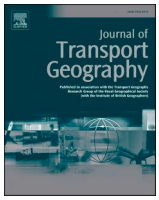
Journal of Transport Geography 128 (2025) 104355

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo



# Quantifying informal public transport using GPS data

Lourens de Beer [a,*], Christo Venter [b], Lourens Snyman [c]

[a] Department of Civil Engineering, University of Pretoria, South Africa
[b] Department of Civil Engineering, Centre for Transport Development, University of Pretoria, South Africa
[c] Department of Geography, Geoinformatics and Meteorology, University of Pretoria, South Africa

ARTICLE INFO

*Keywords:*
Paratransit
Public transport
GPS
Data
Trip volumes

ABSTRACT

Informal public transport modes transport the largest number of passengers in most developing countries. Despite its significance, limited information is available on the extent of its operations, and passenger counts alone do not provide sufficient insight into network coverage or passenger turnover. GPS tracking has emerged as a valuable tool, yet its potential for understanding minibus taxi operations at the road segment level remains underexplored. GPS studies of informal operators have rarely been extrapolated to volume counts per time period, due to statistical problems (non-representative sampling) and small sample sizes. This paper addresses this gap by developing a methodology to determine the minibus taxi vehicle trip count per street segment from GPS data, to map routes, and identify high-traffic corridors, with an illustrative application in the City of Tshwane, South Africa.

The methodology includes data inspection, addressing limitations, and counting trips per street segment using a database and QGIS visualisation. Additionally, the paper outlines detailed steps in QGIS for processing GPS data. We show that the method delivers plausible results at the segment level. The methodology can help to address the global South's need for data-driven interventions in its predominant public transport mode.

## 1. Introduction

Informal modes provide the bulk of public transport in the global South (Cervero and Golub, 2007). Recognising that informal public transport (also called paratransit or popular transport) is critical to achieving sustainability goals, many governments are pursuing efforts to regulate, upgrade, or formalise these modes. A significant barrier to these efforts is the limited information that exists on the extent of operations, passenger demand patterns, and the route networks of informal operators. New data collection methods using Global Positioning System (GPS) tracking and hand-held smartphones have emerged to help fill the data gap. The most common application has been to geolocate points of interest (e.g. passenger stops and ranking locations) and service routes (by tracking a sample of vehicles), leading to the creation of new route maps for passengers (Klopp et al., 2019). Analysis of the GPS data has further provided valuable insights on operational aspects like route typologies (du Preez et al., 2019), operating efficiencies such as speed (Zeeman and Booysen, 2014), delay (Ukam et al., 2024) and energy consumption (Hull et al., 2022), and passenger loads (Saddier and Johnson (2018).

From a transport planning and engineering perspective, existing work has been less useful as it has failed to account for informal operations as vehicular flow. Transport planners are often interested in measuring daily or hourly volumes of vehicles of different classes on particular roadways, as volumes (together with road capacity) determine localised congestion and delay. Volumes are also important for designing infrastructure (e.g. traffic signal settings or intersection layouts), and to identify locations for upgrades (e.g. high-volume corridors for providing public transport priority). Typically, GPS studies of informal operators have not been extrapolated to volume counts due to statistical problems (non-representative sampling) and small sample sizes. As larger and better samples become available, a method is needed to quantify informal operations at the road segment level and make it useful for multimodal transport planning.

This paper introduces a methodology to estimate the number of informal vehicle trips per time period on individual street segments within a city using GPS data. We implement the method on the QGIS software platform, an open-source Geographic Information System (GIS) environment used for creating, viewing, editing, and analysing geospatial data. The method is illustrated using a large dataset of informal

* Corresponding author.
*E-mail addresses:* lourensrdb@gmail.com (L. de Beer), Christo.Venter@up.ac.za (C. Venter), Lourens.Snyman@up.ac.za (L. Snyman).

https://doi.org/10.1016/j.jtrangeo.2025.104355
Received 30 January 2025; Received in revised form 19 May 2025; Accepted 3 July 2025
Available online 11 July 2025

minibus taxis that were tracked in the City of Tshwane, South Africa. We show that the method delivers plausible results at the segment level. We also illustrate its usefulness for transport planning by, firstly, examining variations in trip counts across days of the week, and secondly, using the trip counts to identify and visualise eight high-demand minibus taxi corridors. By going beyond simply mapping routes, the method can thus help guide planners and engineers in spatially targeting priority interventions where they may have greatest impact.

While GPS-to-road matching techniques are well established, the contribution of this study lies in their application to a large dataset of minibus taxi operations in Tshwane. This captures the spatial and temporal flow patterns at a road segment, which has not been done for the informal transport sector. We also demonstrate a methodology for addressing GPS data quality issues and extracting transport planning intelligence that is specific to informal transport operations, outlining a process that can be replicated for estimating trip volumes using open-source GIS tools in other urban contexts in the developing world.

## 2. Literature review

In many cities across the developing world, the primary mode of public transport is informal paratransit, most commonly provided with low-capacity minibuses (Kumar et al., 2021). This system is particularly dominant in African cities, where the share of road-based public transport varies between 65 % in Yaoundé, 72 % in Johannesburg, 82 % in Algiers, 86 % in Accra, and 98 % in Dar es Salaam (Randall et al., 2023). These modes operate largely outside formal government regulation, usually lacking fixed schedules, designated stops, and standardised fare systems. Services are provided by numerous small-scale operators without central planning or coordination. Consequently, there is no central collection of data on routes, services, or operations, which means little information is available regarding vehicle stopping locations, passenger boarding, scheduled trip times, or service frequency. This contrasts with many formalised bus systems that generate digital data using built-in vehicle tracking devices and cameras. Access to such information for informal public transport modes would enable transport planners to design better transport systems and prepare passenger information (Klopp et al., 2015).

### 2.1. Traditional data collection method

Rank surveys – counting the number of minibus taxis entering and leaving a taxi rank and the number of passengers on-board– are the most common method for collecting data on this public transport mode in South Africa (van Zyl and Labuschagne, 2008). These surveys are expensive and inaccurate when significant numbers of vehicles are not rank-based (Gaibe and Vanderschuren, 2010). Additional roadside monitoring surveys may provide snapshots of vehicle and passenger volumes *en route*, but Coetzee et al. (2018a) note that such methods fail to provide sufficient information regarding passenger turnover along routes or the network coverage of paratransit services. This has led to the adoption of GPS tracking techniques to better determine driving patterns and operating characteristics of informal operators.

### 2.2. Advancements in GPS tracking techniques

Initially, equipment installed on an operating taxi, depending on the research required, included (van Zyl and Labuschagne, 2008): on-board computers, GPRS modems and antennas, GPS antennas and receiving modules, door sensors, passenger counters, security cameras, and LCD monitors. These pieces of equipment can be used to obtain automatic vehicle locations and passenger counts. Additionally, they enhance on-board security and provide opportunities for advertising purposes.

Although GPS tracking techniques provide valuable insights, they are expensive. In 2008, the cost of this installation was approximately 3000 USD per vehicle (van Zyl and Labuschagne, 2008). As an alternative, smartphones with built-in GPS tracking functionality can also be used to track vehicles and passengers. Klopp et al. (2015) found that the integrated GPS tracking function of mobile phones effectively captures important public transport information. They also noted a demand for both open data and information in digital and paper formats. The authors emphasised that building networks of trust and collaborations is key to collecting and disseminating data in the era of "big data," especially with the ubiquity of crowdsourcing devices.

### 2.3. Application of GPS in paratransit services

The number of applications of GPS in paratransit services has grown significantly over the last decade (du Preez et al., 2019). For example, Coetzee et al. (2018b) describe the development of a smartphone application to conduct on-board surveys in paratransit vehicles. The application was designed to collect stop locations, route traces, and individual trip boarding and alighting pairs.

The initial intention behind collecting GPS data was to develop route and system maps of informal services, for the benefit of passengers (who lack other sources of information on the extent of informal services), and of authorities (who need information for planning and regulation purposes). Authors like Klopp et al. (2019), Vergel-Tovar et al. (2022), and Zegras et al. (2015), describe efforts to produce such maps in different parts of the world.

However, the use of smartphones and GPS tracking in paratransit services also has the potential to improve efficiency and profitability for operators. van Zyl and Labuschagne (2008) stated that the benefits paratransit operators could gain from implementing GPS tracking outweigh the negative perceptions of being monitored by transport authorities.

