

MAPPING MINIBUS-TAXI OPERATIONS AT A METROPOLITAN SCALE – METHODOLOGIES FOR UNPRECEDENTED DATA COLLECTION USING A SMARTPHONE APPLICATION AND DATA MANAGEMENT TECHNIQUES

J COETZEE*, CHRISTOFF KROGSCHEEPERS, JOHN SPOTTEN

* GoMetro SA (Pty) Ltd, South Africa
ITS Engineers (Pty) Ltd, South Africa
City of Cape Town, Transport and Urban Development Authority, South Africa
Email: justin@gometroapp.com

ABSTRACT

Paratransit forms the basis of public transport systems in many of worlds emerging cities. However, there is limited information available on the extent of operations, the route network and passenger demand. Manual passenger counting techniques alone do not provide sufficient information on the extent of network coverage and passenger turnover for the routes on a network. Exploring techniques and methodologies to collecting data on paratransit operations using mobile technology helps us get an informed picture of the transportation system, and helps us make smarter choices when it comes to improving the existing system. Application of mobile technology to collect paratransit data in emerging markets has enabled the development of new data sources. These data sources help discover the extent of network coverage and enable the profiling of routes on the network from the collected passenger information. The objective of this paper is to demonstrate a methodology on the application of mobile technology to conduct on-board vehicle surveys for paratransit services. The mobile application collects route traces, stop location, and individual trip boarding and alighting pairs. On-board surveys were carried out on a city-wide-basis by field teams making use of a mobile application for data collection on operations and passenger activity along an estimated 800 licensed minibus taxi routes in the Cape Town municipal area using the prescribed methodology. The paper further demonstrates a custom- data management platform for analysing and visualising the data collected using the mobile application.

Introduction

In many of the world's emerging markets, paratransit forms the basis of public transit systems. However, limited open and standardized data is available on the operational characteristics of paratransit services in emerging cities. The public transport system is composed of multiple operators serving multiple routes. A combination of all these routes together forms a complex transport network constantly evolving in response to the system changes in demand and supply. Several benefits are accrued in understanding the network and operations in geospatial terms as it unlocks benefits for all role players involved in the supply of public transport services

(Demissie and Phithakkitnukoon, 2016). For large-scale public transport systems, it is not feasible to record and compile the route network definitions and geospatial characteristics, its operating frequencies, capacities, ridership and travel times using manual paper-based methods or expensive data logging devices (Jiang *et al.*, 2013). This is because the public transport system consists of a large number of agents interacting non-linearly through random complex feedback loops and no clear points of control or static observation (Zegras, Eros and Mehndiratta, 2015).

To model integrated public transport networks, knowledge of public transport routes serviced by paratransit services is required. Schedules for buses, rail and BRT services are known, however the precise network and service frequency of paratransit services has not been established at municipal scale (Dhingra, 2011). Paratransit services are constantly evolving and creating new system structures and new patterns of behaviour. This complexity presents challenges in integrating planning and decision-making processes as well as developing comparative analyses of urban transport systems. To overcome these challenges significant efforts have been channelled towards applying smart phone based solutions to understand operational characteristics in geospatial terms.

In 2014, a South African technology company, GoMetro developed a smartphone application for transport data collection called GoMapp, and subsequently GoMetro Pro. Ndibatya et al (2017) demonstrated a mapping approach through a rigorous field capturing process to provide insights on routes, operations, and characteristics of paratransit services in Kampala, Uganda. The approach utilised the smartphone application developed by GoMetro paired with a rigorous survey design, data collection, and quality assurance methodology.

The initial technology, approach and methodology presented by Ndibatya et al (2017) used a functional back-office that supported the field capturing process. Post-processing the data using a variety of tools before exporting it to .csv or .shp file formats was a very labour-intensive exercise. The post-processing functionality of the approach was functional for hands-on data management and analysis but had limitations on formalising data quality assurance practices and analysis - which had to be carried out as an export from the system into .csv or .shp and updated in software such as Excel or QGIS. This presented challenges at scale encompassing areas of project management and data management to ensure the provision of high quality data.

