Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 82 (2025) 1965–1973

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2023 Montreal 17-21 July 2023

Using GTFS-data to calculate the roadwork caused delays on public transport network

Riku Viri* & Maiju Örmä

Faculty of Built Environment, Tampere University, Tampere, 33014, Finland

Abstract

The reliability of public transport is a significant reason whether people choose it over other transport modes. Roadworks, i.e., mandatory road maintenance, can cause delays on vehicle journeys in parts of the network, thus generating missed transport connections. To sustain the reliability of the public transport system, it is important to understand effects these delays cause. This study explains a method that allows to model delay effects to the network using GTFS-data. The developed method allows a simple workflow to calculate how different delays on different locations affect the reachability of the areas. The method is tested in two case areas in Finland. Based on the case areas in this study, the results indicate that a 3-minute delay occurring for the departures from city centre only seldomly appears as a total travel time delay longer than the calculated delay. However, for some parts of the city that rely on transfer connections, even a 2-minute delay will result in missed connections, and thus, longer travel times.

Keywords: Public transport; Traffic delay ; GTFS ; Estimation

1. Introduction

Public transport system has an important role in city progress and growth by affecting, for example, urban operating efficiency, carbon emissions and energy consumption (Kaewunruen et al. 2021). Public transport system's reliability is a significant reason in whether people choose public transport over other modes of travel such as personal vehicles (Jamous & Balijepalli 2018; Kaewunruen et al. 2021). Reliable public transport system is dependent on a functional road network. Therefore, to sustain and improve the operability of the road network, roadworks, such as maintenance, are a necessity. However, roadworks in cities can cause major delays in transport around the roadwork area, which

---

* Corresponding author. Tel.: +358 50 301 6167.
E-mail address: riku.viri@tuni.fi

10.1016/j.trpro.2024.12.166

can affect the travel times and route choices (Chow et al. 2013; Jamous & Balijepalli 2018). In addition to personal vehicles, this affects also on public transport, causing challenges not only near the area, but also to the complete, connected public transport network (Jamous & Balijepalli 2018).

For larger scale, scheduled roadworks with longer duration, the effects of roadworks can be estimated already when scheduling the public network. However, for short term or unexpected roadworks, this is not often practical. These unscheduled delays can reduce the public transport network reliability (Kaewunruen et al. 2021). Nevertheless, estimations of future closures can be made by utilising the existing data of the same types of closures if available. For example, Yap & Cats (2021) developed a method in order to predict disruptions to different lines in Washington DC metro system by using earlier disruption logs in addition to data obtainable from GTFS-data (General Transit Feed Specification data). Delays can also be estimated by using GTFS and smart card data for public transport as shown by Guridi et al. (2022) for a public transport network in Montevideo. Historical data can also be used to train regression models to predict upcoming train delays as presented by Wang & Work (2015) using Amtrak delay data for the US rail network. Types of predictions can also be further enriched with different machine learning techniques as Arshad & Ahmed (2019) presented by using weather data in addition to estimated Indian train delays.

In case a closure has already been placed and the change in the network has been made, there are many possibilities to gain information about the effects of the change. For example, Yan et al. (2021) used GTFS delay data as a tool to gather information about the delays a school reopening causes for the whole road infrastructure in Sydney. Therefore, it is important to understand, what kind of effects these delays will result and how much they vary based on the duration of the delay. There are some studies that are already close to this aspect, such as the routing model by Bruglieri et al. (2015) that allows the re-routing of itineraries based on possible closures on different stations or cancellations of different lines. On the other hand, Fortin et al. (2016) have also demonstrated a graph-oriented method for analysing and comparing GTFS-datasets.

However, as in the modern world, as this type of data is more and more generated and openly shared, it can be used to gain knowledge and understanding about the possible outcomes that roadworks can cause to the public transport network. GTFS-data (Google 2021) that was used in multiple of the mentioned studies, has been created to share the public transportation data and its associated geographic information in a defined format. GTFS-data is a globally utilized universal format that contains the information about the whole public transport network. GTFS-data offers an appropriate source to calculate network effects also for the aim for this study.

The objectives of this paper are to present a method on how open public transport schedule data can be used to estimate the network-wide effects of different delays, and to analyse how these calculated delays can affect the network-wide results. More closely, we calculate the effects of roadwork caused traffic delays on public transport by using open GTFS-data. The calculation is executed for two different case areas and evaluated with different delay times.

