Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Big data processing and analysis on the impact of COVID-19 on public transport delay

Yuming Ou, Adriana-Simona Mihăiță, Fang Chen

FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY, UNIVERSITY OF TECHNOLOGY SYDNEY, SYDNEY, NSW, AUSTRALIA

1. Introduction

The coronavirus disease 2019 (COVID-19) pandemic triggered at the beginning of the year 2020 [1] has significantly affected people’s daily life around the world in all aspects: personal transport mobility (local, national, and international), daily working schedule, tourism activities, medical operating conditions, etc. Understanding and quantifying the overall impact of such a large-scale disruption will help citizens to mitigate the pandemic and enhance their resilience for future preparation of similar events. To this end, in this chapter, we present the research work that we conducted to study the impact of COVID-19 on public transport at different stages in different areas across a large metropolitan area from Sydney, Australia, particularly in terms of bus delay. The public transport infrastructure in Sydney is very important, as it connects radially the inner and east regions of the city to the large extended regional suburbs. As the city does not have an underground subway system like other large metropolitan cities from Europe or the United States, the bus public transport system is critical in delivering the daily home-to-work journeys. The results not only help understand the changes of travel behaviors caused by the pandemic across the entire city but also benefit the public transport operators to enhance services after the pandemic is over, by providing a benchmark of potential improvement.

In Australia, a few cases of COVID-19 were initially reported before March 2020. However, the number of confirmed cases has rapidly increased during the month of March. There was a signalized spike of reported cases during March 22, 2020, to March 27, 2020, in the state of New South Wales (NSW). Based on this observation, we hypothesize that under the situation of rapid increase of confirmed cases, people will largely reduce their usage of private cars and public transport by working from home and avoiding unnecessary travel. The decision of reducing daily trips was also recommended...
by government authorities to keep social distancing and work from home where possible [2]. The travel behavior shift has led to a change of traffic conditions, which had no precedent. Therefore the impact of such travel restrictions can be quantified in terms of bus delay, which is an important indicator of the overall urban traffic condition. With this hypothesis, we decide to scope our study on the bus delay change in the Sydney region within the time window from February to March 2020. As an observation, previous years of bus delay analysis have been captured continuously by the research team, but for the purpose of this work and case study, we keep the data analysis restricted around the most affected months by the global COVID-19 pandemic. We do make the observation that while great importance has been previously given in the literature to the prediction of bus delays in the network, the current large-scale disruption of COVID-19 represents a unique event that has not been studied before; therefore the challenge is to model and understand the impact of such disruption on not only bus operations but also the traveler’s behavior.

Besides the difference of bus delay between February and March, we also aim to study the public transport differences recorded around different urban areas, in order to have a more comprehensive understanding of the COVID-19 impact on urban mobility. In Sydney, the central business district (CBD) and eastern suburbs are the very popular locations for business, traveling, transport multimodal hubs, and tourism attractions due to their proximity to beaches. This makes the travel behavior in these areas distinct from that in other areas in the city, which have different economical and urban characteristics (several warehouses, depots, remote industrial headquarters, local suburban houses). Consequently, the impact of COVID-19 on these areas is expected to be different from the remote areas, especially during peak hours.

In order to quantify the bus delay changes, the biggest challenge is to correctly and timely estimate the bus arrival times at various bus stops in the network. The information of bus arrival time is currently unavailable, as the bus stations are not interconnected or monitored digitally. However, the real-time positioning of each bus across the network is available via real-time transmitted GPS (Global Positioning System) location points. We therefore need to estimate the bus arrival times by using the real-time transmitted GPS data, which contains the geolocations of buses in the city with their associated time stamps. Nowadays the GPS devices mounted on buses are continuously generating data within a very short time interval of a few seconds, with the purpose to better monitor bus movements. This makes the accurate estimation of bus arrival time possible. On the other hand, due to the extensive and large bus network, this requires having in place a big data processing and analysis framework to run on such a large-volume of geolocation dataset; for example, more than 3 GB of GPS data is generated every day for the entire Sydney bus network. Furthermore, due to the well-known issue of GPS accuracy [3], the GPS data is always associated with an error (a deviation from what the real position of the bus vehicle is in reality). Therefore we are
facing the challenge of mapping the GPS data points to the correct geolocations in order to be able to calculate accurately the individual bus delays and aggregate the results for impact analysis.