Similar approaches using mobile phones to understand various aspects of informal transport networks in emerging cities have been performed. Williams et al (2015), tested the application of the geolocation capabilities of mobile technology to collect route and develop a GTFS data feed for semi-formal transit systems in Nairobi. This exercise demonstrated the possibility of adapting GTFS data for semi-formal transit and offered an approach to institutionalising a process of data updating and sharing. Saddier et al (2016) demonstrated the application of smartphones and digital technologies to map jitney routes in Accra. This approach offered unique perspectives on the methodological and operational requirements of collecting public transport data with limited time and resources. The methodology presented in this paper utilises a custom-made management tool with built-in quality control measures to reduce the post-processing time as well as produce reports to minimise the time spent on analysis. It uses a custom management platform that makes use of machine learning and geospatial techniques

to improve the data post-processing and quality assurance procedures. Figure 1 below shows the overall approach to designing a survey for network discovery and route profile information for paratransit services.

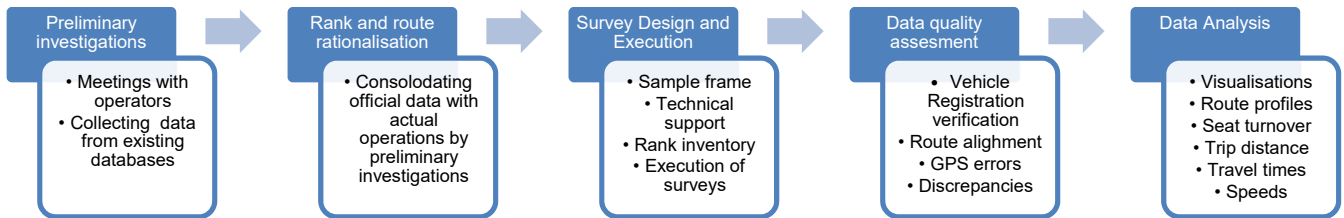


Figure 1 : Steps to conducting network discovery and route profile for paratransit services

The improvements to the first methodology for app-based transport network characterisation includes:

- Capturing passenger origin and destination
- Capturing passenger demographics – Age, Gender and Ethnicity
- Offline data storage on mobile phone
- Automated data cleaning and correction
- Post processing input error detection
- Report generation on the custom development tool

The methodology brings important benefits to city authorities and transport operators related to network discovery and route profiling information. Network discovery enables the understanding of the network coverage of paratransit services, and route profiling enables the understanding of passenger seat turnover on a specific surveyed route. Using the extracted mobility patterns, it will be possible for public transport operators to holistically understand the transport demand and manage their service offerings accordingly. With better information on the mobility patterns of people, city authorities have more opportunities to resolve social and economic problems, and can make informed decisions in terms of future sustainable urban infrastructure planning to better serve citizens.

Methodology

Cape Town is a metropolitan city in South Africa with a population of 4 million. Approximately 51.4 percent of all commuter trips in the Cape Town metro area use paratransit services (Schmidt, 2014). Over the last number of years Minibus Taxis (MBT) have become an increasingly important component of public transport in Cape Town. The latest review of Cape Town’s Comprehensive Integrated Transport Plan (CITP) estimated that there were about 24 000 MBT vehicles registered as of May 2015. MBTs provide connections between various origin and destination pairs within the City, but due to the informal nature of their services. At

the start of the project, it was difficult to know exactly how many routes or passengers are utilising this type of mode. It is estimated that there are around 800 routes carrying approximately 600 000 passengers per day.

To develop a holistic understanding of network coverage, route profiles and passenger information for Cape Town paratransit services, rank counts, and cordon counts only are inadequate. Whereas, on-board surveys provide insight into the supply, demand, and performance characteristics of paratransit services at a per-vehicle resolution. A citywide on-board vehicle mapping exercise to create new transport data on existing operational characteristics of paratransit services in Cape Town is presented.

The mapping exercise was performed for the City of Cape Town`s – Transport and Urban Development Authority – (TDA) under the management of the consultant - ITS Engineers. The objectives of the citywide mapping exercise were to collect:

- Taxi Route Profile information
- Passenger demand
- Passenger link volumes
- Stop boarding and alighting information

The total project data collected over a 12-month period is reported below. As is evident, this is a very sizable dataset, containing 379,643 unique passenger journeys as observed by the mapper, each with unique travel time, origin-and-destination pairs, travel distance and duration. A summary of the data collected during the extensive 12 month period is summarised in Table 1 below.