2. Methodology

The included steps in the methodology are shown in Figure 1. First the route level data is mapped to analyse which lines will have delays. Then the stop times in GTFS-data are edited based on the existence of the delays. This new, modified GTFS-dataset is then compared with the original GTFS-dataset to see what differences are caused by the added delays. All these steps are further explained in this chapter.

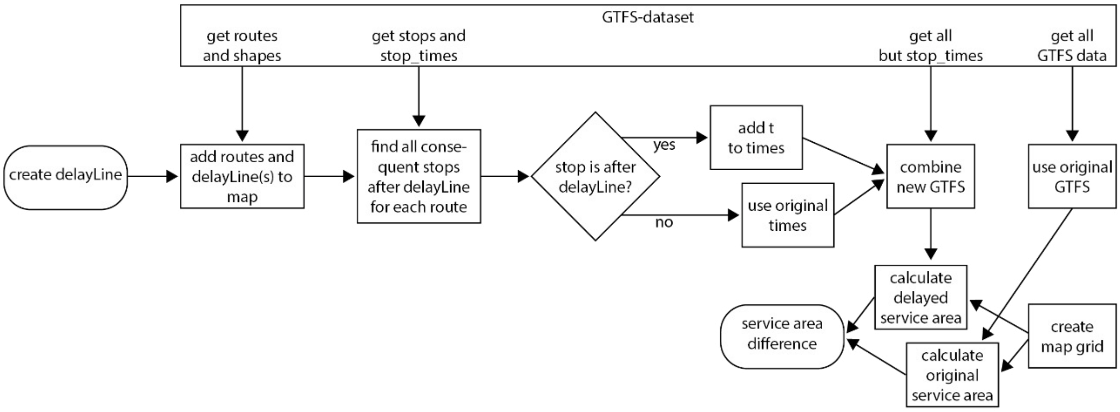

Fig. 1. The overall view of the method used in the study.

GTFS is a format for providing machine-readable public transport routing and scheduling data. In addition to routes, stop locations (stored in *stops*) and departure times from stops (stored in *stop_times*), the dataset contains information about the actual route the vehicle takes between stops (Google 2021). This allows to draw the route on a map. This data is stored in *shapes*. By using this graphical representation of the routes, it is possible to track, which lines use which streets. Shapes consist of sequential point-to-point lines that form the complete line route. Between line stops, there are usually multiple sequential shapes that represent the real, street following route of the line between stops.

To estimate how different types of road closures and delays would affect the network, first a delay area is introduced. For this study, a delay is represented as a line over a street and every route that travels through that line (i.e., uses the street the delay is on) is estimated to have a delay of *t* for all stops after crossing the line. By using the shapes, it can be resolved, on which lines the delays will occur, and on which sequence of that line. However, since the departure times are included in the *stop_times* and there is no direct connection between *shapes* and *stop_times*, stop locations (from *stops*) are used to calculate, which is the closest stop to the location of the delay to every line that has a delay (i.e., crosses the delay line). When the closest stop for these lines is found, its departure time can be adjusted to contain the delay of *t* for that stop and every stop after that (i.e., stops having larger sequence number). Figure 2 below presents a line from where the delay is calculated. It can be noted that for the calculation the delay does not occur when the line is crossed, but on the next stop and the stops after that (as indicated by a red line).

Fig. 2. Delays are calculated directionally starting from the first stop following the calculated delay.

However, as in Figure 2, only bus lines are present, the delay is calculated for every line crossing the delay line. To calculate roadworks only on a certain section, the *route_type* in routes can be also used to track the public transport mode and to limit the delay only to some. In Figure 3 below, the delay line crosses both bus and metro lines, but while the calculated the delay is located on the bridge, the delay can be only limited to the bus lines, allowing the metro to have its original timetable for the calculation.