This chapter is organized as follows. In Section 2, we explain the datasets used for this research and the methods used to collect the data, after which we propose a methodological framework of data processing and analysis to quantify bus delays in Section 3. Section 4 reports a case study on the Sydney metropolitan region across different stages of the COVID-19 pandemic. Finally, we conclude this chapter and provide some directions for future work in Section 5.

2. Data preparation

The data used in this research work is heterogeneous and comes from multiple sources, including COVID-19 case data, bus GTFS (General Transit Feed Specification) data, and LGA (Local Government Area) boundary data. As shown in Table 14.1, the data that we are using has several characteristics: it has a spatiotemporal distribution (data points are transmitted with high time frequency across a very large area of coverage) and arrives in real time and in large volume; this brings various challenges to data processing and analysis. In the following, we detail the abovementioned three types of data and their corresponding collection methods.

2.1 COVID-19 case data

COVID-19 case data contains the information for all confirmed cases in NSW, Australia, including the notification date, the LGA code (where it happened in the city), and the LGA name. The data is available on the official government website Data.NSW [4] and is updated on a daily basis. A sample of the COVID-19 case data is shown in Table 14.2.

To understand the trend of increased confirmed cases, we group the data by their notification date and visualize the results in Fig. 14.1. As shown in Fig. 14.1, there were 10 confirmed cases before March 2020. Starting with March 1, 2020, the number of confirmed cases was rapidly increasing and reaching almost 200 cases daily. The peak occurred during the time window from March 22, 2020 to March 27, 2020 while slowly decreasing afterward.

| Table 14.1 Datasets used in this research work. |
|-----------------------------------------------|
| Data                          | Temporal | Spatial | Real time | Large volume |
| COVID-19 data                 | Yes      | Yes     |           | Yes          |
| Bus GTFS data                 | Yes      | Yes     | Yes       | Yes          |
| LGA boundary data             | Yes      |         |           |              |

COVID-19, coronavirus disease 2019; GTFS, General Transit Feed Specification; LGA, Local Government Area.
Table 14.2  Sample of COVID-19 case data.

| Notification date | LGA code | LGA name         |
|-------------------|----------|------------------|
| 2020-01-22        | 11300    | Burwood (A)      |
| 2020-01-24        | 16260    | Parramatta (C)   |
| 2020-01-25        | 14500    | Ku-ring-gai (A)  |
| 2020-01-25        | 16550    | Randwick         |

COVID-19, coronavirus disease 2019; LGA, Local Government Area.

FIGURE 14.1 Epidemiologic curve of confirmed coronavirus disease 2019 (COVID-19) cases by notification date, New South Wales, Australia.

2.2 Bus General Transit Feed Specification data

GTFS [5] defines a common data format to allow public transit agencies to publish their transit data so that the data can be consumed by various applications. Generally, GTFS is divided into GTFS static and GTFS real-time streams. The former contains public transportation schedules and associated geographic information, while the latter contains the real-time vehicle positions and all trip updates.

Nowadays GTFS has been used as an industry standard for a majority of transit agencies to publish their transit data around the world [6]. As GTFS data contains both scheduled and real-time information about transit operations, it has been actively used for many research problems such as transit accessibility [7–11], transit network analysis [12,13], performance evaluation [14,15], delay prediction [16–18], and transit trip inference [19,20].
In this research, we use the following three types of GTFS data:

- **Real-time bus position data**: the real-time buses’ movements with longitudes, latitudes, and associated time stamps. The real-time bus positions are captured by the GPS devices mounted on the buses. As mentioned before, there are always errors associated with the GPS data. We need to correct the GPS data by a map matching algorithm, which we introduce in Section 3. Fig. 14.2 shows an example of how such GPS point coordinates can be transmitted, with some of them deviating from the road centerline.