Several studies have demonstrated deeper insights into operating behaviour offered by GPS data. Saddier and Johnson (2018), analysing paratransit driver preferences in Accra, Ghana, found that vehicles typically operate with high load factors but conduct limited rotations daily. Operators often spend more time queuing at stations than traveling with passengers onboard, reducing profitability. Other studies have employed GPS tracking devices to determine driving speeds and operational patterns of minibus taxis (Mungadze, 2019; Bulbulia, 2023; Zeeman and Booysen, 2014; Ukam et al., 2024). Ndibatya et al. (2016) studied informal operations in Kampala, Uganda as a "self-organising system", using GPS-enabled smartphone devices to trace routes, identify stops, and analyse drivers' strategies. More recently, GPS data has been used for studying the electrification potential of minibus-taxis in Africa (Hull et al., 2022). Rix et al. (2022) note that for electrification planning, vehicle-based tracking surpasses traditional passenger-based methods, enabling more accurate estimation of energy needs and infrastructure requirements in African cities. Jia et al. (2022) used GPS data collected in Maseru and Gaborone to identify strategies to improve the efficiency and quality of paratransit in these cities.

Voluntary tracking of minibus taxi operators can also reveal insights into potentially dangerous driving behaviours. Booysen and Akpa (2014) found that on the 1200 km route between Cape Town and Mthatha in South Africa, many minibus taxis disregarded the legal speed limit of 100 km/h, with speeds reaching up to 159 km/h. Factors such as route selection, departure time, direction of travel, and whether the driver was also the vehicle owner were found to influence these behaviours.

In conclusion, there is growing evidence that GPS-based methods can provide not only a geospatial understanding of informal transport networks but also a basis for improving public transport planning and accessibility. A key limitation of studies to date is their small sample sizes: the number of paratransit routes surveyed typically vary between less than 10 (e.g. Saddier and Johnson, 2018; Coetzee et al., 2018a) and approximately 315 (Saddier et al., 2017). An exception is Coetzee et al. (2018b), who surveyed approximately 800 minibus taxi routes in Cape Town using GPS-enabled smartphones. Subsequent analysis of this data

by du Preez et al. (2019) remained focused on mapping routes and describing operational characteristics such as route distances, passenger turnovers, and operating speeds – which were then used to develop a typology of informal routes in the city. None of these studies have attempted to use GPS data to identify vehicle volumes that could be useful from a transport planning perspective. Doing so is the focus of the remainder of this paper.

## 3. Case study area

The City of Tshwane (CoT) Metropolitan Municipality governs a large, urbanised region encompassing multiple centres with close economic linkages. These include the Pretoria CBD, which forms the capital core and the main location of employment in the area (Fig. 1), and nine other urban core areas scattered across the region. Radial road and rail linkages connect to outlying residential areas such as Hammanskraal, Soshanguve and Mamelodi, where most of the historically disadvantaged population resides at medium densities. This spatial arrangement creates a large tidal demand for transport over long distances, served by a variety of private and public transport.

A third of all trips are made by car and almost 30 % on foot (City of Tshwane, 2015). Of the remainder, informal minibus taxis are the most important, operating both urban and intercity services and accounting for 22 % of all trips. Formal bus and train services capture less than 10 % of trips. Spatially, Soshanguve in the Northwest region and Mamelodi in the East account for the highest percentage of minibus taxi trips (29.5 % and 25.7 % respectively). While these mode share figures are from 2015 and may not reflect the current post-COVID travel patterns – particularly regarding rail usage – they are relevant to this study as they are contemporaneous with the GPS data collected in 2013–2014.

## 4. Data acquisition

### 4.1. Minibus taxi data

Minibus taxi GPS data was acquired from iSAHA, a company that specialises in transportation data collection and analysis in South Africa. The company conducted on-board surveys on minibus taxis in and around Tshwane using electronic in-vehicle equipment and backroom surveiling and processing (iSAHA, 2015). The main aim of the surveys, conducted in 2013 and 2014, was to inform the implementation of the first two lines of CoT's Bus Rapid Transit system (A Re Yeng). A randomly selected 5 % sample of the 4000 licensed minibus taxis operating within the CoT's network was tracked over a period of 7 days per minibus taxi. The survey can be summarised as follows:

- Total kilometres travelled by all vehicles in the survey: 177203 km
- Number of associations surveyed: 19
- Number of vehicles surveyed: 207
- Number of trips made by all vehicles in the survey: 8997
- Number of stops made by all vehicles in the survey: 72236
- Number of passengers transported throughout the survey: 84539

The dataset is valuable for transport planning purposes as it is significantly larger and more comprehensive than most previous datasets. By tracking individual vehicles across multiple days, the dataset provides the requisite scale and detail to analyse variations across both space and time.

The original GPS survey conducted by iSAHA in 2015 required resources like on-board equipment and post-processing support – amounting to substantial costs (van Zyl and Labuschagne, 2008) –

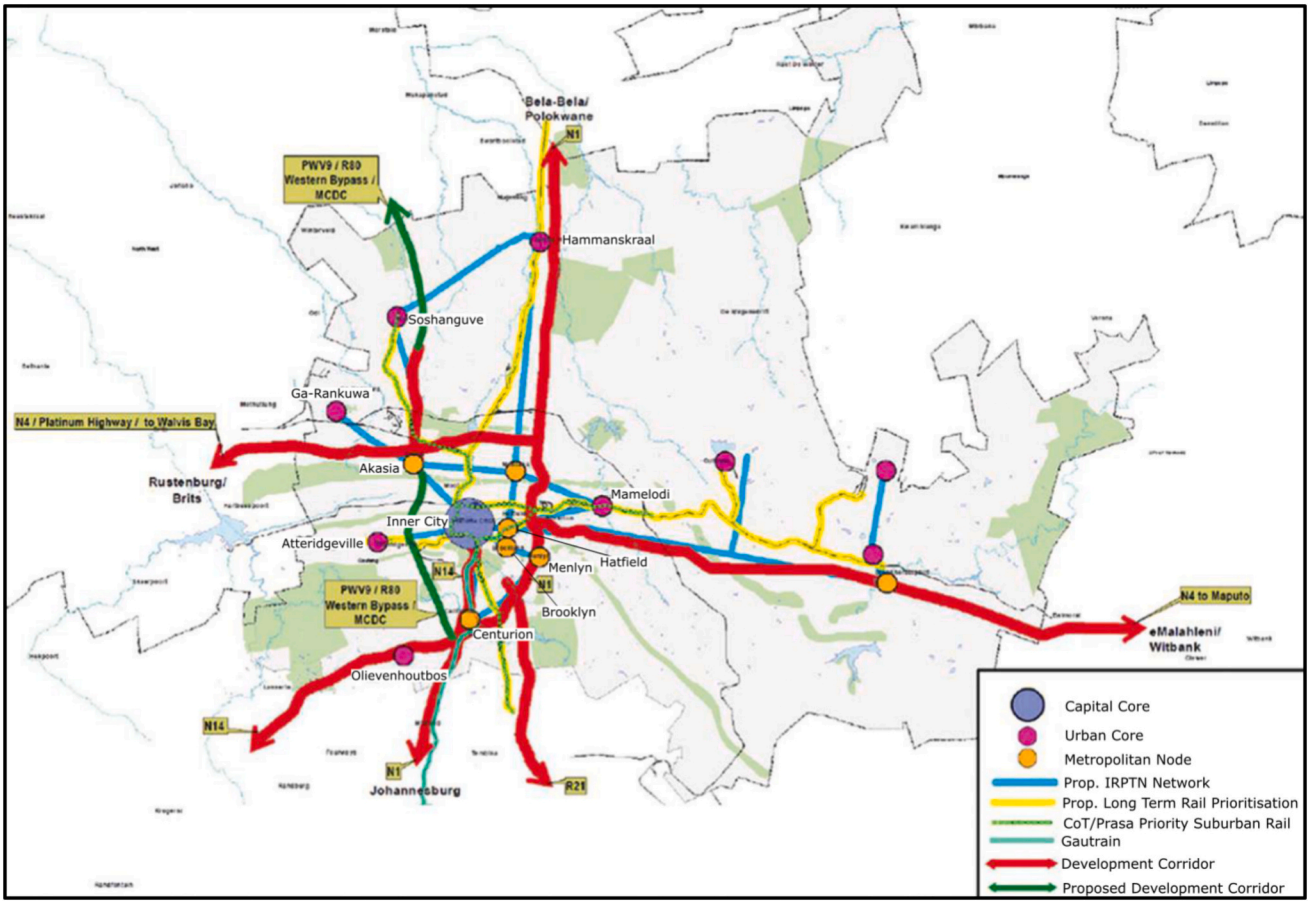



**Fig. 1.** City of Tshwane spatial development concept (City of Tshwane, 2015).

although the exact costs are not available. This may limit the routine replication by local governments in developing countries. The increasing accessibility of smartphone-based tracking offers a more cost-effective alternative, allowing such methodologies to become more scalable.