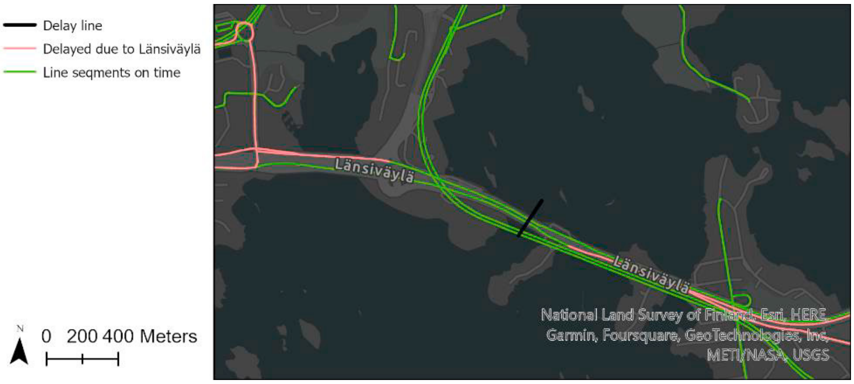

Fig. 3. Delay is limited only to bus lines, while metro has no delay in calculation.

It should be noted that at this point, this model assumes there is more turnover time than delay. Thus, it estimates that next route for the vehicle departs without any delay, and will only gain the delay, if it crosses the delay line during the route.

Based on the calculation, new departure times can be calculated for every stop of every route in the dataset. These new departure times can be used to replace the times in *stop_times* of the GTFS-dataset, resulting in a GTFS-source that shares all the definitions of the original, but with the added delay of *t* for those parts of the routes that have crossed the delayed section.

As the data still respects the GTFS guidelines, any workflow that can calculate travel times using GTFS-data can be used to calculate the effects of the delay. For this study, ArcGIS Network Analyst was used. It allowed to calculate the service area (i.e., time taken to reach other points of the network) for both the original, and the modified version of the data. When the same parameters are used for both, the results can be used to compare the differences in time taken for travel. Since the service areas are formed by the time taken to reach the areas of the network, they can result in a unique shape for every analysis and would be hard to compare directly against each other. Therefore, in this study, the results were laid on the Finnish national 250m×250m statistics grid (SYKE 2021) that allows the comparison between runs, as the comparison can be made in time taken to reach the grid cells. Grids were also used for comparison as some areas may be reachable through multiple different stops, and thus there could be a time difference to reach a certain stop, but the area could still be reached as fast as before by using a different line with a different stop.

For this study, a regular Tuesday was chosen as the day of the analysis for both case areas. For departure times, every minute from 7:50 AM to 8:10 AM were calculated, and the fastest time to reach any given grid cell (in both the original and the modified dataset) was stored to respect the time taken. The fastest of the 20-minute time frame was used to not give any unfair advantage to either of the networks, but to see the actual fastest time during morning to reach the destination grids. These results can be used to calculate the time difference between the networks. In Figures 4 and 5, examples from Tampere Region, Finland, are presented, using GTFS-data provided by Nysse (2021).

To see what kind of results the method could present, two case areas were chosen. Both areas were chosen in Finland, as the data availability was known. First case area was Tampere region (GTFS-data provided by Nysse 2021) and the second area was Finnish capital region (GTFS-data provided by HSL 2022). For both areas, the Finnish national road data (Digiroad 2022) was used as part of the analysis to form the walking network around the stops, as Digiroad-dataset has the spatial data of all Finnish roads and pedestrian paths. The transfers between stops and reaching the areas by walking were calculated by the shortest possible route using the Digiroad network with the walking speed of 5 kmph.

In Tampere region, as first case area, two different studies were conducted. As the first one, the calculated delay was placed on Hatanpään valtatie, which is the main southern corridor to the city centre of Tampere. As the second case study, the calculated delay was placed on Pirkankatu, which is one of the two main western corridors to the city centre. For second case area, the Finnish capital region, only one case was evaluated, but this time with two delays occurring at the same time. One located in Länsiväylä, which is the main western road corridor from Helsinki city centre and other located at Mannerheimintie, which is one of the main northern road corridors from Helsinki, taken by many public transport routes towards north-western parts of the capital region. Here, the delay on Länsiväylä was only limited to the bridge, so buses had delay, but the metro ran unaffected. The block on Mannerheimintie affected all modes, meaning buses and trams in that location.

To estimate the correct delay of t for the calculation, Kaewunruen et al. (2021) present results from models in a study in the UK indicating roadworks causing 82.15- and 123.12-seconds delays for different directions in their case study. In another simulation study in the UK, Jamous & Balijepalli (2018) noted the mean delays of as high as 100 seconds for the worst calculated origin-destination pairs with directional closures and 149 seconds with the total closures of certain connections.