- **Bus timetable data**: contains the scheduled bus trips and scheduled arrival times at bus stops. Fig. 14.3 showcases an example of how a bus timetable is being captured in the urban region; the $x$-axis represents the timeline from the departing time of the bus trips of the bus line 135 (in this case scheduled to depart at 09:55 a.m.) and until the end of the bus stop (10:26 a.m.), while the $y$-axis represents the distance traveled since the departure of the bus. The green (black in print version) line showcases the planned timetable trips, while the orange (gray in print version) line represents the real-tracked movement of the bus during operation. One can notice that the bus in this example was always on time and even arrived earlier until 10:06 a.m., after which it started to record accumulated delay until the end of the trip. Similarly, we conduct the same visualization and tracking for all

![FIGURE 14.2 Example of GPS (Global Positioning System) point coordinates transmitted across a transit line.](image-url)
bus trips across the entire network. The difference between the scheduled and the real-tracked movement of the bus at each stop is identified as the “bus stop delay”.

- **Bus network data**: contains the geolocations of all bus stops (as represented in Fig. 14.4) and the physical geometry (layout) of the bus routes (as represented in Fig. 14.5).

As the bus GTFS data is published in real time, we need to develop a data crawling service that is continuously collecting the data through RESTful data application programming interface (API) provided by the local transport agency: Transport for NSW [21]. Fig. 14.6 illustrates the workflow of data crawling. To collect the real-time bus position data, the data crawling service sends a data pulling request to the data APIs every 5 s. After receiving the data returned from the data API the service then parses the data and checks whether it is exactly the same as the previous data points by examining the hash values. If so then it discards the data, otherwise we append the data to stored data files. The purpose of removing the duplicate records is to save space as well as to reduce the computation cost in the step of delay calculation introduced in Section 3. In the entire Sydney metropolitan region, there are around 24,000 bus stops and more than 25,000 bus trips are being scheduled during a 24-h day, which leads to more than 3 GB of real-time bus position data being collected every day. Apart from the bus position data, the data crawling service also collects timetable and network data daily in a similar fashion, in order to have up-to-date timetable and network data.
2.3 Local Government Area boundary data

The LGA boundary data is downloaded from the Australian Bureau of Statistics website [22], containing the information of LGA geometries and area catchment, which plays the important role to link COVID-19 confirmed cases data to bus data in terms of geolocation. The LGAs in Sydney metropolitan region is shown in Fig. 14.7.

3. Methodology

In this section, we propose our methodology to quantify bus delays based on the collected data described in Section 2. We first introduce the framework and explain the workflow of each step and then provide the details of each step.

3.1 Framework

Fig. 14.8 illustrates our approach to quantify the bus delays, in which every blue (gray in print version) block represents a main algorithmically processing step, every green
(light gray in print version) block is a raw dataset that was collected, and each orange (dark gray in print version) block is the output of each processing step. They are explained as follows:

(1) *Bus GTFS data crawling.* As introduced in Section 2, a data crawling service continuously collects bus GTFS data through data APIs and appends the data to
files. The data collected includes bus position data, bus timetable data, and bus network data as previously described. The data crawling service is running as a standalone process and is independent of other steps.

(2) **Map matching.** For each bus trip that is being collected, bus GPS points are matched to the actual geolocations in the bus network by using our proposed map matching algorithm. This step represents a true research challenge as the GPS points’ deviation can be quite high especially in central urban areas where various complex intersections can be found. Matching the bus GPS locations not only needs to consider the proximity to the road centerline, but more specifically, the direction of the bus trips, the turnings available at each intersection, the last GPS points, the overlaying with other bus trips sharing the same lanes, etc. The method is further described in Section 3.2. The output of this step is the corrected bus geolocations with associated time stamps.

