Generating Historic General Transit Feed Specification for Sydney: 1855 to 2015

Hema Rayaprolu, Bahman Lahoorpoor, and David Levinson

January 21, 2026

1 Abstract

We digitised Sydney's historic public transport networks and constructed General Transit Feed Specification (GTFS) datasets covering the period from 1855 to 2015. The dataset enables systematic, reproducible analysis of the long-term evolution of public transport provision and accessibility in Sydney using contemporary transit analytics. To validate the approach, we compared accessibility metrics derived from the generated GTFS for 2015 with those from the officially published GTFS for the same year. The generated data reproduced access levels within 2% of the official dataset, demonstrating the reliability of the method.

2 Methodology

For the Greater Sydney region, Transport for New South Wales (TfNSW), the state transport authority, collates information from the region's various public transport service operators and publishes GTFS feeds regularly at Transport for NSW (2021). The majority of feeds released since March 2013 have been archived at Open Mobility Data (2021), however, the 2013 and 2014 archives are incomplete, making 2015 the earliest year for which a reliable published GTFS feed is available.

To enable analysis of public transport provision prior to 2015, we generated historic GTFS feeds based on archived route and timetable information. We constructed feeds for bus, train, and tram services in the Greater Sydney region covering the period from 1855 to 2015. These were combined with ferry services derived from the published 2019 GTFS feed. Ferry services were assumed to be constant over the study period, as ferry routes have changed little over time (Sandell, 2021) and their timetables haven't been archived as rigorously.

The resulting historic GTFS feeds were validated by comparing public transport accessibility metrics derived from the generated 2015 feed with those obtained from the officially published 2015 GTFS, and by calibrating key assumptions accordingly.

A GTFS feed comprises a set of comma-separated text files that follow a defined relational structure. A detailed description of the GTFS files and their specifications is provided by Google Developers (2021).

2.1 Historic tram and train GTFS: 1855 to 2015

For the generation of Sydney's historical tram and train GTFS, the focus was limited to the existence of services and network coverage over time, regardless of technological changes in vehicle types or infrastructure.

All stations, stops, and track alignments were georeferenced, with associated opening and closure dates encoded to enable temporal filtering. Schedules were derived from historical sources where available; otherwise, synthetic timetables were generated using assumed average speeds - 20 km / h for trams and 30 km / h for trains - based on known distances. Bidirectional services were assumed throughout, and irregular or special services were excluded due to their minimal impact on overall service frequency.

The tram network data was mainly sourced from historical works by Keenan (Keenan, 1979), while the train data was collected from online repositories such as Wikipedia, the official Sydney Trains website, and reports from the Australasian Railway Association. Where headway or frequency information was missing, consistent service intervals were estimated based on average travel times.

More details on the historic GTFS generation process for trams and trains can be found in Lahoorpoor (2022).

2.2 Historic bus GTFS: 1925 to 2015

To generate historic GTFS for buses, we used archived information on bus route and timetable changes. Historic changes in Sydney metropolitan region's bus routes from 1925 onward have been archived by Henderson (2021) on the website sydneybusroutes.com. The archives provide systematically arranged records of changes in streets traversed, timetables, and operating arrangements for every bus route operated in the region.

Using the archived bus route and timetable information, we generated bus service GTFS for every year between 1925 and 2015. The process was achieved through the following steps:

1. Converting archives recorded in text-format into a machine-readable format to automate data extraction. This was done by manually rearranging and formatting each route variation in the text documents which were then fed into a Python script we developed to extract relevant information.

2. Generating route shapes corresponding to each route variation from the edited archives data. We did this in two steps using the Python package, OSMnx (Boeing, 2017). First, we matched consecutive streets in a route shape's street list with OSMnx street intersections to obtain a list of geo-referenced street intersections corresponding to each route shape. Then, we obtained the shortest path between each pair of consecutive intersections on a route and merged the individual paths to get route shapes.

3. Digitising timetables and estimating service levels. This was done by first manually tabulating route timetables archived alongside route variations. The following information was available for each route variation: off-peak travel time, start time and direction of first and last trips, and headway between trips. Based on these attributes, we calculated the number of trips operated in a day and estimated start times of each trip.

4. Determining stop locations and generating stop-time itineraries. The archives did not contain information on bus stops. Consequently, we assumed all stops serviced by buses in 2019 to have been serviced historically, and established stops where there were none in 2019. We did this by generating points at 400-metre spacing along streets using QGIS (QGIS Development Team, 2021) and the 2019 OSM network (OpenStreetMap contributors, 2021). Stop spacing was set at 400 metres based on the average bus stop spacing in the 2019 GTFS (407 m) and the region's guideline for stop spacing (NSW Government, 2013). Once stops were determined for

each route shape, arrival times at each stop were determined by considering the route speed and the distance travelled to reach the stop.

