MODELLING SPATIO-TEMPORAL VARIABILITY IN INFORMALLY RUN TRANSPORT ROUTES TO IMPROVE JOURNEY PLANNING CALCULATIONS

C KING[1]*, B RYSECK[2], L RALL[1] and S MEHENNI[1]*****

[1] WhereIsMyTransport, 22 Jarvis Street, Cape Town, 8001, South Africa
*Tel: +27 82 861-3947; Email: chris@whereismytransport.com
**Tel: +27 63 425-6801; Email: linley@whereismytransport.com
***Tel: +27 63 425-6801; Email: sam@whereismytransport.com
[2]University of Cape Town, Rondebosch, Cape Town, 7700
Tel: +27 60 727-2145; Email: rysbia001@uct.ac.za

ABSTRACT

Initiatives to map and digitalise informally run transport systems such as minibus taxis are becoming increasingly common in cities across the world (e.g. Digital Matatus, Transport for Cairo, Bus Map Beirut). Many of these projects collect the basic information on the informal routes needed to form a General Transit Feed Specification (GTFS) file, a digital method created to represent scheduled, fixed route networks largely used for the purposes of journey planning. However, a single collected trip on a route would fail to capture the high variability in even a single informally run transport route. Routes and vehicle frequencies fluctuate due to factors such as time of day, vehicle fullness, specialised requests, and route disruptions. Single route collections result in GTFS files that inaccurately translate real world dynamics into a single static piece of information. Furthermore, these GTFS files are fed into journey planning platforms which generate calculations to determine different multimodal route options from a given origin to destination. Without a better understanding of the factors that affect route geometry, the ability to better model real world informally run transport operations is limited, thereby affecting the accuracy of journey planning. The objective of this study is to determine the spatio-temporal variability of a single minibus taxi route in Cape Town, South Africa to inform how real world dynamics should translate into digital journey planning calculations to deliver more relevant information to users. This study finds through on-board data collection and roadside frequency counts that deviations in route geometry, fluctuations in trip duration, variance in route frequency, and average fullness factor along a route all bear implications journey planning tools need to incorporate to provide more accurate information to minibus taxi users.

Keywords: Modelling, Journey Planner, Minibus Taxis

1. INTRODUCTION

Initiatives to map and digitalise informally run transport systems such as minibus taxis (MBT) are becoming increasingly common in cities across the world (e.g. Digital Matatus, Transport for Cairo, Bus Map Beirut). Previously, without publicly available route maps, these systems were navigated based on prior experience or word-of-mouth guidance. Recording digital information on systems has the potential to fundamentally shift public transport users' capability to use the network to better access opportunities across the city.

Translating these networks into paper or digital maps, or even into multimodal journey planning apps, to make the system legible is the purpose of many data collection efforts.

For many of these projects, the output of data collection is a GTFS file, a digital method created originally to represent scheduled, fixed route networks largely used for the purposes of journey planning. GTFS was designed for public transport systems that are fixed route and is unable to accurately represent public transport routes which are highly variable (Eros et al., 2014). Routes and vehicle frequencies fluctuate due to factors such as time of day, vehicle fullness, specialised requests, and route disruptions. Prior research has not yet investigated how many times a variable route needs to be collected to accurately translate real world dynamics into a single static piece of information.

The objective of this study was to determine the spatio-temporal variability of a single MBT route in Sea Point, Cape Town, South Africa to inform how real-world dynamics should translate into digital journey planning calculations to deliver more relevant information to users.

2. LITERATURE REVIEW: ADAPTING JOURNEY PLANNING TO SEMI-FLEXIBLE TRANSPORT SYSTEMS

Recent advancements in data collection tools and processes open the possibility to integrate the previously unrecorded informally run transport systems into multimodal journey planning tools. Much research has been conducted on the tools needed to capture shapes of informally run network routes (Klopp, 2017; Ndibatya, 2016; Oloo, 2018; Zegras, 2014) and how to define these routes in GTFS for compatibility with journey planning apps that use this feed specification (Williams, 2015). These projects collected the information needed to form a GTFS file: Stops, routes, frequencies, and fares on the informally run routes.