(3) **Bus stop extraction.** This step is needed in order to extract the bus stop geolocations from bus network data collected by the data crawling service. The extracted bus stop geolocations are used in the step of arrival time estimation (step 5).

(4) **Scheduled arrival time extraction.** This step is further applied in order to extract the scheduled arrival time at each bus stop for each bus trip from the bus timetable data collected by the data crawling service. The extracted scheduled arrival times are used in the step of delay calculation (step 6).
Arrival time estimation. Based on the bus geolocations with time stamps obtained in step 2 and the bus stop geolocations obtained in step 3, we track the bus movements and estimate the actual arrival times at each bus stop. The method to estimate the arrival time is detailed in Section 3.3.

Delay calculation. By checking the arrival times against the scheduled arrival times, we calculate the delays at bus stops as the difference between planned and actual arrival time. The idea of delay calculation is initially introduced in Fig. 14.3, and we provide the formalized calculation in Section 3.4.

Delay aggregation. Finally, the delays at every bus stop for all trips are aggregated to each LGA level based on their geolocations and time stamps’ time windows, utilizing the LGA boundary data. The aggregation method is detailed in Section 3.4.

3.2 Map matching

As mentioned before, GPS data has an inevitable error that is variable depending on the circumstances, the road network geometry layout, and the continuity of data.
transmission in real time. Many other sources could contribute to GPS errors, such as clock error, signal jamming, weather, and building blocking. An example of GPS errors is shown in Fig. 14.9 in which the dots are the GPS data points sent from the GPS device on a bus while the green (black in print version) line is the actual bus trajectory along the main road. It can be observed that many GPS locations are falling further away from the
green (black in print version) line (road centerline) instead of exactly being on it. Consequently, before using the bus GPS data to estimate arrival times, we need to correct the GPS data through map matching algorithms by matching every GPS coordinate transmitted by the bus to a correct location on the road centerline.

There are various methods that have been used in the literature for map matching [23–26]. One native way is the point-to-curve method, which projects GPS points to their closest edges. This method is simplistic and lacks robustness especially when the road network has a complicated structure such as inside the CBD. An improved method is the curve-to-curve method, which considers the closeness and similarity between the curve formed by GPS points and the candidate path. However, it still has the same problems under the circumstances of large GPS errors and complicated overlay networks. Other approaches include using the geometry and topology of the road network [27], Kalman filters [28], and Fuzzy rules [29]. Generally, map matching algorithms can be divided into two groups: offline and online algorithms. Offline algorithms have an advantage of

**FIGURE 14.9** An example of Global Positioning System (GPS) errors.
processing all GPS points after a trip is finished. On the other hand, online algorithms need to process currently available GPS points before a given time, which makes them have less data to be used and potentially leading to a compromise of accuracy.

To achieve a high accuracy of GPS data correction in real time, our map matching method is based on a Hidden Markov model (HMM) \[30,31\]. HMMs usually model a system by considering their unobserved states and their observations. In the system, one hidden state can change to any other hidden state by following a state transition probability. Instead of the hidden states, one can observe the values generated from the hidden states with emission probabilities. In this work, we model the road segments on which the bus is as the hidden states and the GPS readings as the observations, as shown in Fig. 14.10. Under this setting, the emission probability is defined in Eq. (14.1) as

\[
P(GPS_t | Seg^i_t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(gps^i_t - gps_t)^2}{2\sigma^2}}
\]  

in which $GPS_t$ is the bus GPS reading at time $t$, $Seg^i_t$ is the road segment $i$ that the bus is on at time $t$, $gps^i_t$ is the projection of $GPS_t$ on $Seg^i_t$, $gd$ is the great circle distance between two geolocations, and $\sigma$ is the stand deviation of the GPS error.