3. Results and main contributions

The network in Tampere (first case area) is mostly based on buses and there are only some areas where public transport has its own segregated paths. Therefore, minor delays may occur based on other road users and a 1-minute delay would not be important to investigate, as there are no planned transfers with as short of a transfer window. As both Kaewunruen et al. (2021) and Jamous & Balijepalli (2018) present possible results of higher than 2-minute delays, a 3-minute delay was chosen for inspection for both case studies. However, since there is also evidence of shorter delays, the latter case is also run with 2-minute delay. To see whether the results would change significantly with a minute longer delay, a 4-minute delay is also investigated in the second case. For Helsinki (second case area), a 2-minute delay was used for both locations.

3.1. Case 1 – Hatanpään valtatie

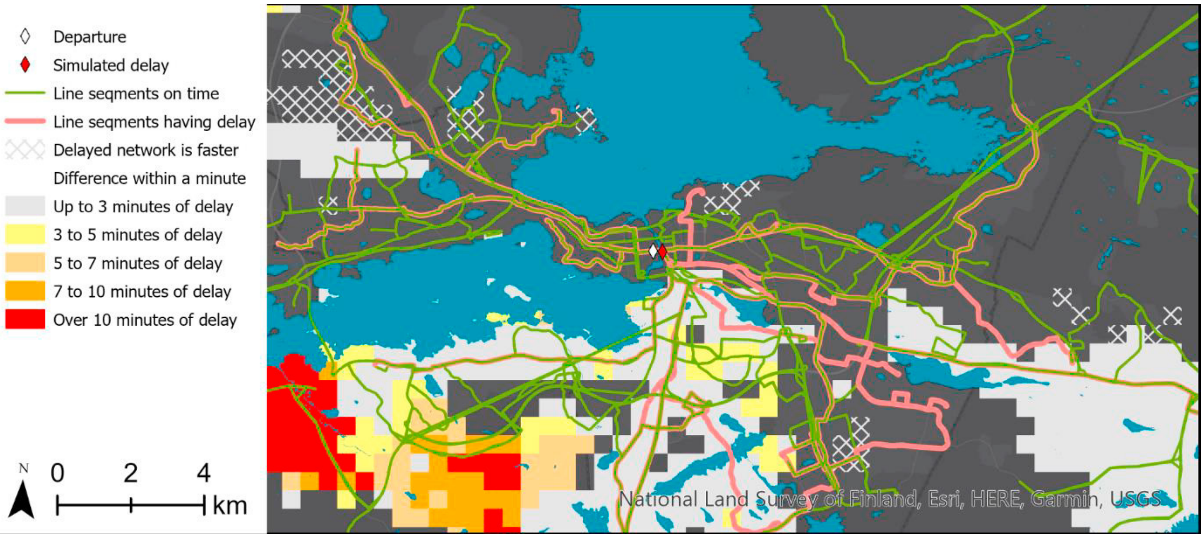

Fig. 4. Difference of time taken to destinations from Keskustori, Tampere when Hatanpään valtatie has a 180-second delay.

As seen in Figure 4, travelling to most of the southern parts of the city are affected by the delay. Some areas, as indicated in grey, have only a 3-minute delay caused by the vehicle going through the delay location. For some areas (indicated in yellow to red hues), the actual delay is larger, since the 3-minute delay of certain vehicles did not allow certain transfers between lines to happen, causing a larger delay. There are also locations, as indicated in cross-hatched overlay that can be reached faster. This is possibly due transfer that is normally not possible, but due to only the second vehicle being late, it is possible in the delayed network. This seems to apply to the north-western parts of the city, as the lines are already delayed before the departure point. The north-eastern part of the city is virtually unchanged, as only a few lines from the city centre go through the delayed area and, therefore, the service appears to be normal.

3.2. Case 2 – Pirkankatu

Fig. 5. Difference of time taken to destinations from Keskustori, Tampere when Pirkankatu has a 180-second delay.