Beyond creating the tools needed to collect basic components to construct a GTFS file, only limited research has been done into translating real world dynamics of the informally run transport system into journey planning tools for commuters. Ndibatya et al. (2014) used an adaptive model to predict MBT stopping behaviour in Stellenbosch, South Africa to provide passengers with the most probable point along a route at which to catch a MBT given the time of day and the day of the week. This work entailed placing trackers on ten vehicles operating between two large centres of activity and recording date and times, location, speed and travel direction to determine stop locations. These stop predictions could have the potential to improve journey planning accuracy, if aggregated by routes and translated into segments to reflect the more flexible hailing behaviour of vehicles.

To make collected data from data collection projects across whole city networks compatible for use with journey planning apps which rely on GTFS, Williams et al. (2015) created a methodology for translating the collected stop and route data into the GTFS format. Continuous stops were suggested as a method to adapt GTFS to more accurately provide journey planning information that reflects the reality of the informally run system.

For scheduled, fixed-route transport, route geometry is not required as stop locations are the usual method for plotting the route in the GTFS format. As informally run transport such as MBT does not have scheduled stops but allows users to request drop-off and pick-up points along the route, Williams et al. (2015) outlined a method for continuous stops to provide a solution for simulating hailing along a route while still conforming to traditional GTFS use. Continuous stops was proposed as an additional field to the stop times and routes table within the GTFS that would allow for hailing or disembarking from a vehicle along any point along its route.

This method has the advantage of allowing a passenger to hail a vehicle at any point and reduces the work required to define many stops at small intervals along a route. It is important to note, that route geometry is a prerequisite to defining continuous stops in the case of data collected during informally run transport collection project. However, hailing is not a feature of GTFS-based journey planners and route geometry is optional in constructing a GTFS file. To build hailing into a journey planner using GTFS, route geometries are required.

Without a better understanding of route variability across the many dimensions of GTFS including frequencies, stopping points, and travel times, the ability to better model real world informally run transport operations is limited, thereby affecting the accuracy of journey planning.

3. RESEARCH PROBLEM AND OBJECTIVES

Though great strides have been made to record and codify informally run transport networks into formats that can be used in traditional, existing journey planning tools, there has been less research done into how to accurately depict the spatio-temporal reality of the network in a multimodal journey planning platform. Added considerations around informally run transport for journey planning include the effect of temporal dimensions of demand flows on frequency of vehicles, possibility to hail a vehicle along a given route point, and overall trip duration. Route geometry can also vary during the day based on demand as well as traffic, affecting frequencies. Frequency of vehicles along a route varies based on the number of vehicles serving the route during different times of the day, and deviations drivers may take that shorten the route.

Rather than investigating the cause of variations, this study limits its scope to understanding how the variations themselves could affect journey planning calculations. Factoring such spatio-temporal considerations into journey planning calculations could present more reliable information and thereby inform considerations for improving adoption and retention of users on journey planning apps for informally run transport systems.

To better inform how real world dynamics should translate into digital journey planning calculations to deliver more relevant information to users, this study intends to determine the spatio-temporal variability of a single MBT route in Cape Town, South Africa.

This is broken down into four research objectives. These are to determine variability in (1) route geometry, (2) trip duration, (3) fullness factor, and (4) route frequency over the course of the weekday during AM-peak, PM-peak as well as off-peak hours.

Study hypotheses are formed based on the way that current journey planning calculations are being made. In terms of route geometry, we hypothesise that rank locations remain the same throughout the day regardless of AM, PM, and off-peak travel periods. Further, we hypothesise that it is possible to hail a vehicle on an alternative route that deviates from the trunk route. We hypothesise that vehicle behaviour at the beginning and end of a route affects the expected trip duration. In regards to fullness factor of a MBT, we hypothesise that it is possible to hail a vehicle at any point along the route. Lastly, we hypothesise that vehicle frequency along a route varies based on time of day and route directionality.