Furthermore, the transition probability is defined in Eq. (14.2) as

\[
P(Seg_{t+1}^j | Seg^i_t) = \frac{gd(GPS_t, GPS_{t+1})}{rd(GPS_t, GPS_{t+1})}
\]

in which $rd$ is the distance between two geolocations along the segment path.

Given a sequence of GPS readings as the observations, we can utilize the Viterbi algorithm [32] to find out the most likely sequence of road segments as the hidden states.
3.3 Arrival time estimation and delay calculation

After obtaining the actual bus geolocations, we use them together with bus stop geolocations to estimate the arrival times at bus stops. The arrival time $A_t^s$ at the bus stop $s$ for bus trip $t$ is computed using Eq. (14.3) as

$$A_t^s = \frac{rd(GPS^i_s, GEO_s)}{rd(GPS^i_m, GPS^m_s)}(T_{GPS^i_m} - T_{GPS^i_m}) + T_{GPS^m}$$

(14.3)

in which $GEO_s$ is the geolocation of bus stop $s$, $T_{GPS^i_m}$ is the time stamp of the last GPS reading $GPS^i_m$ before bus stop $s$, and $T_{GPS^i_n}$ is the time stamp of the first GPS reading $GPS^i_n$ after the bus stop $s$.

3.4 Delay calculation and aggregation

The delay $D_t^s$ at bus stop $s$ for bus trip $t$ is calculated using Eq. (14.4) as

$$D_t^s = A_t^s - SAt_t^s$$

(14.4)

in which $SAt_t^s$ is the scheduled arrival time at bus stop $s$ for bus trip $t$. A negative $D_t^s$ means that the bus will arrive at the stop early than the scheduled time, while a positive $D_t^s$ indicates that the bus will be late.

Finally, after all delays have been calculated for all bus stops and all bus trips, we use LGA boundary data to fuse the COVID-19 case data with bus delay data. By checking whether the bus stop geolocations are within a same LGA boundary, we divide the bus stops into groups in which all bus stops are located in the same LGA. We further calculate the median, the 5th percentile and 95th percentile for each LGA group and each 30-minute time window from 5:00 a.m. to 22:00 p.m. We also group the number of confirmed cases of COVID-19 by their locations of LGA. In this way, the COVID-19 case data and bus delay data are linked to each other.

4. Case study

In order to verify our methodology, we conducted a case study to quantify the impact of COVID-19 in terms of bus delay in the Sydney metropolitan region. As previously mentioned, the time window of the case study is from February to March 2020, so all results presented here will be referring to these 2 months of study.

4.1 Area selection

We first started with data processing on the COVID-19 records to visualize the accumulated number of confirmed cases in March for each LGA. Fig. 14.11A presents the results of accumulated infected cases by each LGA; the higher the number of reported cases, the darker the LGA area becomes. There were six LGAs with an accumulated
number of cases reaching more than 80, which are grouped into four areas in terms of geolocation including:

- CBD and eastern suburbs,
- Northern Beaches area,
- the Southerland shire area, and
- the Blacktown area.

The area around city CBD and eastern suburbs ranks at the top of accumulated cases in March 2020, and it represents a higher priority for government agencies because of the importance and location of all major business hubs, together with all services and multimodal transport hubs.

Following the methodology proposed in Section 3, we further processed the bus data and quantified the variation of bus delay from February to March for each LGA. By “variation,” we refer to the reduction (or increase) in the aggregated bus delay for each LGA. The results are visualized in Fig. 14.11B. The darker the color of each LGA, the bigger the delay reduction registered for that particular LGA. Upon a detailed investigation, it can be observed that there were significant changes in three areas including

- CBD and eastern suburbs,
- Northern Beaches area, and
- the Southerland shire area.