In the case of a delay in Pirkankatu the largest changes will happen in the western parts of the city, as the delay occurs on one of the main public transport routes from the city centre to the west. On the large share of areas, the total delay is the size of the calculated delay, as towards west there are only a few areas need transfers. These transfers would be missed on the delayed network. However, the same south-western areas that are affected in the first case, are affected here as well. As the trunk line has the 3-minute delay, some transfers are missed, and, therefore, the delay to reaching these areas is larger. However, some transfers here are faster, since there has been a 6-minute waiting time to transfer that now drops to 3-minute waiting time. As only the first line has a delay, the actual connection is faster. The same also applies in the areas that can be reached faster on the north-western parts. In the eastern parts, there is not a large difference. As the rural areas in the north-eastern parts of the city are serviced through a transfer, there seems to be some delay to these areas.

3.3. Sensitivity analysis for previous cases

When a 120-second delay is applied on the case 2 instead, the results change a bit. For both north-western and north-eastern areas, the results remain the same. Mainly, the results only change by the duration of the delay and some transfers are, therefore, missed causing a larger delay in the same areas as previously. However, there is a large shift in the south-western areas. While the results of the far south-western cluster remain the same, the eastern cluster disappears, and instead of delays, it is reachable even faster than in the regular, not delayed network. It seems that if the trunk line to the area runs 2 minutes late, the transfers are not yet missed (as with a 3-minute delay), but happen without any waiting time, thus creating a faster journey.

When a 240-second delay is applied, the results seem to be aligned with the 180-second delay. For the areas with a 3-minute delay in Figure 4 are now reached with a 4-minute delay, and for the areas where there is a connection and larger delay in Figure 4, the same applies with a 4-minute delay. However, there are still areas (as in Figure 4) in south-western parts that can be reached faster, as even the 4-minute delay is not large enough to result in missing some transfers.

It seems that in general, at least for this case, the areas with delays do not drastically change when the delay is slightly adapted, but the delay in certain central parts can affect large areas. Since the case area of the study, the centre of Tampere, is located in a narrow land area between two lakes, the number of connecting streets to the city centre are highly limited from the west, thus not allowing the routes to be divided more into different streets. Still, if the network is investigated from the city centre, the delays seem to only affect their part of the city. However, since a ±10-minute time elasticity was allowed in departure time, an already delayed route at the departure point would not affect the time used to travel. Without this departure elasticity, there would have been larger areas with delays, but the results would have been more difficult to interpret, as some differences would only be caused as timetables for some lines would favour either the regular or the delayed network.

3.4. Case 3 – Länsiväylä and Mannerheimintie

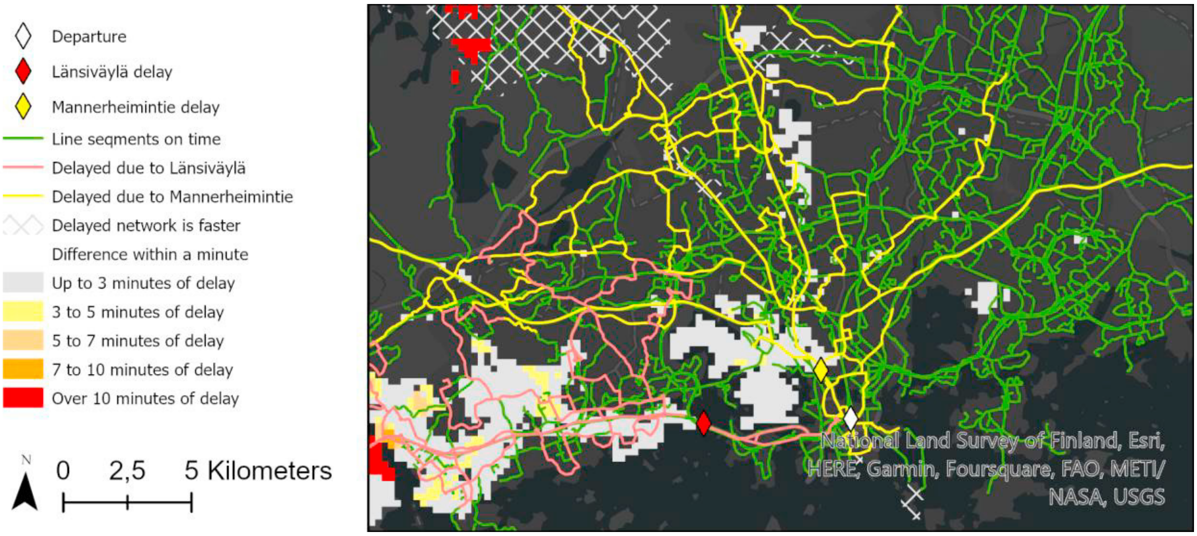

Fig. 6. Difference of time taken to destinations from front of main railway station, Tampere when Länsiväylä (red diamond) and Mannerheimintie (yellow diamond) both have a 120 second delay.