4. STUDY SITE

The study focuses on a licensed route which runs about 5.6km in length from Cape Town Station in the Central Business District to Sea Point, a suburb along the Atlantic Seaboard (see Figure 1). The route runs through several predominantly higher-density, upper-middle-income suburbs including De Waterkant, Green Point and Three Anchor Bay, ending in Sea Point – a suburb of roughly 13,000 people with a density of 8,418.16 people per km^2 (Statistics South Africa 2011). The demographic makeup is a majority white population with black Africans making up one-fifth of the population (Ibid.). The licensed route follows a main two-lane road broken with traffic lights through mixed-use areas and with a high concentration of commercial and entertainment buildings along the main road in Sea Point. There are several larger shopping centres along this route. Scheduled MyCiTi buses run three bus lines along this route. At about a 350m distance, High Level Road, a road with far fewer traffic lights, runs parallel to this main road through residential areas. Several routes beginning in low-income areas in the peripheries of Cape Town such as Nyanga and Philippi Lower contribute to the number of vehicles on the Sea Point route, joining the route primarily via Granger Bay Boulevard Circle at York Rd, about 2.6km from the Cape Town Station.

Figure 1: Licensed Cape Town to Sea Point MBT route

5. METHODOLOGY

The study was conducted in two parts. In the first part of the study, route geometry, trip duration and fullness factor were collected a minimum of three times in both outbound and inbound directions on the route connecting the taxi ranks in Sea Point Main Road and Cape Town Station. The second part of the study, route frequency, entailed counting vehicles per interval minute and the direction it came from and is going to along three points along the route. Due to resource constraints this was limited to two points along the main route and one point along the alternative route to capture variation in frequency at key points along the route where vehicles have an option to exit the main route. Observing major points along the route where taxis take shortcuts or cut across streets (such as

Glengariff Road in Three Anchor Bay) in the first part of the study informed the positioning of data collectors along probable route junctures.

This study made use of WhereIsMyTransport's Collector App – a fit-for-purpose app designed for the collection of GTFS data on informally run transport routes. Specifically, three components of the app were used – the GPX traces, time stamps, and boarding and alighting fields that indicate seating capacity.

Route characteristics such as land-use, alternative road options, demographic composition of areas routes run through, were likely to strongly influence the spatio-temporal variability of a given route. Therefore, results from this study were limited to improving accuracy along the specific study route. Due to resource and time constraints this study was limited to weekdays excluding Fridays, and this study assumed that data collected during study days reflect the variability seen on non-collection days.

6. FINDINGS

The findings listed below are organised by hypotheses and their related data results. Key terminology includes rank, terminating point, route, and departure point. A *rank* is defined as a point where vehicles depart from at the end of a route, sometimes defined by physical infrastructures such as a station, but also can be without any infrastructure and have been formed over time through a shared understanding of the full route's endpoint between drivers and passengers. A *terminating point* differs in that it is a point along a route where a MBT may choose to turn around. A *route* is the path along a road network that a vehicle will follow with relative accuracy and has a starting rank and an ending rank or terminating point. In the context of journey planning, a *departure point* is the point where a journey planner will direct a passenger to start their trip.

6.1 Route Geometry

We hypothesised that rank locations remained the same throughout the day regardless of AM, PM, and off-peak travel periods. This hypothesis was shown to be false. Rank location was observed through cleaned route paths collected with the app to change during the day affecting route start and end points (see Figure 2). Data shows that multiple starting ranks can exist. Further, data shows that a rank can have no official location as it is merely a turnaround point for the vehicles, thereby allowing vehicles to terminate their route early or to extend their route as a courtesy to passengers wanting to travel slightly further than the designated terminating point. In one observed case during the PM-peak, a taxi travelling in the direction of Cape Town dropped off all passengers in front of the Cape Town Station, despite it not being the official rank or terminating point.