5. Compiling GTFS feeds by year. The historic GTFS feeds were formed by combining the generated routes shape, trip, and stop time information for each route variation in each year. Agency for all bus routes was assigned as 'Sydney Buses Network' as in the 2019 GTFS feed. The feeds contain service information for typical weekday services as well for Saturday and Sunday services, which was duly reflected in calendar service entries.

A detailed account of the step-by-step process including assumptions and limitations is available in Rayaprolu (2023).

2.3 Validating 2015 GTFS using access

To ensure reliability and accuracy of the generated historic GTFS, we validated the feeds generated by comparing 2015 feeds with those published.

As the information sources and development processes for the two feeds are different, we found the datasets were not readily comparable. Although routes could be compared by matching route numbers in most cases, the published routes have several short-workings and other variations that are not entirely captured in the archives and the generated GTFS. As a result, we chose to use person-weighted average access (PWA) across the region, provided by the two networks to calibrate the assumptions made while generating the GTFS.

Access was measured as the cumulative number of opportunities reached by using public transport within a certain time threshold. Jobs are often used to measure access as commuting to and from work tends to be the most significant of daily trips. However, as spatially disaggregated employment data were unavailable historically, we measured access to population instead. We used the historic population distribution established by Lahoorpoor and Levinson (2021) for the Greater Sydney region.

With population counts by mesh block and GTFS for each year, we estimated the population that could be reached within a certain time threshold by public transport for each year. To do this, we queried travel time isochrones for each mesh block using Open Trip Planner (OTP) (OpenTripPlanner, 2021). OTP builds a network graph based on the supplied street network and GTFS, and computes point-to-point travel times and isochrones for a given point, time cutoff, mode, day and time. We supplied OTP the GTFS generated, and queried isochrones for each mesh block centroid for 8:00 am on a weekday. The street network remained unchanged (downloaded from OpenStreetMap contributors (2021) in July 2020) as historical street networks are not consistently available Turner et al. (2023). The isochrones obtained were overlapped with population in the corresponding year to estimate access. Population within a mesh block was assigned to its centroid for computational ease.

The 30-minute PWA difference at 08:00 am between the generated 2015 GTFS and the published 2015 GTFS was 1.9%.

We also computed access at every minute between 08:00 am and 08:30 am and measured the difference in PWA over the 30-minute period to normalise time-specific access variations. Overall, the average PWA from generated 2015 GTFS between 08:00 am and 08:30 am is 1.4% less than that from the published GTFS. When compared spatially at the mesh block level, most of the core and inner areas of the region have mostly small differences.

Together, these results indicate that the generated 2015 GTFS closely reproduces both temporal and spatial accessibility patterns observed in the published data, supporting its use for historical accessibility analysis.

References

Boeing G (2017) OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* 65: 126–139.

Google Developers (2021) GTFS Static Reference. URL https://developers.google.com/transit/gtfs/reference. Accessed: 2021-12-22.

Henderson R (2021) Sydney bus routes: A history of private and government bus routes in Sydney. URL https://sydneybusroutes.com. Accessed: 2021-07-14.

Keenan DR (1979) *Tramways of Sydney*. Transit Press Sydney. ISBN 0909338027.

Lahoorpoor B (2022) *Terraces, towers, trams, and trains : Examining the growth of Sydney using empirical models and agent-based simulation*. PhD Thesis, The University of Sydney.

Lahoorpoor B and Levinson D (2021) An empirical model of land use and public transit co-developments in Sydney. Working Paper 06, The University of Sydney.

NSW Government (2013) Sydney's bus future. Simpler, faster, better bus service. URL https://mysydneycbd.nsw.gov.au/sites/default/files/user-files/uploads/bus-future-final-web.pdf. Accessed: 2021-07-22.

Open Mobility Data (2021) Greater Sydney GTFS. URL https://transitfeeds.com/p/transport-for-nsw/237. Accessed: 2021-07-14.

OpenStreetMap contributors (2021) Planet dump retrieved from https://planet.osm.org. URL https://www.openstreetmap.org.

OpenTripPlanner (2021) Multimodal Trip Planning. https://www.opentripplanner.org. Accessed: 2021-07-22.

QGIS Development Team (2021) *QGIS Geographic Information System*. QGIS Association. URL https://www.qgis.org.

Rayaprolu H (2023) *The Co-Evolution Of Public Transport Access And Ridership*. PhD Thesis, The University of Sydney. URL https://ses.library.usyd.edu.au/handle/2123/31254.

Sandell R (2021) UrbanFerryist. URL https://twitter.com/UrbanFerryist/status/1476468691625594880. Tweet, 30 December 2021: A pictorial reminder that Sydney's ferry network of 1909 was very similar to what it is today - main change is the cessation of car ferries following the opening of the Harbour Bridge.

Transport for NSW (2021) TfNSW Open Data Hub and Developer Portal. URL https://opendata.transport.nsw.gov.au. Accessed: 2021-07-14.

Turner H, Lahoorpoor B and Levinson DM (2023) Creating a dataset of historic roads in sydney from scanned maps. *Scientific Data* 10(1): 683.