It is notable that the area of CBD and eastern suburbs ranks as the top of the most affected area by a significant change in bus delay and also in the number of accumulated
infection cases. These interesting findings motivate us to further investigate the understanding of the impact that the COVID-19 pandemic has had on different urban areas. We strategically choose two areas with different characteristics for comparison. One area (called area A) is the area of CBD and eastern suburbs due to these interesting findings. Another area (called area B) is the area of western suburbs that has a small number of confirmed COVID-19 cases. These two selected areas are shown in Fig. 14.12.

4.2 Bus delay analytics and comparison

Figs. 14.13 and 14.14 show the bus delay statistics on areas A and B in February and March, respectively, including the median bus delay for each area, the 5th and the 95th percentiles of bus delay from 5:00 a.m. to 22:00 p.m., recorded across all bus stops in the area. Blue (black in print version) lines stand for the delays in February 2020, and orange (gray in print version) lines stand for the delays in March 2020. It can be observed that the change of the bus delay was more significant in area A than in area B, especially in the upper bound of delay in peak hours. The overall average decrease of delay was 4.4 min (36%) in area A while only 1.3 min (15%) in area B in peak hours (7:00 a.m. to 9:00 a.m. and 16:00 p.m. to 18:30 p.m.). Furthermore, the results indicate that there is a significant improvement of bus delay in the area of CBD and eastern suburbs, which can
reach even a 9.5-min delay reduction around afternoon peak hour (17:30 p.m.). This implies that travelers in central areas are now waiting almost 10 min less for their bus to arrive to the stop, which were previously delayed by severe traffic congestion most likely.

However, the area of western suburbs tends to maintain the same level of bus delay despite the new conditions and does not record a significant improvement of the bus
delay reduction after the travel restrictions and stay-at-home social distancing rules. This can be explained by several factors, mostly related to the number of business hubs in that area and the type of business that are present, which consist of many warehouses, bus depots, large storage areas, and suburb houses. This explains that the activity in this area continues as normal and is not affected by a decrease in the number of workers traveling daily or a slight decrease in the number of personal cars on the roads.

To further illustrate the difference of delay reduction in the two selected areas, the distributions of delays from 17:30 to 18:30 are also visualized in Fig. 14.15 in which the blue (dark gray in print version) color is used for February dataset and orange (light gray in print version) color is for March dataset. Once again, it can be observed that the distribution of delays in area A shifts to the left and has a smaller tail, which confirms the overall decrease of delay in this area, while the distribution of delays in area B suffers very slight changes, remaining almost the same as previous COVID-19 travel restrictions.

To reinforce the abovementioned findings, Fig. 14.16 shows the delay heat map of each bus stop delay recorded for area A in the morning peak hours of February and March. Each dot in Fig. 14.16 stands for a bus stop. Green (light gray in print version) color means smaller delay (buses arriving on time), while red (dark gray in print version) color means larger delay (buses being always late). The heat map reveals the locations where the bus delay has significantly improved, such as the CBDs, transport interchanges, shopping/entertainment centers, and beaches. Once again, this reinforces the positive impact that the travel restrictions have had in area A and a significant improvement of the bus service in this area. The current conditions can be taken as a new ideal standard of bus operations when traffic patterns and congestion levels will be reestablished to regular levels later in 2020.

Finally, the comparison results imply that the COVID-19 pandemic has had a different impact on public transport travel behavior in different areas in the city, and this
can be used for all bus operators making optimization decisions in different regions across large metropolitan areas, under all varying traffic conditions: regular, pandemic, or area-restricted travel plans that the local transport agencies might impose when needing to isolate specific strategic locations in the city. The current big data solution can be used periodically for reassessing the abovementioned findings and for making again reinforced travel planning decisions.

5. Conclusion and future work

In this chapter, we proposed a methodology to quantify the changes of bus delay in order to study the impact of COVID-19 pandemic on public transport travel behaviors across a large metropolitan area from Australia. The research involved big data real-time processing and detailed data analytics. The main datasets were collected from multiple sources and had various characteristics such as large volume, real time, and spatial and temporal features. These bring true challenges to the data processing and analysis.