Table 1 : City of Cape Town Mapping metadata

Number of Trips	28,981
Number of Routes	1,018
Number of Unique Vehicles	10,636
Number of km mapped	338,237
Number of unique passenger journeys	379,643
Number of stops	159,529
Fare Revenue measured	R4,024,534.58

This data can be used to inform reform programs for paratransit operators looking at introducing new passenger technology-based services. In addition operators can develop and explore new operational business models using a systematic methodology to assist with decision making.. The steps can be classified into three broad stages. These three stages are in line with the overall approach taken to deliver the objectives of the paper as shown in Figure 2.

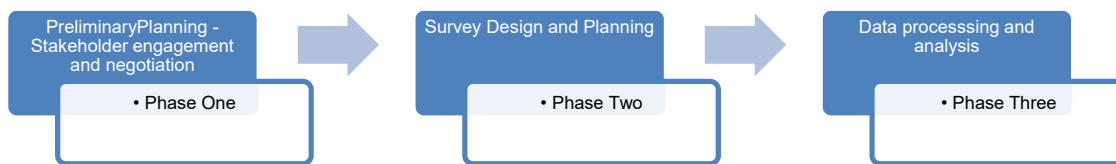


Figure 2 : Steps followed to implement the survey methodology

Phase One - Preliminary planning - Stakeholder engagement and negotiation process

The objective of this phase is to obtain an as-is understanding of the existing operational characteristics such as locations of gazetted taxi ranks and official route names. Prior to commencing mapping, formal consultation is carried out with the governing structures of the minibus taxi operators operating in the survey area. Through this process it was discovered that it is important to build confidence and trust in the technology and the on-the-ground operational team in order to ensure willingness to participate. A thorough broad-based communication strategy is required to develop and build support around any data collection activities for network and operational discovery surveys (Gaibe and Vanderschuren, 2010). This is followed by interviews with the minibus taxi associations operating on the official routes to confirm all the gazetted ranks and routes as well as identify non-gazetted taxi ranks and routes. During this phase of the project, challenges included accessibility of some taxi ranks attributed either to security related issues and or uncooperative taxi operators. To overcome these challenges the project management team held additional one-on-one meetings and workshops demonstrating and clarifying the objectives of the data collection process with concerned operators. This phase articulated the purpose and objective of the survey and ensures mutual agreement on the execution and duration of the survey between the taxi associations.

Phase Two - Survey planning, design and execution

Survey Planning and design

Phase 2 used output information collected from the stakeholder engagement process. Based on the baseline operating characteristics identified in phase one and the specified data requirements, the survey was designed to determine the target population, sampling frame, sampling size and sampling method. Based on the design parameters a pilot study was conducted to test all aspects of the survey plan and to ensure that the operational arrangements are feasible and provide the required data compatible with the proposed survey objectives. This is an iterative stage and the pilot underpins the success of the proposed survey plan. The survey design accounted for the key defining parameters such as sample size estimation of the vehicles, survey areas and, survey periods as identified from the initial stakeholder engagement process. The sampling of the total fleet size is required because it is logistically expensive and practically not feasible to survey all vehicles in the population.

Therefore, the survey is designed to observe selected vehicles within the population to infer insights about the characteristics of the population. The effectiveness of the mapping exercise is dependent upon choosing the appropriate sample. Therefore, the sampling design is a fundamental part of the methodology and should include the following elements:

- Define target population
- Define sampling unit
- Selection of sampling frame
- Choice of sampling method
- Consideration of likely sampling errors and biases
- Determination of sample size

Four days of the week were picked to conduct the on-board surveys, which comprised of two-week days (Tuesday/Wednesday/Thursday), a Friday and Saturday. For each day, data was collected in the morning peak(4am-11am), afternoon peak(3pm -5pm) and off peak periods(11am – 3pm). Based on the operating baseline data from the city of 800 routes and 120 official gazetted ranks, each rank was identified as an origin and all destinations from that rank were identified based on the baseline data. For each origin and destination pair the survey design required a minimum of three outbound trips and three inbound trips from each origin-destination point for the morning and afternoon peak periods and a minimum of two inbound trips and two outbound trips for the off peak periods.

The comprehensive survey managed to collect data on 550 routes from 85 ranks across the City of Cape Town public transport network. Security issues were experienced on 40 of the 550 surveyed routes that the survey discovered and confirmed. The survey discovered that certain routes and ranks in the baseline data of the city records were no longer operational. 250 routes do not exist

The difference in the outputs of the comprehensive survey and baseline data collected from the City can be attributed to survey challenges highlighted in phase one concerning mapper security at ranks and uncooperative associations. In addition, the survey discovered that certain routes in the baseline data were no longer operational.