For the analysis in Helsinki region, two simultaneous delays are placed, one in Länsiväylä only affecting to bus transport and one in Mannerheimintie affecting all modes. Both areas are calculated to cause the same 2-minute delay to public transport. At first, many western destinations from the city centre will have a slight delay, which most likely just occurs about the vehicles arriving a couple of minutes later. However, since the metro line is calculated not to have any delay, the areas next to the metro stations or areas connected through transfers by metro do not have any delays. However, this only applies to areas right next to the stations, the direct bus is a somewhat faster as the metro does a little detour through Otaniemi area. There are also some transfers to the suburbs of Espoo, where the first connection is missed, and thus, the delay increases slightly.

For the northern parts of the region, the delay on Mannerheimintie seems to have a much lesser effect on the areas. Even as Mannerheimintie is one of the major corridors, there are also some other streets that carry the public transport to the northern areas, meaning only some small areas are directly affected. It can also be seen that some north-western areas are faster to reach. This is probably due to the commuter train running on schedule and allowing transfers to buses that run late, which would otherwise be missed. All in all, there are very few areas that will have effect larger than the delay itself for the Helsinki region case.

4. Conclusion and future works

The method presented in this study allows estimating how certain delays, such as those caused by roadworks, will affect the whole network. The method allows to easily vary both the location, the duration, and the amount of the delays to estimate the total network effects. Based on this method, it is possible to investigate, whether it is profitable to use resources to try to cut the duration of the delay caused to the vehicles, or whether a three-minute delay has the same effects as two-minute delay would. However, it should be noted that this method only pays attention to the scheduled timetables and the actual results may vary. There may, for example, be other random delays that are not known, or some parts of the journey, where the vehicle is faster than the schedule, allowing it to catch up the time lost in the delay. Also, as the data does not have the clear information of which journeys share the same vehicle, it could not be estimated, whether a delay on a part of the network would continue for the next route of that same vehicle. In this study, it was assumed that the turnaround time of each vehicle is larger than the calculated delay, and therefore, the vehicle can always start the next journey on time despite its late arrival in the last stop of the route. In addition, GTFS-dataset has information about the direction of the vehicle routes. This would allow the delay to be limited to only lines heading in a certain direction, i.e., westbound traffic will be delayed, but northbound traffic is not. In this study, however, a bidirectional delay was assumed in all cases.

In both calculation cases for Tampere region in this study, the central-departure results followed the same pattern. For most of the areas where the delay affects, the time taken to travel is only delayed by the amount of the calculated delay time of t. For both cases, the Pirkkala-area south-west from the city centre stands out, as some areas in there are reachable fastest with a transfer when departing from the city centre. The delay may deny the use of the insisted transfer, thus forcing the passenger to have either longer waiting times or longer walking distances. In this case, the sensitivity analysis of a minute did only affect the results minorly, as in most where the delay occurred, the actual delay was the same as the calculated time. For Helsinki region, the results were even smaller, as most of the areas having delays, only had the delay time of t without any compounding effects, even though there were two separate delays in place. However, it should be noted that for further research, also departure locations different from the city centre should be used. This could vary the results, due to many of the transfers occurring while travelling through the centre, thus possibly generating more missed transfers and longer delays.

The work presented here is based in Finland and the data available in Finland, but generally the concept is expandable to any city or region providing the same type of data. It should, however, be noted that some areas may use for example timed points along the route (i.e., stops where the vehicle will wait, if it is ahead of the timetable), which this method does not take in the account, as it is not known based on the data, how much the vehicle may be ahead on these time points, and therefore, it cannot be estimated how large of a delay could be spurted. For this study, these time points are not generally used on the networks, so this should not cause any major challenges with the results.