To verify our methodology, a case study was carried out across the Sydney metropolitan region from February to March 2020. The findings show that over the month of March 2020, the COVID-19 pandemic has significantly impacted people’s travel behaviors especially in the central and coastal areas. The effect of travel restrictions reduced the overall traffic congestion in the central urban areas, which is usually affected by major delays on all transport modes, especially during morning/afternoon peak

FIGURE 14.16 Bus delay heat maps for area A in the morning peak hours of February and March. Red (dark gray in print version) color means larger delay.
hours. The analytics platform revealed a significant drop in bus delays in March 2020 around the central and eastern suburbs in Sydney, reaching even a lower delay record of almost 10 min lower during the afternoon peak hours. This is a significant improvement for bus operations in such a central location.

The proposed methodology enables us to quantify the travel behavior changes caused by major events such as the COVID-19 pandemic. This is helpful to understand and mitigate the impact in different areas with different conditions. The quantified delay reduction also reveals the potential of better transport performance, which could be used for the benchmark of transport performance improvement after the pandemic.

We keep collecting the data continuously and will extend this research by studying the travel behaviors after the pandemic, when travel restrictions will be lifted. It will be interesting to see how the travel behaviors restore to the normal conditions and if the regular traffic congestion will bring the bus delays to the same levels as prior to the pandemic. A future extension would be to enable the impact of travel restrictions across multiple travel modes in the city and quantify the interchange travel time reductions. Also, we are currently exploring various mobility restriction measures that can be enforced under critical situations such as restricting completely the accessibility of travelers to specific suburbs (total lockdown of various areas). This would need to be quantified in terms of public routing as well as intermodal changes.

References

[1] WHO, World Health Organisation, 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen.

[2] PM, Prime Minister of Australia, 2020. Available from: https://www.pm.gov.au/media/update-coronavirus-measures-270320. (Accessed 30 April 2020).

[3] S.D. Williams, Y. Bock, P. Fang, P. Jamason, R.M. Nikolaidis, L. Prawirodirdjo, M. Miller, D.J. Johnson, Error analysis of continuous GPS position time series, J. Geophys. Res. Solid Earth 109 (B3) (2004).

[4] NSW, Government. Data. NSW, 2020. Available from: https://data.nsw.gov.au/data/dataset/covid-19-cases-by-location/resource/21304414-1ff1-4243-a5d2-f52778048b29.

[5] Google, Google Transit APIs, 2020. Available from: https://developers.google.com/transit.

[6] Google, Public Feeds Wiki, 2020. Available from: https://code.google.com/archive/p/googletransitdatafeed/wikis/PublicFeeds.wiki.

[7] S. Farber, M.Z. Morang, M.J. Widener, Temporal variability in transit-based accessibility to supermarkets, Appl. Geogr. 53 (2014) 149–159.

[8] S. Farber, L. Fu, Dynamic public transit accessibility using travel time cubes: comparing the effects of infrastructure (dis) investments over time, Comput. Environ. Urban Syst. 62 (2017) 30–40.

[9] K. Fransen, T. Neutens, S. Farber, P. De Maeyer, G. Deruyter, F. Witlox, Identifying public transport gaps using time-dependent accessibility levels, J. Transport Geogr. 48 (2015) 176–187.

[10] A. Owen, D.M. Levinson, Modeling the commute mode share of transit using continuous accessibility to jobs, Transp. Res. Pol. Pract. 74 (2015) 110–122.
[11] A. Guthrie, Y. Fan, K.V. Das, Accessibility scenario analysis of a hypothetical future transit network: social equity implications of a general transit feed specification—based sketch planning tool, Transp. Res. Rec. 2671 (1) (2017) 1–9.

[12] Y. Hadas, Assessing public transport systems connectivity based on Google transit data, J. Transport Geogr. 33 (2013) 105–116.