Survey execution

Upon reaching consensus on the project requirements with all stakeholders, enumerators are mobilised to meet the survey design specifications. A project management and oversight environment is developed based on the survey design parameters. This includes all vehicles to be mapped, user accounts of mappers and project managers of the data collection exercise. The mapping methodology also focuses on capacity building that empowers operators, drivers, and owners with knowledge on how the technology works. Operators are encouraged to supply volunteers from their operating areas to allow access to opportunities for the economically able in those communities. The execution of the methodology is dependent on the ability to monitor and evaluate the activities involved in the entire data collection process. An operating manual is designed for each survey to ensure strict adherence to prescribed data collection specifications. The database administrator implements data quality management to ensure good data quality for the proper working of algorithms and tools. All datasets are strictly checked

according to the quality management procedures, which require a minimum set of meta-information.

GoMetro Pro App

The methodology is hinged on two interfaces, the mobile mapping platform, and the data management platform. The GoMetro Pro app is a smartphone application, which allows enumerators to collect public transport taxi operational data with an emphasis on paratransit operations. The data collected on the mobile smartphone is stored and managed on the GoMetro Pro platform.

The Figure 3 below shows the main characteristics of the GoMetro Pro app, which collected public transport data for urban mobility. The application collects vehicle, route trace, location of stops and passenger information on-board in real time. The trip information can be stored locally on the mobile device, exported or uploaded onto the cloud server for storage and analysis. The app presented by Ndiabanya et al (2017) did not yet collect the demographic data for all boarding and alighting passengers. The feature improvements now allow this data to be collected.

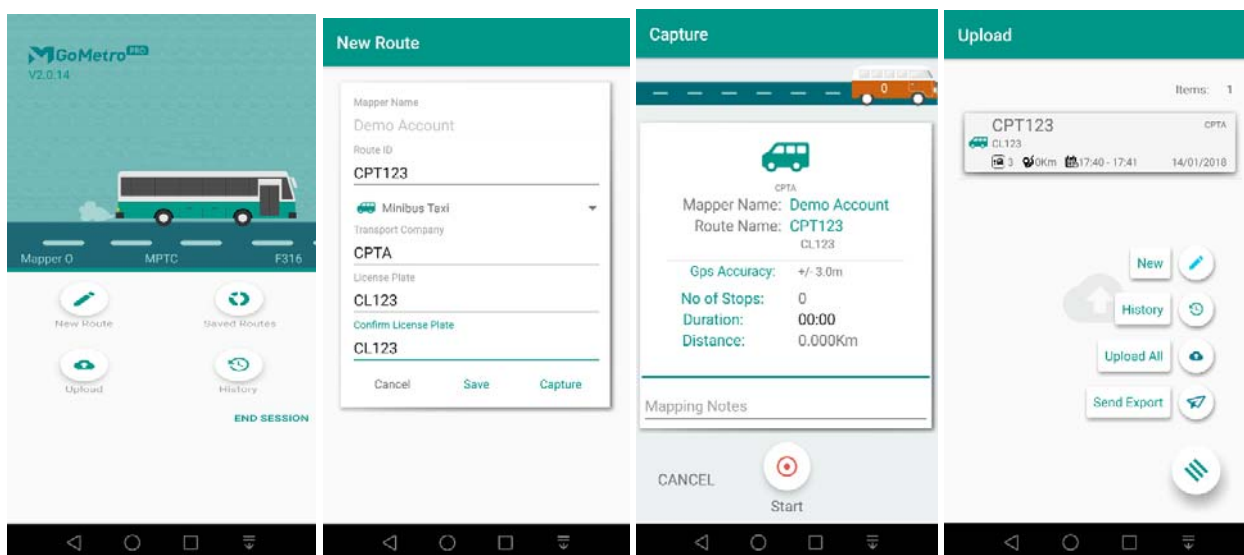


Figure 3 :Mobile application showing simple login and data collection interface.

GoMetro Pro Data Manager

The GoMetro Pro data manager platform compiles, processes and analyses all surveyed data collected using the GoMetro Pro app. The data management system includes three layers: geographic layer, transport data layer, and the data collection quality layer. These three layers are built on a stable and compact cloud-based architecture to improve the robustness of the custom management platform. The project management tool is a new contribution to the GoMetro Pro mapping software, has been specifically developed for data management of large-scale field survey work, and has the following capabilities.