A limitation was that the data was collected 8 years ago and could be considered outdated. However, for the present purposes of developing and testing a methodology for quantifying taxi service, the data was adequate.

The minibus taxi GPS data were received in pre-processed vector format (shapefiles) and included both trip data, represented as line features, and stop data, represented as point features. Each stop point represents a position in space and time where a passenger or several passengers boarded or alighted the vehicle. This is not typical, as GPS data are more commonly provided in raw, point-by-point formats. This GIS pre-processing influenced the methodological approach, as it necessitated an adaptation from working with raw GPS data to integrating GIS shapefiles for trip analysis. However, once the shapefiles have been approximated by point features (as described below), the rest of the method we present is also applicable to raw GPS data.

The stopping and trip locations were mainly focused in and around Tshwane but also included a small number of intercity trips to neighbouring towns. These were removed, to narrow the focus to intra-metropolitan services. Fig. 2 shows the stopping and trip locations made in Pretoria and the surrounding areas. As is typical of informal services, routes and stops are widely dispersed across the area, although frequencies could vary considerably across time and space.

The attribute data for the stops- and trip shapefiles contained information on each stop or trip. The shapefiles already differentiated between the stops and trips, therefore a stay and move identification process was not required (Asakura et al., 2014). An example of the stop data for a single stopping point for a minibus taxi is shown in Table 1.

Each taxi trip consisted of a spatially referenced line feature indicating the route travelled from the origin (start point) to destination (end point) location. An example of the data of a single minibus taxi trip made is shown in Table 2. The trip data variables are similar to the stop data variables but include additional fields such as the trip start and end times, the duration and distance travelled, as well as the fare each passenger is charged, and the income made per trip. Only start point and end point latitude and longitude coordinates were provided for the start and end positions of the trip.

**Table 1**
Single stopping location data for one minibus taxi.

| Variable | Variable code name | Example |
|---|---|---|
| Object Identification | OBJECTID | 32 |
| Stop Identification | Stop_ID | G0012014080603032 |
| Trip Identification | Trip_ID | G0012014080603 |
| Minibus Taxi Code | TaxiCode | G001 |
| Date | Date | 06-Aug-14 |
| Stop Sequence | Stop_Seque | 32 |
| Trip Sequence | Trip_Seque | 3 |
| Stop Time | Stop_Time | 07:25:13 |
| Number of Passengers Boarding | On | 3 |
| Number of Passengers Alighting | Off | 2 |
| Number of Passengers on Board | Pass_on_bo | 15 |
| Minibus Taxi Function | Function | Transfer |
| Minibus Taxi Association | Associatio | BAZAAR |
| Geographical Area | Area | 6 |
| Weekday | weekDay | 2 |
| Trip Day | tripDay | G00120140806 |

### 4.2. City of Tshwane road data

GIS data of the road network for the City of Tshwane was sourced from the CoT Metropolitan Municipality. The dataset consists of arcs, nodes and vertices. Arcs are defined as line (or road) segments that are connected at intersections by nodes. Vertices define the contour or shape of arcs (Lloyd, 2010).

## 5. Data inspection and limitations

### 5.1. Data inspection

The data were inspected to ensure that it was accurate, complete, and suitable for its intended purpose.

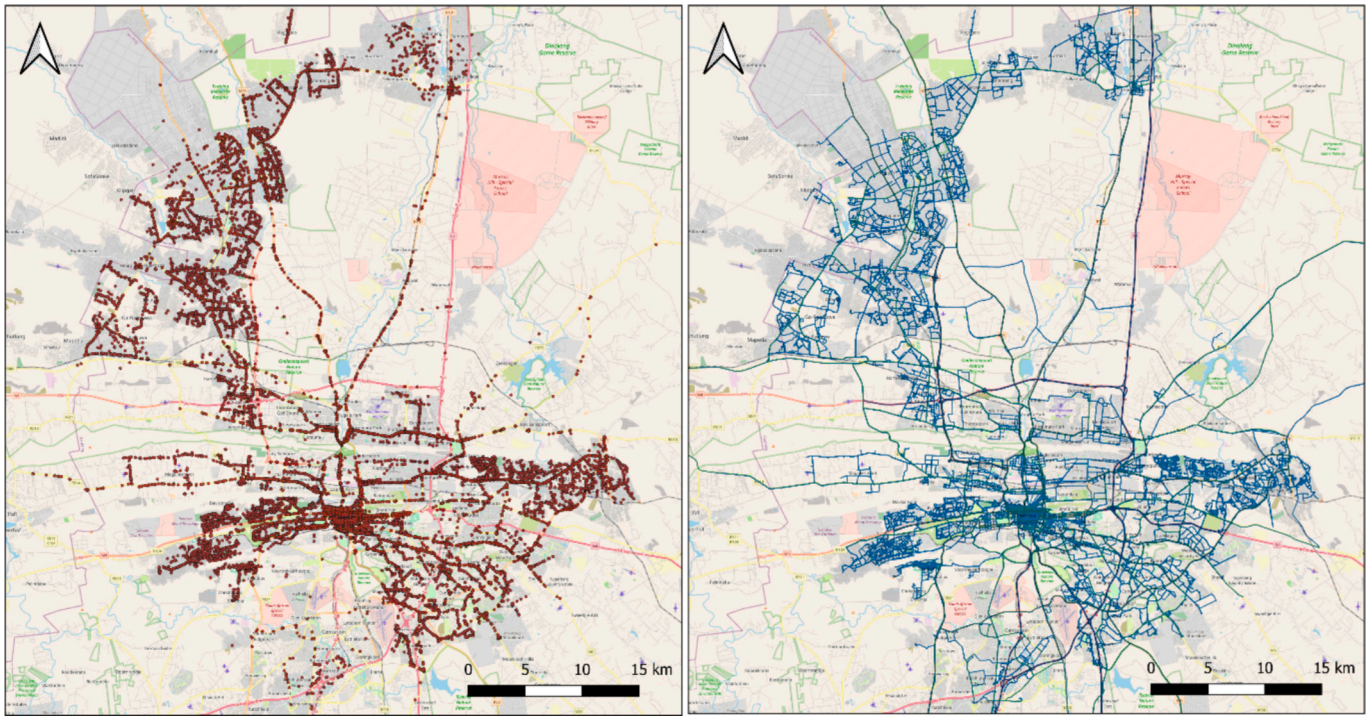



**Fig. 2.** All minibus taxi (a) stop and (b) trip locations recorded in Pretoria and surrounding areas excluding long-distance trips.

**Table 2**
Single trip data description for one minibus taxi.

| Variable | Variable code name | Example |
|---|---|---|
| Object Identification | OBJECTID | 282 |
| Trip Identification | Trip_ID | G0012014080603 |
| Minibus Taxi Code | TaxiCode | G001 |
| Date | Date | 06-Aug-14 |
| Weekday | Weekday | Wednesday |
| Trip Sequence | Trip_Seque | 3 |
| Trip Start Time | Trip_Start | 06:52:25 |
| Trip End Time | Trip_End_T | 08:12:20 |
| Trip Duration | Trip_Durat | 79.92 |
| Total Distance | Total_Dist | 40.945353 km |
| Trip Fare | Fare | 16 ZAR |
| Trip Income | Trip_Incom | 256 ZAR |
| Total on Board Trip Sequence | Total_On_S | 18 |
| Trip Sequence | Trip_Seq | 3 |
| X Coordinate Start Point | Start_X | 28.12802167 |
| Y Coordinate Start Point | Start_Y | −25.46852667 |
| X Coordinate End Point | End_X | 28.193725 |
| Y Coordinate End Point | End_Y | −25.74003667 |
| Length of Shape | Shape_Leng | 0.409453535 |
| Minibus Taxi Association | Associatio | BAZAAR |

We also examined the spatial and temporal distribution of the GPS data. Comparing the raw minibus taxi trip data (Fig. 2) with the land use patterns of the CoT (Fig. 1), it appears that minibus services are mostly concentrated in areas to the north and east of the CBD. This corresponds to the areas with highest population, and with highest minibus-taxi use. Less dense coverage is provided toward the west and south of the city.