We hypothesised that it is possible to hail a vehicle on an alternative route that deviates from the trunk route. While we found that vehicles did deviate from the trunk route, it is still unclear whether there is a high enough probability of hailing on these to include such deviations in the journey planner. Route geometry varied most strongly during AM and PM-peak travel periods, with taxis deviating from the licensed route along the Main Road to take alternative paths between ranks. Deviations were observed when a taxi filled up at the departure rank and, to minimise trip duration in order to quickly return to fill up with more passengers, took the path of least resistance to the ending rank or terminating point. Since the Main Road has a higher concentration of traffic lights and vehicles than High Level Road, full taxis opted to take High Level Road as an express route. In other cases, particularly in the AM-peak travel period, taxis quickly returned back empty on High Level

Road from Sea Point to Cape Town to refill passengers in the city centre rather than search for passengers travelling in the direction of Cape Town along Main Road. While we found that it is possible to hail a MBT along the alternative route on High Level Road, it was much less likely due to a low frequency of vehicles in the AM-peak, no vehicles in the off-peak and full vehicles during the PM-peak travel periods. This finding affects the route frequency which will be further discussed in the final findings section.

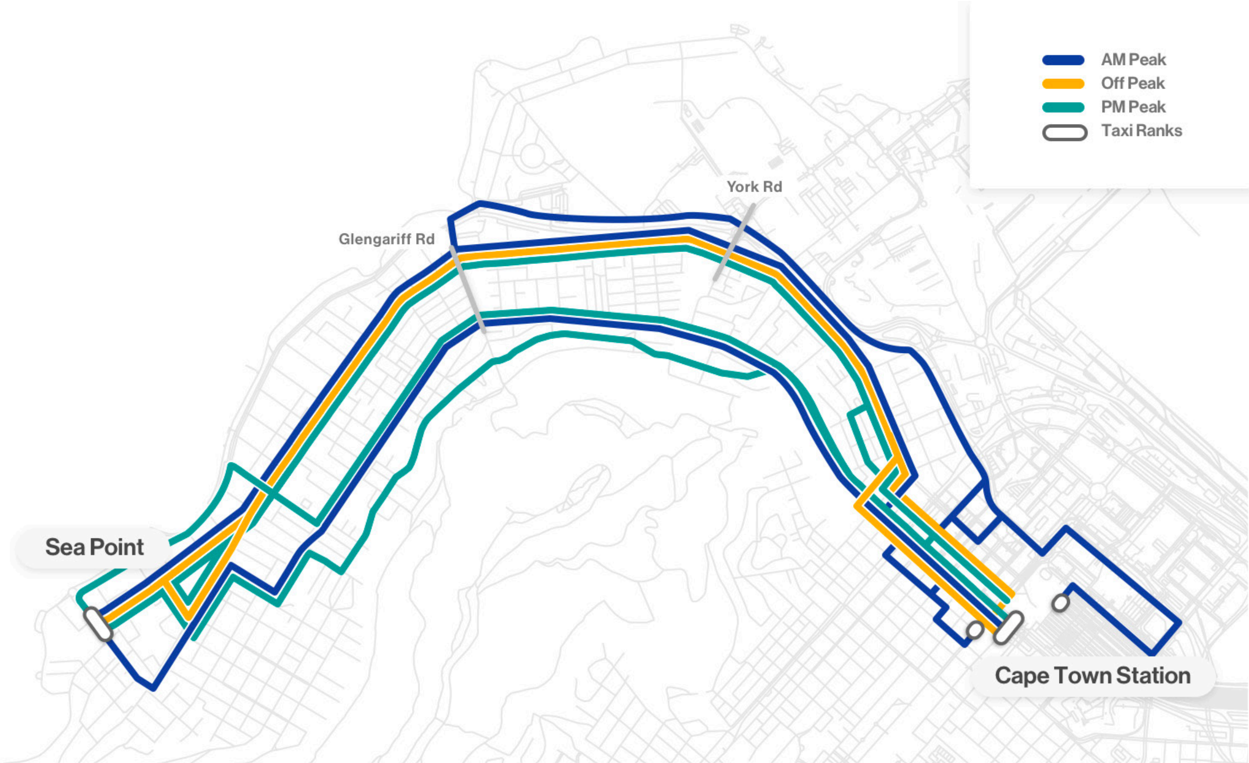

Figure 2: Observed route geometries of Cape Town to Sea Point route in AM, PM and off-peak travel periods