- Trips uploaded from any device for a project can be viewed in a list format
- Each trip has a unique Trip ID assigned to it
- Each trip can be sorted by vehicle registration, trip ID, route description, mapper name, company name, start time, travel time, distance mapped, revenue earned, date

mapped, seat discrepancies, fares, number of stops, number of passengers, mapping notes and GPS error rate as a percentage of a route.

- The unique passenger boarding and alighting co-ordinates, time of boarding and time of alighting, individual travel distance and travel times, as well as demographic information such as age range, gender, and race is available per trip.

The data management system allows data to be selected and exported on a trip-by-trip basis in the following formats for further analysis:

- csv files of all trips in a selection
- KML files of trips surveyed
- Visualisations of all trips surveyed

The platform allows for analysis to be performed during the survey checks for data completeness and fix any GPS issues that are a result of poor/ lost GPS signal during operations. Information on breakdowns and any service disruptions can be consolidated for analysis from the trip comments or notes.

Phase Three - Data processing and analysis

Phase 3 involves second level data cleaning and processing for reporting. The algorithm and machine learning techniques work on the three levels of the management platform, geospatial layer, network layer, and the data layer.

The analysis derived from the data management platform include trip visualisations, passenger transit patterns, and network characteristics. These can be visualised and supported in an interactive visualisation environment as shown in Figure 4 with an example of a route profile showing the routes and stops. The information extracted from the custom developed platform includes passenger demand, revenue generation, asset utilisation levels; network utilisation levels and several trip related demographics. Figure 5 shows the total number of passengers demand recorded on-board a vehicle throughout a typical survey day.

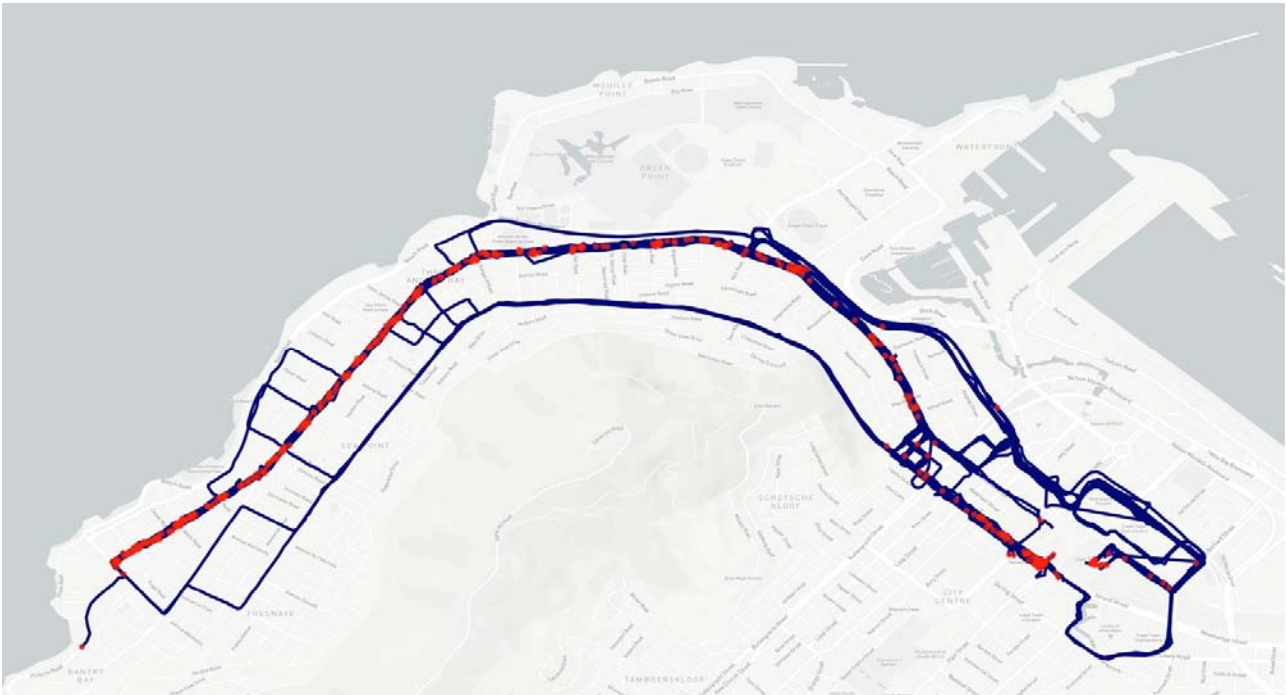


Figure 4 :Typical results map for route profiling (from Seapoint to Cape Town CBD)