The number of minibus taxi trips made per hour of the day is illustrated in Fig. 3. The trip profiles are consistent with the operating patterns displayed by minibus taxis in South African urban areas, i.e. there are pronounced morning and afternoon peak periods but a reduced number of trips in the inter-peak period, resulting in a supply peak-to-base ratio of 2.3. The horizontal distance between the two lines reflects the general timing pattern of the trips. When the duration of a trip is less than an hour the lines overlap, as trip start- and end times are plotted in hourly bins. For trip durations longer than an hour the horizontal distance increases. For the morning and afternoon peak periods the average trip duration is longer than an hour whereas during the late morning and early afternoon, average trip durations are less than an hour. This is consistent with general congestion patterns observed within the city (City of Tshwane, 2015).

### 5.2. Sample data limitations

In order to determine the number of minibus taxi trips on each road segment within the city, a key step is to allocate GPS data points to the correct arc representing that segment in the GIS dataset. During the data exploration process, some issues with the trip data were identified which could influence the method of calculation as well as the overall results. These included:

1. Some GPS tracks are not aligned with any underlying GIS street centerlines, because those roads do not exist in the database. This is the case where taxis traverse open lots to connect between roads, or use unnamed informal roads. Giliomee et al. (2023) also discuss this issue in the context of simulating taxi movements, concluding that it is a difficult to solve but rare occurrence. Such events amounted to approximately 1 % of all GPS routes in our sample. Visual inspection showed that these cases are limited to township areas and informal settlements, at invariably low taxi frequencies. They were manually removed from the database as the study was limited to the formal road network.
2. Of the remaining GPS-tracked routes, many data points did not align perfectly with the underlying GIS street centerlines due to inaccurate GPS readings in the original source data which could be attributable to high-rise buildings preventing the GPS devices from making accurate recordings (Merry and Bettinger, 2019). This was especially prevalent in the Pretoria CBD (as illustrated in Fig. 4). This issue had to be addressed such that the outcome reflected the minibus taxi trip count on the road network as accurately as possible.
3. Irregular vertices (or spikes) were observed in the trip data indicating unrealistic deviations or inconsistencies on the route. These spikes are larger than the GPS drift described above and could cross multiple road segments in some instances. This could have been caused by the operator switching off the GPS device and turning it on later in the trip, or by random errors in GPS coordinates caused by satellite errors. In this case the latter is likely the cause. A single trip (shown in red) was isolated in Fig. 5 to illustrate the inaccuracy of the route that was recorded. Whilst the majority of the trip fell on the road network, a spike was observed as the vehicle approached the

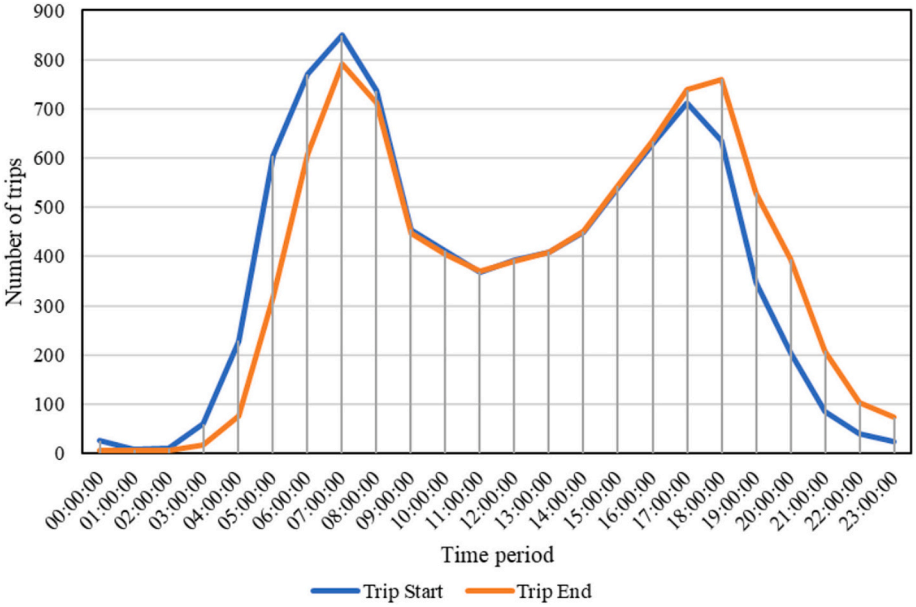



**Fig. 3.** Number of minibus taxi trips recorded per hour.

**Fig. 4.** Example of GPS-tracked routes not aligning with GIS street centrelines.

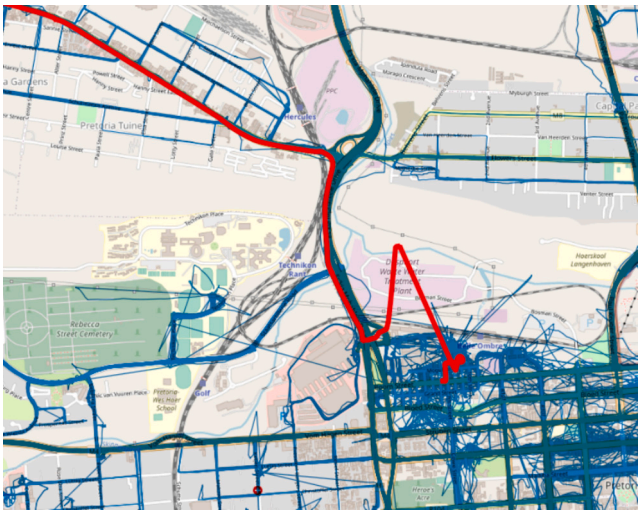


**Fig. 5.** Example of GPS error in the data.

CBD. It would therefore be necessary to either relocate that portion of the trip to the correct location or exclude it from the data

4. The trip count could not be estimated per direction. This was due to the GIS road layer not having separate road segments per travel direction. Even if the road network had separate segments per direction, the segments would lie too close to each other and the vehicle trip lines would overlap both directions, causing inaccuracy in the directionality. This would require manually snapping the trip to the correct road segment which would be a time-consuming process. The trip count was therefore estimated as the sum of the counts for both directions, which proved to be adequate for the purpose of this research

The shortcomings of the data were corrected and dealt with at the same stage as when the trip count per street segment was calculated. The steps taken to correct the errors in the data as well as determining the trip count per street segment are discussed in the following section.

## 6. Calculating trip count per street segment

This section explains the process that was followed to derive the number of taxi trips per road segment. Based on data limitations discussed in Section 5, the data cleaning and trip count calculation process was divided into the following steps:

1. Convert taxi trip lines to points,
2. Assign points to the nearest road segment,
3. Remove noise and determine trip counts.

### 6.1. Step 1: Convert taxi trip lines to points

To link the taxi trips to segments of the CoT road network, each trip (consisting of a single line) was split, or converted to point features. There were two reasons for converting the lines to multiple points:

1. To allow all the individual road segments that were traversed during a single trip to be identified, and
2. To allow for the subsequent removal of portions of the trip data that did not align with the road network data or where irregular vertices (or spikes) were observed.

The *Points Along Geometry* function in QGIS, which includes a parameter to set the distance between points, was used to split the trip lines into multiple points. The function created points at regular intervals along the line geometries. The derived points would still retain data such as the Trip_ID from the original line feature. Additional attributes were also added such as the distance along the geometry and the angle of the line at the point.

The appropriate distance between points when converting taxi trip lines had to be determined. If the distance was too long, relevant road segments that formed part of a taxi trip could be missed or skipped since there would be no associated point in close proximity. This would have resulted in incorrect trip counts, especially where minibus taxis use short road segments such as in the CBD. If the distance between points was too short, the running time for the programme became too long.

To address the first problem, it was decided to use the minimum link length in the database as a guide. The shortest link on the road network was 20 m, so this was used as the distance between points when generating point features from lines. Fig. 6 illustrates a minibus taxi trip split into 20 m segments and the segments being reduced to points.