6.2 Trip Duration

We hypothesised that vehicle behaviour at the start and end of a route affects the expected trip duration. Data shows that vehicles travel at different speeds along different segments of the route during the different travel periods (see Tables 1 and 2). We observed drivers travelling at fast speeds in the AM-peak travel period when departing for Sea Point with a full load and returning at slower speeds when departing for Cape Town in order to increase their chances of picking up passengers. The data further showed drivers increasing their speed in the middle segment of the route, between York Road and Glengariff Road, during all travel periods and in both directions.

Table 1: Vehicle speed analysis for MBT travelling to Sea Point

Road Segment	AM-Peak Average Speed (Kph)	Off-Peak Average Speed (Kph)	PM-Peak Average Speed (Kph)
Cape Town to York Rd	28.53	16.19	19.17
York Rd to Glengariff Rd	29.72	36.09	30.02
Glengariff Rd to Sea Point	25.08	14.75	15.81

Table 2: Vehicle speed analysis for MBT travelling to Cape Town

Road Segment	AM-Peak Average Speed (Kph)	Off-Peak Average Speed (Kph)	PM-Peak Average Speed (Kph)
Sea Point to Glengariff Rd	22.22	11.25	15.04*
Glengariff Rd to York Rd	36.09	20.30	33.33*
York Rd to Cape Town	20.08	13.19	22.33*

* PM peak results are based off two trips. Vehicles opted to take alternate routes via High Level Road four out of a possible seven trips. Due to time constraints, we were unable to keep collecting until a successful third trip occurred along the Main Road.

6.3 Fullness Factor

We hypothesised that it is possible to hail a vehicle at any point along the route. While true, we found that a passenger's chances of hailing a vehicle at the beginning of the route in the direction of Sea Point and not at the rank are much lower due to the vehicle's fullness. The data shows that the first expected stopping point for a vehicle in the direction of Sea Point in the AM-peak travel period is at Cape Quarter, around 1.6km into the route (see Table 3). During all travel periods in the direction of Sea Point, MBTs observed left the rank once they were full and on average ran over capacity, meaning that the first place a seat would likely open up was at Cape Quarter. In the off-peak travel period, this distance reduced to around 1.5km and in the PM-peak travel period to 1.3km. On the return trip in the direction of Cape Town and during all travel periods, vehicles depart without a full load and begin picking up passengers immediately (see Table 4).

Table 3: Travel distance and fullness results for MBT travelling to Sea Point

	AM-Peak	Off-Peak	PM-Peak
Avg. Distance to First Stop (km)	1.63	1.50	1.31
Min. Distance of First Stop (km)	1.30	1.40	0.85
Avg. Fullness at Start (number of people)	17	16	17
Avg. Fullness after First Stop (number of people)	17	15	16

Table 4: Travel distance and fullness results for MBT travelling to Cape Town

	AM-Peak	Off-Peak	PM-Peak
Avg. Distance to First Stop (km)	1.64	0.86	0.25
Min. Distance of First Stop (km)	0.13	0.35	0.13
Avg. Fullness at Start (number of people)	1	2	8
Avg. Fullness after First Stop (number of people)	1	5	12

6.4 Route Frequency

We hypothesised that vehicle frequency along a route varies based on time of day and route directionality. Data shows that vehicle frequency changes depending on the time of day and route direction. AM and PM-peak travel periods in either direction on both route segments have mean waiting times of generally lower than that of off-peak travel period times. Prolonged waiting times were observed along Main Road during the latter half of the AM-peak travel period while parallel to this a spike in vehicles travelling along High Level Road was observed. There was an observed maximum waiting time of 6 minutes during the AM-peak travel period near York Road that may have been due to a police roadblock that was set up during the collection period, thereby incentivising taxis to deviate from the Main Road and seek alternate routes.