The GTFS-data also supports the definition of forced transfers (i.e., if the first vehicle for the connection is delayed, the second one will wait for it), which were not used in either of the case networks. However, since the method only modifies the stop times of the original GTFS dataset, these should be supported for the analysis, if the selected analysis method for the GTFS dataset supports it. It should also be noted that for Finnish cases, Digiroad (Digiroad 2022) was used as the main network to estimate walking and transfers, but that type of data may not be generally available for other areas. However, OpenStreetMap should have the same type of spatial definition, so an OSM-extraction should work for the walking and transfer calculation in the same manner as Digiroad.

Currently, the approach is only limited to the use of the GTFS-data. However, through the data collected by other actors in transport, the calculation could be further developed. Different measuring systems, such as traffic lights can collect information on delays or level of congestion directly from the network in real time (see: Tampere 2022). The real time location of buses can also be tracked to measure the actual caused delay (see: ITS Factory 2022) as well as to estimate how much the lines could have ability to spurt on when timed points are used. These could be used to better adjust the delay and calculation method used in the later calculations.

Acknowledgements

Declaration of interest: none.

References

Arshad, M. & Ahmed, M. (2019). Train Delay Estimation in Indian Railways by Including Weather Factors Through Machine Learning Techniques. Recent Advances in Computer Science and Communications. 12.

Bruglieri, M., Bruschi, F., Colorni, A., Luè, A., Nocerino, R. & Rana, V. (2015). A real-time information system for public transport in case of delays and service disruptions. Transportation Research Procedia. 10. 18th Euro Working Group on Transportation, EWGT 2015. pp 493-502

Chow, A. H.F., Santacreu, A., Tsapakis, I., Tanasaranond, G. & Cheng, T. 2013. Empirical assessment of urban traffic congestion. Journal of advanced transportation, Vol 48, pp. 1000-1016.

Digiroad. 2022. Digiroad data. Finnish Transport Infrastructure Agency. Online, available at: https://vayla.fi/en/transport-network/data/digiroad/data, cited 24th November 2022

Fortin, P., Morency, C. & Trépanier, M. (2016). Innovative GTFS Data Application for Transit Network Analysis Using a Graph-Oriented Method. Journal of Public Transportation. 19(4). pp. 18-37

Google. 2021. Reference | Static Transit | Google Developers. Online, available at: https://developers.google.com/transit/gtfs/reference, cited 24th November 2021.

Guridi, S., Olmedo, S. & Nesmachnow, S. (2022). Big data analysis for delay estimation of public transportation systems. Proceedings of the V Ibero-American Conference on Smart Cities (ICSC-CITIES 2022).

HSL. 2022. Open data, Public transport network and timetables (GTFS). Available at: https://www.hsl.fi/en/hsl/open-data

ITS Factory. 2022. Tampere Public Transport SIRI Interface. Available at: http://wiki.itsfactory.fi/index.php/Tampere_Public_Transport_SIRI_Interface_(Realtime_JSON_at_data.itsfactory.fi)

Jamous, A. & Balijepalli, C. 2018. Assessing travel time reliability implications due to roadworks on private vehicles and public transport services in urban road networks. Journal of Traffic and Transportation Engineering (English Edition), Vol. 5(4), pp. 296-308.

Kaewunruen, S., Sresakoolchai, J. & Sun, H. 2021. Causal analysis of bus travel time reliability in Birmingham, UK. Results in Engineering, Vol. 12, 5 p.

Nysse. 2021. Tampere Public Transport GTFS feed, dataset accessed through ITS Factory Data Sources at http://data.itsfactory.fi/, cited 24th November 2021.

SYKE. 2021. YKR-aluejaot, Suomen ympäristökeskus. Available at: https://www.opendata.fi/data/fi/dataset/ykr-aluejaot

Tampere. 2022. Tampere Traffic Lights API, Open Data Tampere. Available at: https://data.tampere.fi/data/en_GB/dataset/tampereen-liikennevalorajapinta

Wang, R. & Work, D. B. (2015). Data driven approaches for passenger train delay estimation. 2015 IEEE 18th International Conference on Intelligent Transportation Systems

Yan, W., Yao, H., Chen, L., Rayaprolu, H. & Moylan, E. (2021). Impacts of School Reopening on Variations in Local Bus Performance in Sydney. Transportation Research Record. 2675(9). pp. 1277-1289

Yap, M. & Cats, O. (2021). Predicting disruptions and their passenger delay impacts for public transport stops. Transportation 48, pp. 1703-1731