If, however, the process begins with GPS points rather than GIS lines, Step 1 can be skipped, although the points will not be 20 m apart, as their spacing is determined by GPS frequency and the vehicle's speed.

### 6.2. Step 2: Assign points to the nearest road segment

In order to associate the derived taxi trip point features with the correct road segment, the distance from each point to their nearest road segment was calculated, in order to snap the point to that road. The *Join attribute by nearest* function was used for this, as it links the attributes of one feature set in a spatial vector dataset, in this case the points that

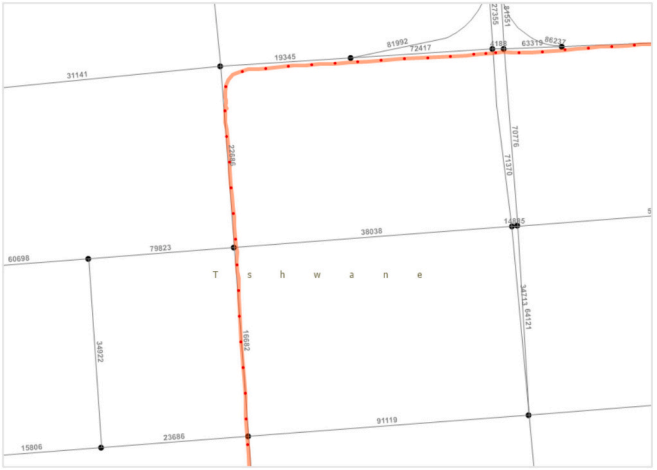



**Fig. 6.** Minibus taxi trip split into 20 m segments.

were generated, to the attributes of the closest feature in a second set, the road segments.

### 6.3. Step 3: Remove incorrect points and determine trip counts

The next step was to remove portions of the trip data that did not align with the road network data or where irregular vertices (or spikes) were observed. A point was regarded as incorrect either:

When only one point from a single trip was located on a street segment (illustrated with the blue circle in Fig. 7) – this was viewed as a spike since the distance between points was short enough that each road segment would have at least two points, or When a point was too far from the street segment (illustrated with the red circle in Fig. 7). A trial-and-error approach was used to determine the optimal distance and a suitable length of 20 m was chosen as it was determined that this was the shortest link length on the road network.

Microsoft Access was used as a database to store the derived taxi trip points in multiple tables. Although QGIS provides a built-in database manager where the user can create, edit, and delete tables, an external database proved easier and quicker to use. The queries were used to sum all the unique points on a road segment and discard the incorrect points. Unique points were then converted to trip counts and sorted by day of the week. Once the queries were completed to address the problems, the subsequent tables could then be exported to QGIS from which the final maps were produced.

## 7. Minibus taxi trip count outputs

This section discusses the trip count per street segment outputs generated from the minibus taxi trip data.

### 7.1. Bandwidth plots of taxi counts

The basic output from the method described above is a GIS vector file of road segments, with the number of taxi trips counted appended as an attribute to each segment. This can be filtered to show the trips made per street segment per time of day or per day of the week. The data can also be used to illustrate taxi volumes through a bandwidth plot, with the width of lines plotted proportional to the quantum of trip counts.

Fig. 8 illustrates the minibus taxi trip count classification across the entire CoT. It should be noted that these trip counts are from all the minibus taxi trips that were recorded over the duration of the project, that is, over a rolling seven-day period, and do not correspond to daily volumes. Jenks classification was used to define trip count intervals as it splits up data by grouping similar values that "minimise differences between data values in the same class and maximise the differences between classes" (Slocum, 2009; ArcGIS Pro, 2023). This method is best used with data that is unevenly distributed but not skewed toward either end of the distribution.

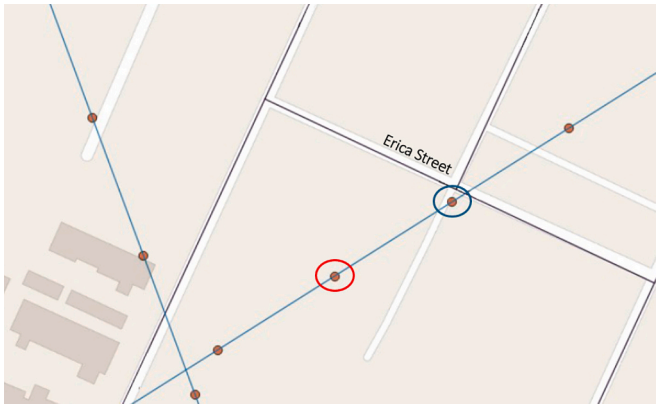



**Fig. 7.** Single points lying on a road segment.

A few observations are pertinent: two long minibus taxi corridors with a trip count between 446 and 787 were observed, one connecting Soshanguve in the North to Pretoria's CBD and the other, in the East, connecting Mamelodi to the suburbs in Pretoria East. These corridors mirror the development corridors shown in Fig. 1. Secondly, the highest trip counts (in the band between 787 and 1212 vehicle-trips) are concentrated in Pretoria's CBD, or the capital core (within the green square).

Fig. 9 shows the taxi trip count classification within the CBD in more detail. The high level of taxi activity in the CBD is evident, consistent with the high concentration of jobs, shopping malls and smaller shops in the area. Taxi activity is concentrated more in the north of the area where two large minibus taxi ranks are located, and many operators hold during less busy parts of the day. The route volumes also match the passenger stop locations (which were not used in generating counts). Looking at the distribution of stops in the figure, it appears that the boarding and/or alighting locations of passengers are spread-out throughout the CBD. This concurs with results obtained by Ankunda and Venter (2025) who found that minibus taxi transfers were spatially efficient resulting in short walking and waiting times, as operators tend to oblige passengers' requests for stops close to their destinations.

### 7.2. Output reasonableness check

To determine the effect of the data cleaning process on the accuracy of the results a comparison was made between the trip count output and the raw data. As the data cleaning removed some points deemed to be erroneous, one can expect the trip count results to be slightly lower than the true number of tracked trips per road segment. For each interval in the Jenks classification, two random road segments were selected inside and outside the CBD, and the trip count was compared to the raw trip data extracted from the trip shapefile. The raw trip data were manually examined such that the number of trips could be counted. The comparison for the resulting 15 locations (there were only seven count intervals outside the CBD as the highest interval was not observed) are shown in Fig. 10a and Fig. 10b respectively. Within the CBD, accuracy varied between 84 % and 100 %; in all cases there was an underestimation of the trip counts. The average accuracy of 89 % indicates an average error in trip counts of 11 %, owing to the CBD being the area with the greatest amount of noise and spiked vertices. Outside the CBD, the error was smaller, with accuracy varying between 92 % and 104 % with an average of 99 %. Whether this level of error is acceptable depends on how the results are to be used. There are no strict validation criteria available, like the GEH statistic often used to compare modelled and observed traffic counts in traffic models (Kabashkin et al., 2018). For the purpose described below – the identification of major travel corridors – the error is very unlikely to significantly affect the outcome.

### 7.3. Temporal variations

The minibus taxis trips were segmented by day of the week in 24-h periods to determine any significant patterns in the flow over the duration of a week. Figs. 11–14 illustrate the minibus taxi trip counts from Monday to Sunday across the region. Overall, Monday saw the highest number of minibus taxi trips and Sunday the lowest. The two main corridors already identified remained the two busiest corridors throughout the week. Interestingly, Thursday saw a significant decrease in the number of trips made (approximately 10 % lower than the other weekdays). The reason for this is that fewer vehicles were tracked on this day but the spatial distribution of trips between corridors remain comparable across the days of the week.

