1 **Leveraging GTFS data to calculate an open-source Transit Supply Index**
2
3
4
5 **James Reynolds**
6 *Corresponding Author*
7 Research Fellow
8 Public Transport Research Group, Institute of Transport Studies, Department of Civil
9 Engineering, Monash University, Victoria, Australia
10 `james.reynolds@monash.edu`
11
12 **Graham Currie**
13 Professor
14 Public Transport Research Group, Institute of Transport Studies, Department of Civil
15 Engineering, Monash University, Victoria, Australia
16 `graham.currie@monash.edu`
17
18
19
20
21 Word Count: 684 words + 0 table(s) × 250 = 684 words
22
23 Submission Date: June 30, 2023

1 **ABSTRACT**
2 TBC.
3
4 *Keywords*: Transit, Public transport, GTFS, Benchmarking,

INTRODUCTION

Transit service level indicators include those in the Transit Capacity and Quality of Service Manual (TCQSM) (1), the Transit Score metric and many more. Practitioners, researchers and advocates seeking to use such metrics may face two inter-related challenges: firstly, there is the problem of calculating the metrics themselves for a specific location; secondly, is the challenge of explaining the metrics, their meaning and importance those who are not specialists in transit, such as politicians, other decision-makers or the general public.

The TCQSM specifies Levels of Service (LOS) between A and F across a range of factors including service span, frequency, speed, and the proportion of population serviced. Previous research by Wong (2) overcame some challenges of using the TCQSM, by using Python, PostgreSQL and R software and GTFS feeds as input to automate the calculation of daily average headways, route length and stop numbers. This indicators, however, are route based and so do not include any consideration of geographic or population coverage. Further metrics addressing these topics and much detail about their calculation and meaning are included in the TCQSM, such as the Service Coverage Area (pp. 5-8 to 5-21). However, these appear highly detailed, may required bespoke GIS or other analysis, and it might be challenging to explain these measures (beyond the fact at A is good and F is bad) to non-technical decision-makers, stakeholders or others who might be involved. Transit Score provides a similarly easily understood rating scale, scoring locations out of 100 (3). However, the algorithm is patented and effectively a black box, meaning that it is not possible to calculate scores independently to understand how the metric might change with cahnges to the transit system or surrounding environment.

The Supply Index developed by Currie and Senbergs (4) may provide a metric that is relatively easy to calculate, open (rather than a black box), and relatively simple for a non-technical audience to understand, engage with and use. This Index is based on calculating the number of transit arrivals at stops within an area of interest, with an adjustment made for the amount of the area of interest that is within a typical walk access distance of each stop. However, it does not appear to have been widely used, perhaps in part because it still required an analyst to obtain sources of timetable and geographic data. Since the publication of Currie and Senbergs (4) such data has become much easier to obtain with more than 10,000 agencies now providing timetable and network data using the General Transit Feed Specification (GTFS) format (5). A gap, however, is that there is not yet a method for calculating the Currie and Senbergs (4) Supply Index directly from GTFS data.

This paper reports the development of R code to calculate the Supply Index of Currie and Senbergs (4) directly from GTFS data. The code is developed using data from a single case: the GTFS for Victoria in Australia, which includes Greater Melbourne. Cross-case comparison to Toronto, Canada, and Washington DC, USA, is also undertaken to test the results and gain understanding of how the Supply Index might be useful for practitioners, researchers and advocates. The motivation for this research is to better understand how GTFS data might be used to produce benchmarking metrics that can be calculated using open-source code, that can be used to access proposed network changes and which may be relatively easy for non-technical specialists to understand.

RESEARCH CONTEXT

Even a brief search shows that there is a very large number of metrics available for benchmarking transit services, for example: the Transit Cooperative Research Program (TCRP) Report 88

1 provides an extensive guidebook on developing a performance-measurement system (6); online
2 databases are provided by the Florida Transit Information System (FTIS) (7) and the International
3 Association of Public Transport (UITP) (8) have online databases, while the Transport Strategy
4 Centre of Imperial College London runs extensive annual benchmarking programmes across over
5 100 transit provides around the world (9).
6   The TCQSM and Transit Score may provide contrasting examples: with respect to the first
7 challenge, TCQSM metrics may require large amounts of network, service, population and other
8 data to be assembled before the indicators can be calculated; whereas Transit Scores are readily
9 available on the Walk Score (3) website for locations with a published GTFS feed (eliminating the
10 need for any calculations). With respect to the second challenge, the meaning of the Transit Score
11 appears easy to explain (the closer to 100, the better), but as the score is calculated by a patented
12 algorithm (effectively a black-box) it may not be easy to understand or explain the connection
13 between real-world conditions and the score, or what might need to be done to improve the score
14 and service levels. Nor does it appear to be possible for Transit Scores to be generated for proposed
15 changes to networks. The TCQSM, in