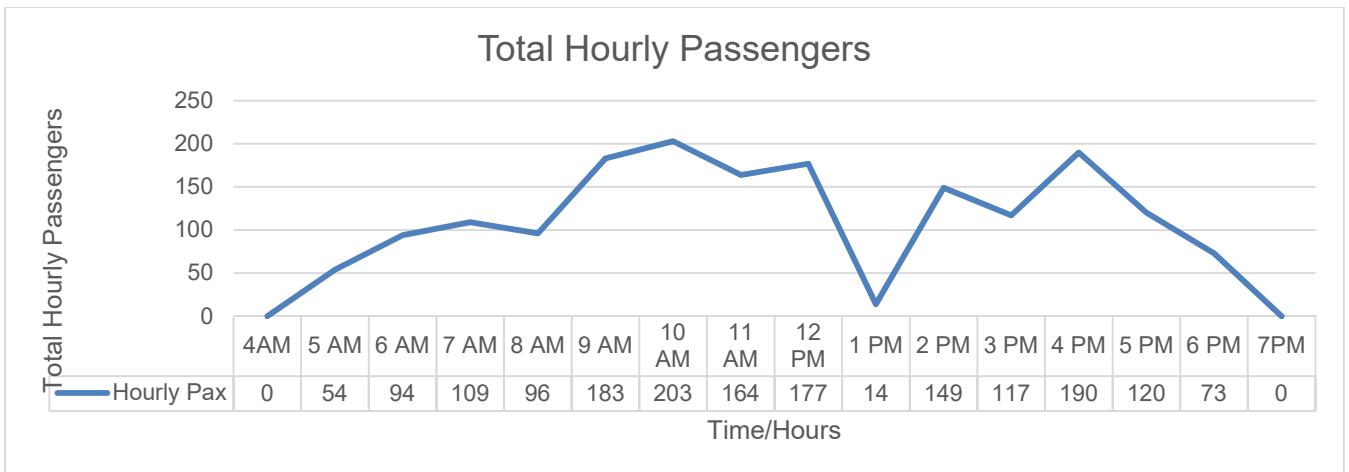


Figure 5 : Hourly passenger demand for a certain route and direction for a particular survey day

Discussion of Results

The approach presented in the paper helps discover the extent of paratransit network coverage and the passenger demand profiles on the routes forming the network. In addition to unlocking value to stakeholders, such as planners and operator associations in the provision of public transport, the methodology presented in this paper is a starting point for designing an effective data management system and a conceptual framework for developing a comprehensive platform to develop transportation models and other analytical tools needed for planning and system analysis. The minibus taxi network is made up of a collection of nodes such as taxi ranks or major informal stops acting as major source nodes and sink nodes of travellers in a

network, and can be digitally modelled and quantified as a network, route operations, and individual vehicle basis.

Lessons learnt from this initiative range from the kind of systems required regarding on board vehicle data collection and the need to have consistent definitions of paratransit route and network information for comparing minibus taxi services in cities and regions. Developing data management capabilities improve the quality of transport data collected by streamlining the post-process and application of data (Williams *et al.*, 2015), (Saddier *et al.*, 2016). Tools for transportation data management improve the collection, management and reporting of transport data. More importantly is the impact that good quality data has in informing the decision-making processes of planning authorities responsible for public transport systems. The custom management tool contributes innovatively in supporting the lifecycle management of informal paratransit data. Supporting this is a clear definition of indicators, and consistent good quality data. A challenge related to consistency is that of defining route codes that align with the cities naming convention and the inherent complex nature and organisation of mapped paratransit services as they are analysed.

The most significant outcome from the mapping project is the new information on vehicle seat turnover and individual passenger boarding and alighting time and location. It is not possible to understand passenger movement information to this degree using cordon counts and rank surveys. Cordon and rank counts do not provide an accurate representation of the passenger turnover in the public transport network, nor can they establish the vehicle kilometres travelled by a vehicle. The methodology presented in this paper overcomes these shortcomings by using a custom-designed mobile application in the hands of an on-board enumerator. The data collected using the methodology is a new and unique data set for paratransit operations, and it is recommended that this on-board vehicle survey data should be utilised to supplement cordon and rank counts, where these are performed.