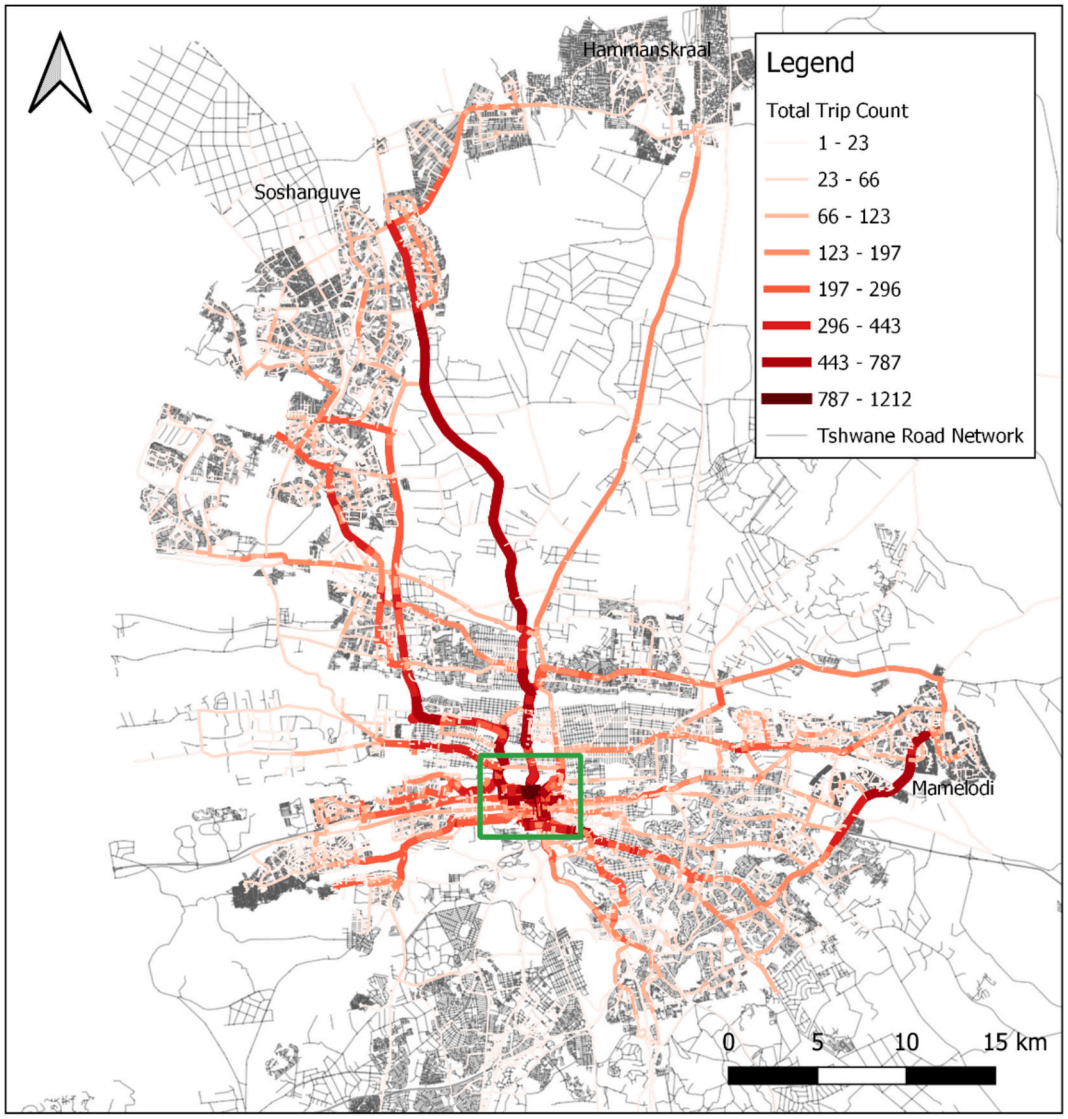



**Fig. 8.** Minibus taxi trip raw count classification for the Tshwane Metropolitan Municipality.

## 8. Application of trip volumes

### 8.1. Turning counts into volumes

The total trip counts recorded for all the sampled taxis were first categorised and averaged to reflect the trip count for each day of the week. Since a 5 % sample size was used to track the minibus taxis, the number of trips were scaled up by a factor of 20 in order to estimate the total trip volume across each corridor more accurately and to reflect the full scope of the minibus taxi activity in the network.

### 8.2. Identification of major taxi corridors

From the output maps the eight most significant minibus taxi corridors could be identified. Such information provides ridership evidence of their strategic importance in the network and can be useful to transport planners for prioritising infrastructure investments, focusing on areas that would benefit most from improvements such as dedicated lanes or public transport stations. We selected corridors based on their observed daily trip volumes. A corridor was included if it showed a high trip volume, represented by the dark red band, indicating between 1520 and 2440 trips per day. Three such corridors met this criterion. Additional corridors were then selected using the orange and yellow bands, which represent daily trip volumes between 940 and 1520 and 580–940, respectively. The corridors were once again plotted (Fig. 15) using a Jenks natural breaks classification to aid with visual interpretation. Table 3 details the start and end locations for the selected minibus taxi corridors as well as the daily trip volume identified along the corridor.

All corridors except for Corridors 2 and 8 end in the CBD and begin in the outskirts of the city. It once again confirms the importance of the CBD as a major trip destination, and the arguments in favour of decent public transport investment in this node. Corridor 2 transports passengers from the eastern township of Mamelodi to Silver Lakes and other such affluent estates and ending in Garsfontein. In terms of volumes, Corridor 2 is the primary corridor as it serves a critical link between a densely populated, economically disadvantaged area and affluent suburbs with few other public transport options available.

## 9. Conclusions

This paper presents a novel methodology for quantifying trip volumes for informal public transport from GPS data, providing new insights into network patterns and high-demand areas. The method is illustrated using GPS data for minibus taxi services from the City of Tshwane in South Africa. We provide detail on how to deal with data cleaning issues that typically emerge when dealing with public transport

