehicles. Furthermore, obtaining global real-time information at the beginning of the trip reduces passengers delay dramatically, although still being inferior to receiving such along the trip.

The computation times of the model are low enough to run in real-time or to process the passenger delays of the previous day overnight, allowing operators to gather insights about how yesterday's operation actually influenced the door-to-door trips of passengers. Furthermore, by using the framework with artificial delays as in e.g. Landex & Nielsen, 2006, rather than historical delays, the model can be used to evaluate timetables that have not even been put into operation yet. When used in this regard, the proposed model could be a valuable tool for public transport planners.

REFERENCES

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, pp. 224–239.

Cats, O. & Jenelius, E. (2014). "Dynamic Vulnerability Analysis of Public Transport Networks: Mitigation Effects of Real-Time Information". *Networks and Spatial Economics* 14.3-4, pp. 435–463.

Comi, A., Nuzzolo, A., Crisalli, U. & Rosati, L. (2016). "A New Generation of Individual Real-time Transit Information Systems". In: *Modelling Intelligent Multi-Modal Transit Systems*. Ed. by A. Nuzzolo & W. H. K. Lam. Boca Raton, FL, USA: CRC Press. Chap. 3, pp. 80–107.

Delling, D., Pajor, T. & Werneck, R. F. (2015). "Round-Based Public Transit Routing". *Transportation Science* 49.3, pp. 591–604.

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 for the 14th Conference on Advanced Systems in Public Transport and Transit Data 2018*.

Estrada, M., Giesen, R., Mauttone, A., Nacelle, E. & Segura, L. (2015). "Experimental evaluation of real-time information services in transit systems from the perspective of users". *Proceedings of the Conference on Advanced Systems in Public Transport (CAPST)*, pp. 1–20.

Hickman, M. D. & Bernstein, D. H. (1997). "Transit Service and Path Choice Models in Stochastic and Time-Dependent Networks". *Transportation Science* 31.2, pp. 129–146.

Horni, A., Nagel, K. & Axhausen, K. W., eds. (2016). *The Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press.

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* 19.12, pp. 3603–3613.

Landex, A. & Nielsen, O. A. (2006). "Simulation of disturbances and modelling of expected train passenger delays". In: *Computers in Railways X*. Vol. 1. WIT Transactions on The Built Environment, Vol 88. Southampton, UK: WIT Press, pp. 521–529.

Leng, N. & Corman, F. (2020). "The role of information availability to passengers in public transport disruptions: An agent-based simulation approach". *Transportation Research Part A: Policy and Practice* 133, pp. 214–236.

Nielsen, O. A., Landex, O. & Frederiksen, R. D. (2009). “Passenger delay models for rail networks”. In: *Schedule-Based Modeling of Transportation Networks*. Boston, MA: Springer US. Chap. 2, pp. 27–49.

Nuzzolo, A., Crisalli, U., Comi, A. & Rosati, L. (2016). “A mesoscopic transit assignment model including real-time predictive information on crowding”. *Journal of Intelligent Transportation Systems* 20.4, pp. 316–333.

Nuzzolo, A., Russo, F. & Crisalli, U. (2001). “A Doubly Dynamic Schedule-based Assignment Model for Transit Networks”. *Transportation Science* 35.3, pp. 268–285.

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

Paulsen, M., Rasmussen, T. K. & Nielsen, O. A. (2021). “Impacts of real-time information levels in public transport: A large-scale case study using an adaptive passenger path choice model”. *Transportation Research Part A: Policy and Practice* 148, pp. 155–182.

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

Rambha, T., Boyles, S. D. & Waller, S. T. (2016). “Adaptive Transit Routing in Stochastic Time-Dependent Networks”. *Transportation Science* 50.3, pp. 1043–1059.

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.

Wagner, D. & Willhalm, T. (2007). “Speed-Up Techniques for Shortest-Path Computations”. In: *STACS 2007*. Vol. 4393 LNCS. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 23–36.

Wahba, M. & Shalaby, A. (2011). “Large-scale application of MILATRAS: case study of the Toronto transit network”. *Transportation* 38.6, pp. 889–908.

Yao, X., Han, B., Yu, D. & Ren, H. (2017). “Simulation-Based Dynamic Passenger Flow Assignment Modelling for a Schedule-Based Transit Network”. *Discrete Dynamics in Nature and Society* 2017.

Zargayouna, M., Othman, A., Scemama, G. & Zeddini, B. (2018). “Multiagent Simulation of Real-Time Passenger Information on Transit Networks”. *IEEE Intelligent Transportation Systems Magazine* 12.2, pp. 50–63.