Future Work

The transport registry minimum requirements with respect to the needs of route-based data collection and surveys must include the fleet size, frequencies, and service capacity per mode between origin and destination on each route. The current requirement for a Current Public Transport Record is for the route to be described in words. There is no geospatial referencing of the routes in a geographic information system format for group analysis or cross-referencing. It is proposed that the Current Public Transport Record should require .shp files as an essential component – and not just route descriptions in words. The Current Public Transport Record minimum requirements should require geospatial data and not text-based records only. These network characteristics have been successfully measured using the methodology presented in this paper for minibus taxis and can be collected in any municipality. To understand the demand and supply attributes of a public transport system requires knowledge of all vehicle operations in the entire public transport network. That is to say, surveys to understand the performance of public transport systems should also include neighbouring municipal areas bordering the City of Cape Town. In addition, passenger boarding and alighting data for buses operating in the Cape Town area should be collected. Data collected from on-board surveys can also augment smart card data from the BRT system to gain further insight into the passenger demand characteristics in and around the city. This provides a better overview of the demand and supply

attributes of the public transport system, which is key to providing an integrated demand responsive service. Future work investigates the applicability of new driver technologies to understand vehicle operations and introduce new passenger services. The objective is to develop a framework for an incremental and inclusive approach to organically grow the applicability of new transport technologies in emerging cities.

Conclusion

The need to improve the availability and quality of transport data has been widely acknowledged. Given the enormous challenges being faced by urban mobility and transport policy makers, fresh and valid data of different kinds is extremely important for measuring the success of policies and strategies. Transport demand estimation is a challenging task, particularly in emerging markets mainly because of the time-consuming data collection requirements. Paratransit service providers have difficulty in making important decisions related to service efficiency without accurate demand estimation. In this paper, an in-depth understanding of the factors influencing public transport data collection at scale is shown. A custom developed management system to assist with data management for large quantities of transport data is presented. The custom management tool allows post-processing and analysis to be streamlined using machine learning and geospatial techniques enabling real-time data collection and assessment. The methodology determines passenger demand for paratransit services using a smartphone application and custom data management tool. The data generated with this methodology is the same O-D data that is collected by smart card systems. With better information on the mobility patterns of people, city authorities have more opportunities to resolve social and economic problems, and can make informed decisions in terms of future sustainable urban infrastructure planning to better serve citizens.

References

Demissie, M. G. and Phithakkitnukoon, S. (2016) 'Inferring Passenger Travel Demand to Improve Urban Mobility in Developing Countries Using Cell Phone Data: A Case Study of Senegal', *IEEE Transactions on Intelligent Transportation Systems*, PP(99), pp. 1–0. doi: 10.1109/TITS.2016.2521830.

Dhingra, C. (2011) *Measuring Public Transport Performance : Lessons for Developing Cities*.

Gaibe, H. and Vanderschuren, M. (2010) 'An Investigation into the methodology of mini-bus taxi data collection as part of the Current Public Transport Record: A Case Study of Stellenbosch in the Western Cape', pp. 16–19.

Jiang, T. *et al.* (2013) 'Sustainable Transport Data Collection and Application: China Urban Transport Database', *Mathematical Problems in Engineering*. Hindawi, 2013, pp. 1–10.

Ndibatya, I., Coetzee, J. and Booysen, T. (2017) 'Mapping the informal public transport network in Kampala with smartphones', *Civil Engineering -Siviele Ingenieurswese*. South African

Institution Of Civil Engineering (SAICE), 2017(Volume 1), pp. 35–40.

Saddier, S. *et al.* (2016) 'Mapping the Jitney Network with Smartphones in Accra, Ghana', *Transportation Research Record: Journal of the Transportation Research Board*. Transportation Research Board of the National Academies, 2581, pp. 113–122.

Schmidt, I. (2014) 'NHTS Provincial Report Western Cape Profile Stats SA Library Cataloguing-in-Publication (CIP) Data NHTS Provincial Report – Western Cape Profile', *Statistics South Africa. Pretoria: Statistics South Africa*.

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*. Pergamon, 49, pp. 39–51.

Zegras, C., Eros, E. and Mehndiratta, S. (2015) 'Digitalized Transportation Data as Disruptive Innovation: Prospects in Mexico City', (January).