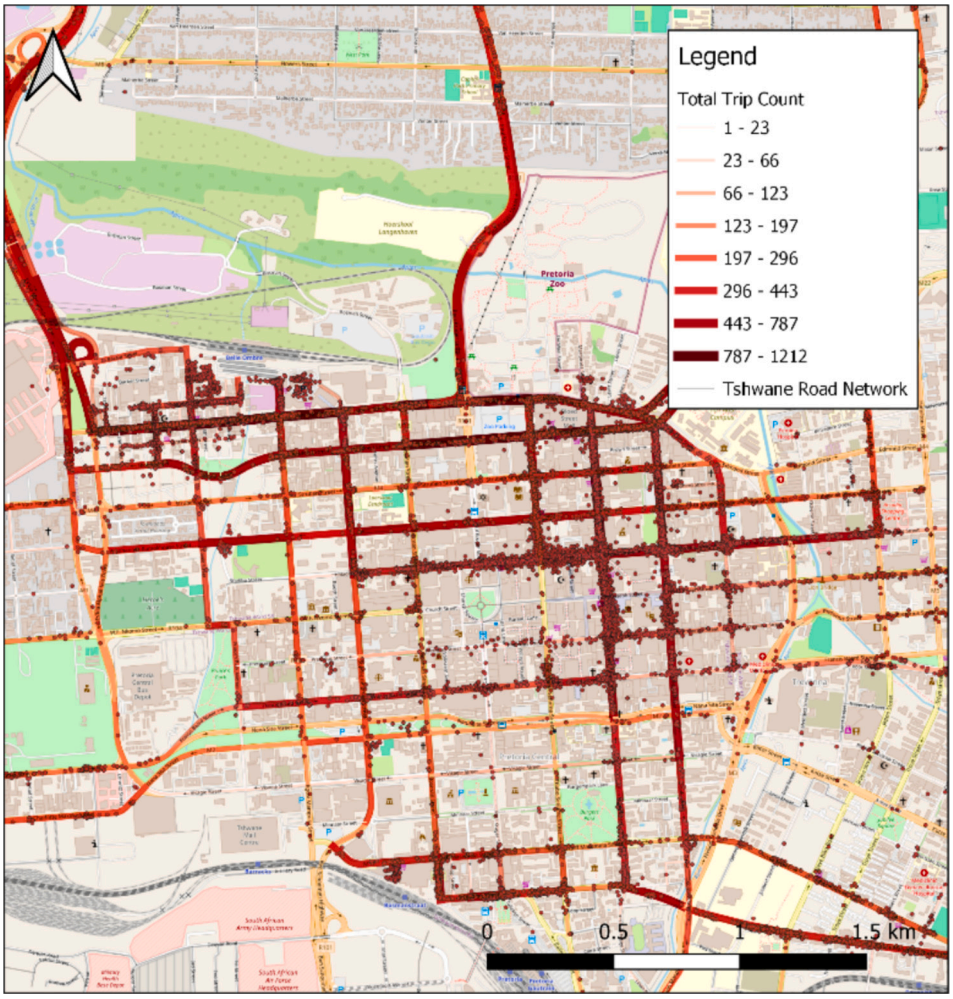



**Fig. 9.** Minibus taxi raw trip count classification for the Pretoria CBD overlaid with stop locations.

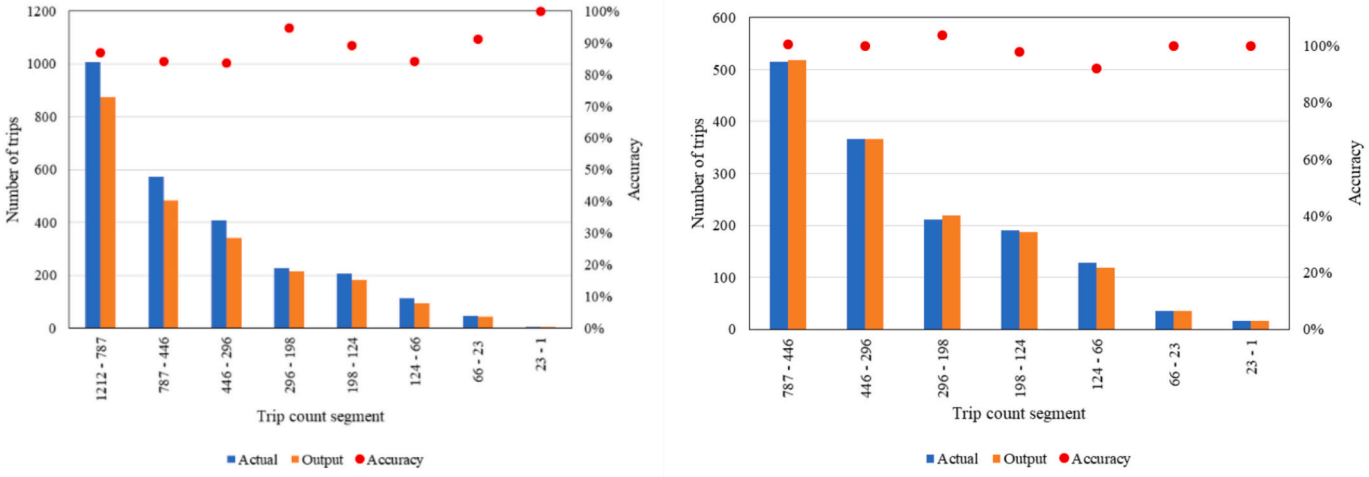



**Fig. 10.** Trip count segment comparison (a) CBD and (b) outside CBD.

GPS data, especially when the data are provided in GIS shapefile format rather than raw GPS points. A reasonableness check of the results confirmed spatial and temporal variations that match operational trends, such as peak activity periods during morning and evening commutes and reduced activity in peripheral areas outside core demand zones. Variations across space revealed that trip counts were significantly higher in the CBD and major corridors compared to suburban areas, reflecting the concentrated demand for mobility in economic hubs. Key limitations in the data, including the use of informal roads, GPS-related errors (spiked vertices), and directional inconsistencies, were identified. To improve accuracy, the methodology suggests techniques to filter noise and spiked vertices. The data cleaning