16 **General Transit Feed Specification (GTFS)**
17 **Social needs and the Suppy Index**
18 [1]

19   he SI_{CCD} can then be compared between different CCDs to give an indication of the relative
20 supply of transit, adjusted for accessibility.
21   An advantage of the Supply Index is that it is a relatively simple number to calculate,
22 understand and explain. It is based on the number of bus/tram/train arrivals per week at stops
23 within the CCD, which is multiplied by a factor allowing for the amount of the CCD that is within
24 walking distance of each stop [2].
25   Currie and Senbergs (4) calculated the SI for various CCDs in Melbourne using a timetable
26 database provided by the Victorian Public Transport Authority (PTA). This predated the widespread
27 availability of GTFS data, which provides a standardised format for timetable data that is produced
1 by many transit systems. A question, therefore, is how to calculate the SI using GTFS data so that
2 SI_{CCD}s can be calculated for current services in Melbourne or other places.

$$SI_{CCD} = \sum_N \left(\frac{Area_{Bn}}{Area_{CCD}} * SL_{Bn} \right) \quad \textbf{Formula 1}$$

where:

SICCD	= Supply Index for the CCD
CCD	= CCD under analysis
N	= number of walk access buffers to stops/stations in each CCD
Bn	= Buffer n for each stop/station in each CCD
Area	= square kilometre spatial area of the CCD
SL	= Service Level Measure (number of bus/tram/train vehicle arrivals per week)

[1]

[2]This is the $Area_{Bn}$ divided by $Area_{CCD}$ part of the calculation. Thresholds for access of 400m for bus and tram stops, and 800m for railway stations are used in the calculation of the $Area_{Bn}$ variable.

3 **METHODOLOGY**
4 **RESULTS**
5 **R code**
6 **Case studies**
7 *Melbourne*
8 Overall
9   Wheelchair accessibility?
10   The metro tunnel - adding services

11 *Toronto*
12 *Washington DC*
13 *Cross-case comparison*
14 **DISCUSSION**
15 **CONCLUSIONS**
16 **AUTHOR CONTRIBUTION STATEMENT**
17 The authors confirm contribution to the paper as follows: study conception and design: A. Anony-
18 mous, D. Zoolander; data collection: B. Security; analysis and interpretation of results: A. Anony-
19 mous, B. Security; draft manuscript preparation: A. Anonymous. All authors reviewed the results
20 and approved the final version of the manuscript.

References

1. Kittleson & Associates, Parsons Brinckerhoff, KFH Group, Texas A&M Transportation Institute, and ARUP. *Transit Capacity and Quality of Service Manual, Third Edition*. Transportation Research Board, Washington DC, third edition, tcrp report 165 edition, 2013. URL http://www.trb.org/Main/Blurbs/169437.aspx.

2. James Wong. Leveraging the general transit feed specification for efficient transit analysis. *Transportation Research Record*, 1(2338):11–19, 2013. doi: 10.3141/2338-02.

3. Walk Score. Transit score methodology. 2023. URL https://www.walkscore.com/transit-score-methodology.shtml.

4. Graham Currie and Zed Senbergs. Identifying spatial gaps in public transport provision for socially disadvantages australians: the melbourne needs-gap study. In *Australasian Transport Research Forum*. Australasian Transport Research Forum, 2007.

5. MobilityData. *General Transit Feed Specification (GTFS)*, undated. URL https://gtfs.org/.

6. Paul Ryus, M Connor, S Corbett, A Rodenstein, L Wargelin, L Ferreira, Y Nakanishi, and K Blume. Tcrp report 88: a guidebook for developing a transit performance-measurement system. Technical report, 2003.

7. Florida Transit Information System. Urban integrated national transit database, 2018. URL http://www.ftis.org/urban_intd.aspx.

8. International Association of Public Transport (UITP). Mobility in cities database 2015, 2015. URL uitp.org/publications/mobility-in-cities-database/.

9. Imperial College London. Transport strategy centre (tsc); applied research, undated. URL https://www.imperial.ac.uk/transport-engineering/transport-strategy-centre/applied-research/.