Route directionality during these periods of the day further affects expected frequencies. From the observed averages and 90% confidence intervals (see Tables 5 and 6), there is a pattern in expected frequencies that differs based on route direction at different times of the day. During the AM-peak travel period expected waiting times between vehicles are longer in the direction towards Sea Point than that towards Cape Town. However, during off-peak and PM-peak travel periods, expected waiting times in the direction towards Sea Point are shorter than those towards Cape Town.

Table 5: Waiting time results for MBT travelling to Sea Point

	AM-Peak		Off-Peak		PM-Peak	
Observation Point	**Glengariff/ Main Rd**	**York/ Main Rd**	**Glengariff/ Main Rd**	**York/ Main Rd**	**Glengariff/ Main Rd**	**York/ Main Rd**
Mean Waiting Time (minutes)	0.33	1.31	1.26	1.77	0.36	0.63
Max. Waiting Time (minutes)	3	6	4	5	3	3
Sample Standard Deviation (minutes)	0.674	1.543	1.13	1.63	0.74	0.82
90% Confidence Interval (minutes)	0.33 +/- 0.20	1.31 +/- 0.59	1.26+/- 0.43	1.77 +/- 0.68	0.36 +/- 0.22	0.63 +/- 0.26

Table 6: Waiting time results for MBT travelling to Cape Town

	AM-Peak		Off-Peak		PM-Peak	
Observation Point	**Glengariff/ Main Rd**	**York/ Main Rd**	**Glengariff/ Main Rd**	**York/ Main Rd**	**Glengariff/ Main Rd**	**York/ Main Rd**
Mean Waiting Time (minutes)	0.26	1.03	1.44	2.05	0.59	0.91
Max. Waiting Time (minutes)	2	6	4	6	3	4
Sample Standard Deviation (minutes)	0.53	1.52	1.19	1.96	0.88	0.89
90% Confidence Interval (minutes)	0.26 +/- 0.15	1.03 +/- 0.55	1.44 +/- 0.47	2.05 +/- 0.86	0.59 +/- 0.28	0.91 +/- 0.31

Further, we observed a change in vehicle frequency at points within a route where multiple routes combined or split up. Expected waiting times were consistently recorded as longer at the route segment prior to the York Road traffic circle where vehicles from other routes entered. For example, as seen in Table 5 during the AM-peak travel period, the maximum time between vehicles in the direction of Sea Point was recorded as 6 minutes prior to the circle as opposed to after the circle where the maximum time was recorded as 3 minutes. With a 90% confidence interval, during the AM-peak travel period it is predicted that a vehicle will come within at least 114 seconds prior to the traffic circle, whereas after the traffic circle a vehicle would be predicted to come within at least 32 seconds. It is important to note that while off-peak waiting times are longer than those during the peak period, greater fullness factors during peak times may mean passengers waiting along the route cannot be accommodated, thereby increasing their wait time. Further research would be needed to investigate whether waiting time is further affected by fullness factors during peak periods.

7. CONCLUSIONS AND RECOMMENDATIONS

This multi-part study in Cape Town highlighted several observations that bear learnings that can improve the reliability of journey planning calculations and provide more accurate information to MBT users.

In terms of route geometry, rank locations change between AM, PM, and off-peak travel periods. A journey planner needs to allow for multiple routes and starting or terminating ranks or points during the different travel periods. While it is possible to hail a taxi on an alternative route, due to the low likelihood, a journey planner needs to continue to direct passengers to the trunk route for pick-ups and drop-offs. Further research needs to be conducted to determine if this is the case across all major routes in the network that have alternative path options as this would have an impact on how data is collected and ultimately processed by journey planners.