[13] N. Wessel, M.J. Widener, Discovering the space—time dimensions of schedule padding and delay from GTFS and real-time transit data, J. Geogr. Syst. 19 (1) (2017) 93–107.

[14] T. Braz, M. Maciel, D.G. Mestre, N. Andrade, C.E. Pires, A.R. Queiroz, V.B. Santos, Estimating inefficiency in bus trip choices from a user perspective with schedule, positioning, and ticketing data, IEEE Trans. Intell. Transp. Syst. 19 (11) (2018) 3630–3641.

[15] N. Wessel, J. Allen, S. Farber, Constructing a routable retrospective transit timetable from a real-time vehicle location feed and GTFS, J. Transport Geogr. 62 (2017) 92–97.

[16] F. Sun, Y. Pan, J. White, A. Dubey, Real-time and predictive analytics for smart public transportation decision support system, in: 2016 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, 18 May 2016, pp. 1–8.

[17] F. Sun, A. Dubey, C. Samal, H. Baroud, C. Kulkarni, Short-term transit decision support system using multi-task deep neural networks, in: 2018 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, 18 Jun 2018, pp. 155–162.

[18] J. Wu, L. Zhou, C. Cai, F. Dong, J. Shen, G. Sun, Towards a general prediction system for the primary delay in urban railways, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, 27 Oct 2019, pp. 3482–3487.

[19] S.A. Zahabi, A. Ajzachi, Z. Patterson, Transit trip itinerary inference with GTFS and smartphone data, Transp. Res. Rec. 2652 (1) (2017) 59–69.

[20] N. Nassir, A. Khani, S.G. Lee, H. Noh, M. Hickman, Transit stop-level origin—destination estimation through use of transit schedule and automated data collection system, Transp. Res. Rec. 2263 (1) (2011) 140–150.

[21] TfNSW, Open Data Hub, 2020. Available from: https://opendata.transport.nsw.gov.au.

[22] ABS, abs.gov.au, 2020. Available from: https://www.abs.gov.au/websitedbs/d3310114.nsf/home/digital+boundaries.

[23] M. Hashemi, H.A. Karimi, A critical review of real-time map-matching algorithms: current issues and future directions, Comput. Environ. Urban Syst. 48 (2014) 153–165.

[24] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, Y. Huang, Map-matching for low-sampling-rate GPS trajectories, in: Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems, vol. 4, 2009, pp. 352–361.

[25] C.E. White, D. Bernstein, A.L. Kornhauser, Some map matching algorithms for personal navigation assistants, Transp. Res. C Emerg. Technol. 8 (1–6) (2000) 91–108.

[26] S. Brakatsoulas, D. Pfoser, R. Salas, C. Wenk, On map-matching vehicle tracking data, in: Proceedings of the 31st International Conference on Very Large Data Bases, 2005, pp. 853–864.

[27] D. Srinivasan, R.L. Cheu, C.W. Tan, Development of an improved ERP system using GPS and AI techniques, in: Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems vol. 1, IEEE, 2003, pp. 554–559.

[28] L. Li, M. Quddus, L. Zhao, High accuracy tightly-coupled integrity monitoring algorithm for map-matching, Transp. Res. C Emerg. Technol. 36 (2013) 13–26.

[29] M. Ren, H.A. Karimi, Movement pattern recognition assisted map matching for pedestrian/wheelchair navigation, J. Navig. 65 (4) (2012) 617–633.
[30] A. Varga, R.K. Moore, Hidden Markov model decomposition of speech and noise, in: International Conference on Acoustics, Speech, and Signal Processing, IEEE, 1990, pp. 845–848.

[31] S. Fine, Y. Singer, N. Tishby, The hierarchical hidden Markov model: analysis and applications, Mach. Learn. 32 (1) (1998) 41–62.

[32] G.D. Forney, The viterbi algorithm, Proc. IEEE 61 (3) (1973) 268–278.