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 + ?? table(s) $\times$ 250 = 434 words
22
23 Submission Date: June 30, 2023

24 **ABSTRACT**
25 TBC.

26

27 *Keywords*: Transit, Public transport, GTFS, Benchmarking,

INTRODUCTION

Transit service level indicators include those in the Transit Capacity and Quality of Service Manual (TCQSM) (**?**), 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 **?**) 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 (**?**). 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 **?**) 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 **?**) 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 (**?**). A gap, however, is that there is not yet a method for calculating the **?**) Supply Index directly from GTFS data.

This paper reports the development of R code to calculate the Supply Index of **?**) 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 provides an extensive guidebook on developing a performance-measurement system (**?**); online databases are provided by the Florida Transit Information System (FTIS) (**?**) and the International

72 Association of Public Transport (UITP) (**?**) have online databases, while the Transport Strategy
73 Centre of Imperial College London runs extensive annual benchmarking programmes across over
74 100 transit provides around the world (**?**). The Fielding Triangle (**?**) provides a framework for
75 understanding how such metrics combine service inputs, service outputs and service consumption
76 to describe cost efficiency, cost effectiveness or service effectiveness measures. At a larger scale,
77 **?**) and **?**) discuss some of the traffic, mobility, accessibility, social equity, strategic planning and
78 other rational decision-making frames that might underlie such transit metrics, while **?**) extends
79 this into models of how institutionalism, incrementalism and other public policy models might
80 apply to decision-making processes. Further examples are provided by **?**), who develop a mea-
81 sure of accessibility in the context of policy development and social equity for Latin American
82 Bus Rapid Transit (BRT) based networks, and the street space allocation metrics based around 10
83 ethical principles from **?**).

84   However, many of these metrics are difficult to calculate, complex to explain or understand,
85 and likely not well suited to communication with those who are not transit planners or engineers, or
86 otherwise technical specialists. However, where pre-calculated metrics are immediately available it
87 may not be possible for generate metrics for proposed system changes or know exactly how scores
88 are calculated. The TCQSM and Transit Score may provide contrasting examples: with respect
89 to the first challenge, TCQSM metrics may require large amounts of network, service, population
90 and other data to be assembled before the indicators can be calculated; whereas Transit Scores
91 are readily available on the **?**) website for locations with a published GTFS feed (eliminating
92 the need for any calculations). With respect to the second challenge, the meaning of the Transit
93 Score appears easy to explain (the closer to 100, the better), but as the score is calculated by
94 a patented algorithm (effectively a black-box) it may not be easy to understand or explain the
95 connection between real-world conditions and the score, or what might need to be done to improve
96 the score and service levels. Nor does it appear to be possible for Transit Scores to be generated
97 for proposed changes to networks. The TCQSM, in contrast is open-source, in that **?**) provide a
98 manual describing all the metrics and how to calculate them. However, the calculations themselves
99 appear to be complex, which may be a barrier to use by practitioners, researchers, advocates or
100 others who are not transit scheduling specialists. While **?**)
101   `https://github.com/transitland/transitland-processing-animation`

102 **General Transit Feed Specification (GTFS)**
103 **Social needs and the Suppy Index**
104 [1]

$$SI_{CCD} = \sum_N \left(\frac{Area_{Bn}}{Area_{CCD}} * SL_{Bn} \right)$$ **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]

105   he SI_{CCD} can then be compared between different CCDs to give an indication of the relative
106 supply of transit, adjusted for accessibility.

107   An advantage of the Supply Index is that it is a relatively simple number to calculate,
108 understand and explain. It is based on the number of bus/tram/train arrivals per week at stops
109 within the CCD, which is multiplied by a factor allowing for the amount of the CCD that is within
110 walking distance of each stop [2].

111   **?**) calculated the SI for various CCDs in Melbourne using a timetable database provided
112 by the Victorian Public Transport Authority (PTA). This predated the widespread availability of
113 GTFS data, which provides a standardised format for timetable data that is produced by many
114 transit systems. A question, therefore, is how to calculate the SI using GTFS data so that SI_{CCDs}
115 can be calculated for current services in Melbourne or other places.

116 **METHODOLOGY**
117 **RESULTS**
118 **R code**
119 **Case studies**
120 *Melbourne*
121 Overall
122   Wheelchair accessibility?
123   The metro tunnel - adding services

124 *Toronto*
125 *Washington DC*
126 *Cross-case comparison*
127 **DISCUSSION**
128 **CONCLUSIONS**
129 **AUTHOR CONTRIBUTION STATEMENT**
130 The authors confirm contribution to the paper as follows: study conception and design: A. Anony-
131 mous, D. Zoolander; data collection: B. Security; analysis and interpretation of results: A. Anony-
132 mous, B. Security; draft manuscript preparation: A. Anonymous. All authors reviewed the results
133 and approved the final version of the manuscript.

134 **ACKNOWLEDGEMENTS**

---

[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.