In calculating trip duration, a journey planner cannot rely on traditional distance-based or road speed-based routing calculations for expected travel times. Further, knowing the average travel time of the route is not enough to provide an accurate travel time either. If a passenger boards mid-way through a route, the majority of the route's slow pick-up phase could have passed, resulting in less time remaining than would be expected when considering the entire route's average duration. The application of these findings can have a drastic effect on journey planning results. For example, when calculating the estimated time of arrival (ETA) of a passenger trip, a scheduled transit system using continuous stops will be limited to using the average speed of the trip, whereas a navigation engine such as Google Maps might use a combination of the road speed, traffic and historic figures to determine a more accurate result. However, a MBT has unique driving patterns, which when taken into consideration for an ETA, yield even more accurate results than a navigation engine could. For example, the ETA for a trip from York Road to Glengariff Road during the off-peak travel period, using the average speed of the entire route and a continuous stopping behaviour allowing hailing at any point would be 249 seconds. Using a navigation engine such as Google Maps would result in a trip of 180 seconds. However, using average speed changes within the route would provide an estimation of 119 seconds, roughly 34% more accurate than a standard navigation engine. We recommend that a journey planner calculate expected travel times based on the point at which a passenger hails the vehicle using expected travel speeds along the remaining segments.

Though in theory, an MBT can pick up and drop off passengers at any point along a route, in practice the probability of hailing a taxi with available seating capacity varies across the route, but is far less likely if a passenger is within the first kilometre of a rank. Therefore, journey planning calculations need to take into account the probability of hailing to direct a passenger to the nearest likely hailing point where a vehicle will pass with an open seat based on the rank that the vehicle is departing from.

Due to the non-fixed timing nature of MBT, vehicle frequencies, and ultimately passenger waiting times, need to be modelled as non-linear confidence base intervals rather than fixed, interval-based times. This will help journey planners provide accurate estimations of expected waiting times at different times of the day. The traditional model for route-based frequencies assumes a linear frequency along the entire route. This needs to be adapted to allow for a change in frequency within a route. Routes need to be segmented to accommodate non-linear variation in frequencies in different directions, particularly along segments where factors affect the number of vehicles servicing the route such as intersections where route alternatives can be taken and where several routes overlap.

To understand how many times data needs to be collected to represent reality, further research needs to look at real-time information on a route versus the information predicted by the journey planner. This can lend deeper insight into the advantages and disadvantages of additional data collection versus developing a system-wide real-time solution. Further research should also take into consideration other factors affecting spatio-temporal dimensions including weekends, severe weather conditions, and large events.

8. REFERENCES

Eros, E et al. 2014. Applying the General Transit Feed Specification to the Global South: Experiences in Mexico City, Mexico–and Beyond. *Transportation Research Record*, 2442 (1), pp. 44-52.

Klopp, J and Cavoli, C, 2017. The Paratransit Puzzle: Mapping and Master Planning for Transportation in Maputo and Nairobi. In: T Priya and K Lucas, ed., *Urban Mobilities in the Global South*, 1st ed. New York: Routledge, pp. 95-110.

Ndibatya, I et al., 2014. An Adaptive Transportation Prediction Model for the Informal Public Transport Sector in Africa. *IEEE ITSC*, Qingdao, China.

Ndibatya, I et al., 2017. Mapping the informal public transport network in Kampala with smartphones. *Civil Engineering : Magazine of the South African Institution of Civil Engineering*, 25 (1), pp. 35-40.

Oloo, F, 2018. Mapping Rural Road Networks from Global Positioning System (GPS) Trajectories of Motorcycle Taxis in Sigomre Area, Siaya County, Kenya. *ISPRS International Journal of Geo-Information,* 7, pp. 309.

Stats SA (Statistics South Africa). 2011. Census 2011. [online] Available at: http://www.statssa.gov.za/ [Accessed 12 October 2018].

Williams, S et al., 2015. The digital matatu project: Using cell phones to create an open source data for Nairobi's semi-formal bus system. *Journal of Transport Geography*, 49 (C), pp. 39-51.

Zegras, C et al., 2014 Tracing a path to Knowledge? Indicative user impacts of introducing a Public Transport map in Dhaka, Bangladesh. *Cambridge Journal of Regions, Economy and Society*, 8, no. 1, pp. 113-129.