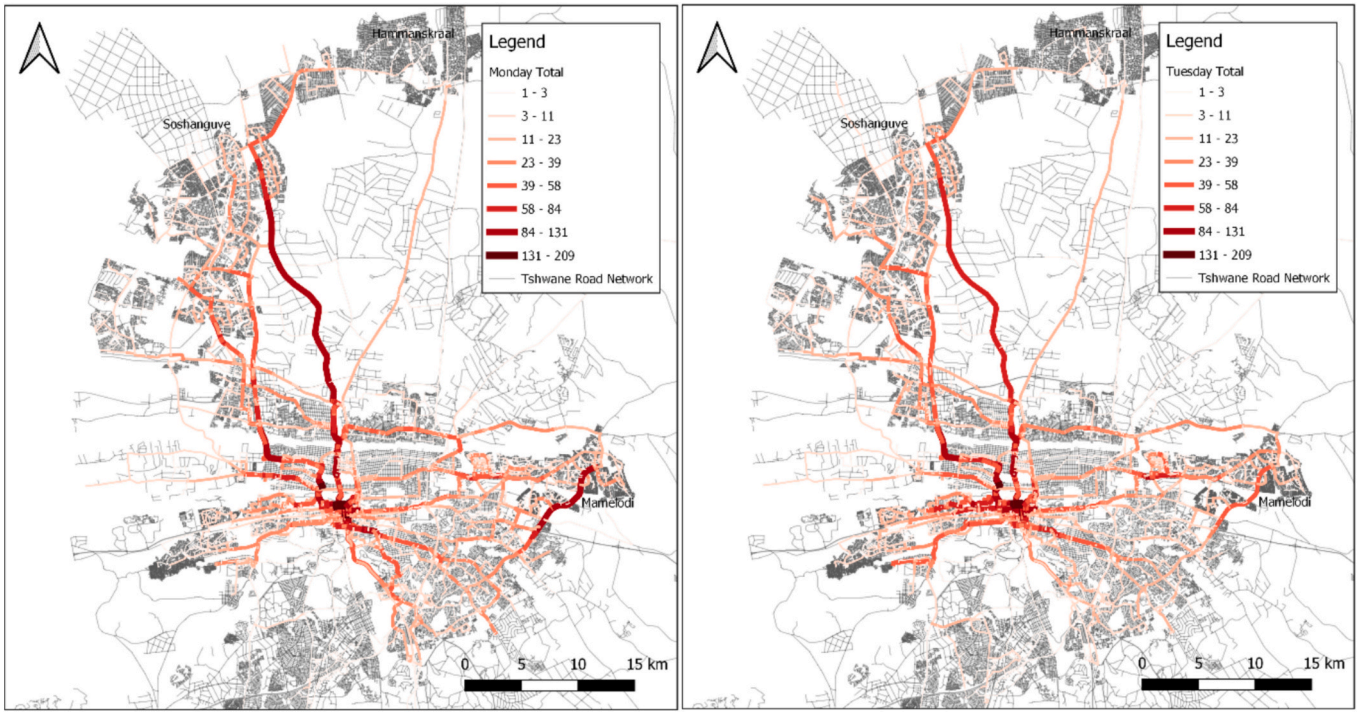



**Fig. 11.** Minibus taxi raw trip counts by day of the week (Monday and Tuesday) – City of Tshwane.

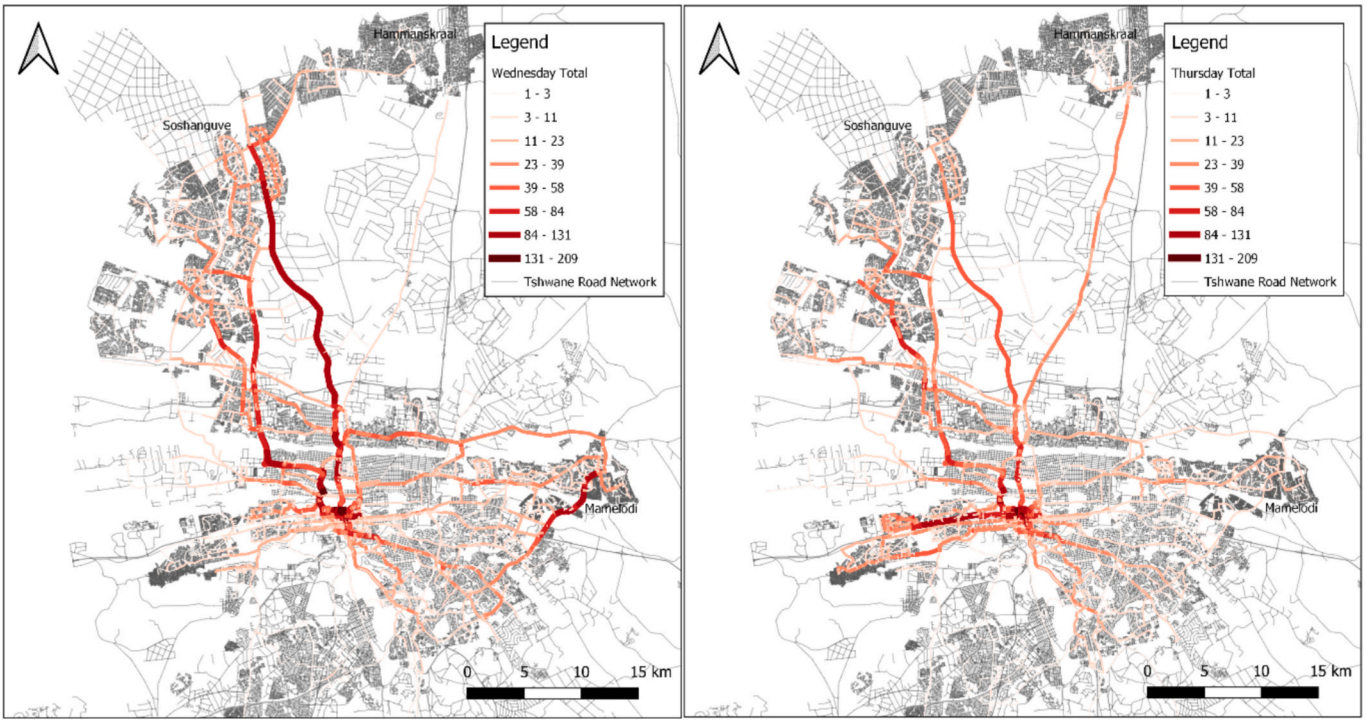



**Fig. 12.** Minibus taxi raw trip counts by day of the week (Wednesday and Thursday) – City of Tshwane.

method introduced errors of less than about 10 % when deriving vehicle volumes on road segments. Further work is needed to refine the cleaning process, especially on high-volume roads where the errors are higher. Work is also needed to consider the directionality of trips when determining volumes per road segment, as this was not included at this stage. Directional volumes are important for transport planning as they indicate peaking across the day and vary much more significantly than overall volumes.

Although this study focused on vehicle trips, the dataset includes some passenger boarding and alighting information that could support estimation of passenger flows in future work. Due to limitations in the spatial granularity of the data, we did not attempt to interpolate passenger turnover at the street-segment level.

The methodology shows how volumes are derived from trip counts per street segment, producing actionable outputs with practical applications in transport planning and traffic engineering. Such results are

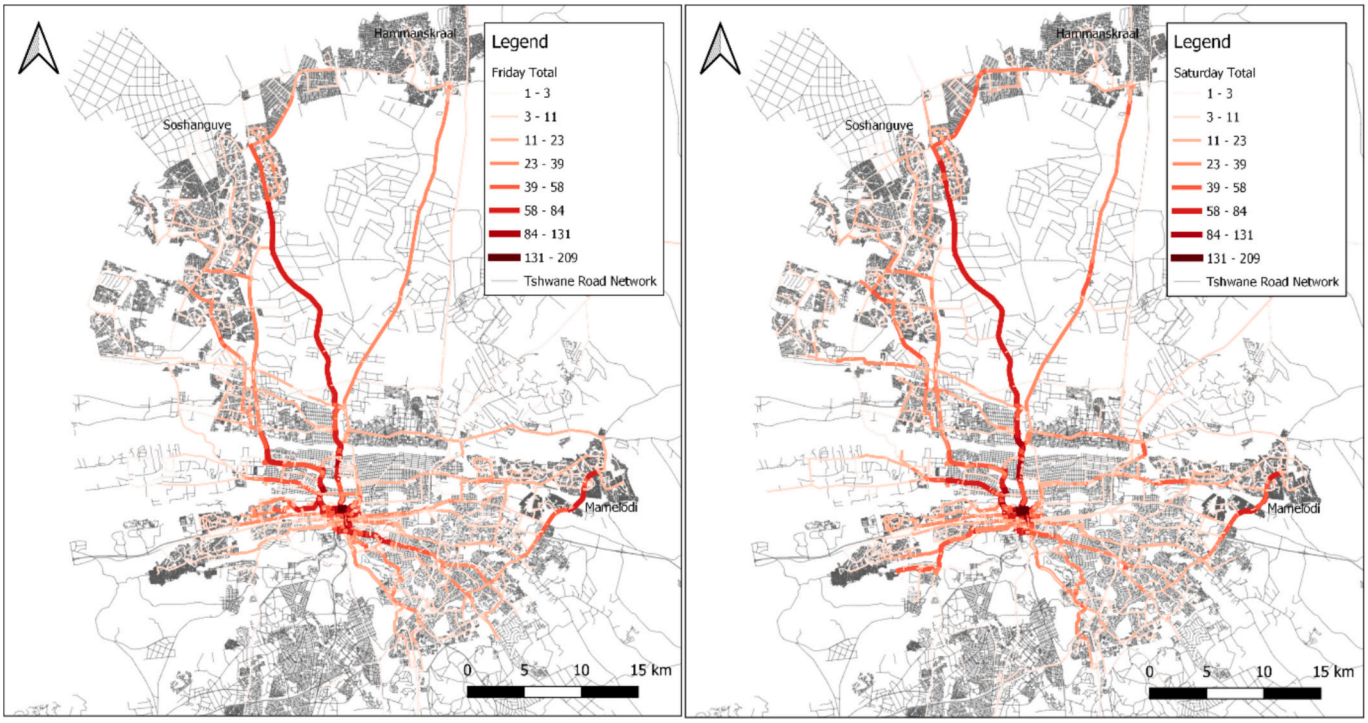



**Fig. 13.** Minibus taxi raw trip counts by day of the week (Friday and Saturday) – City of Tshwane.

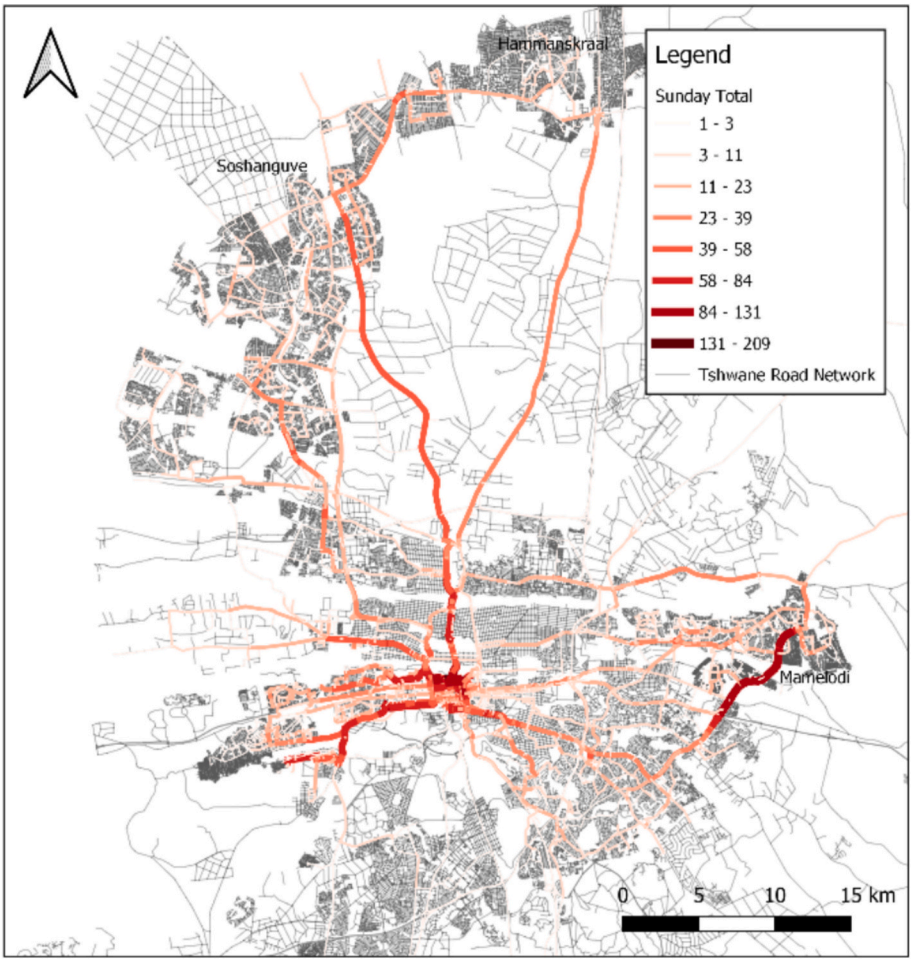



**Fig. 14.** Minibus taxi raw trip counts by day of the week (Sunday) – City of Tshwane.

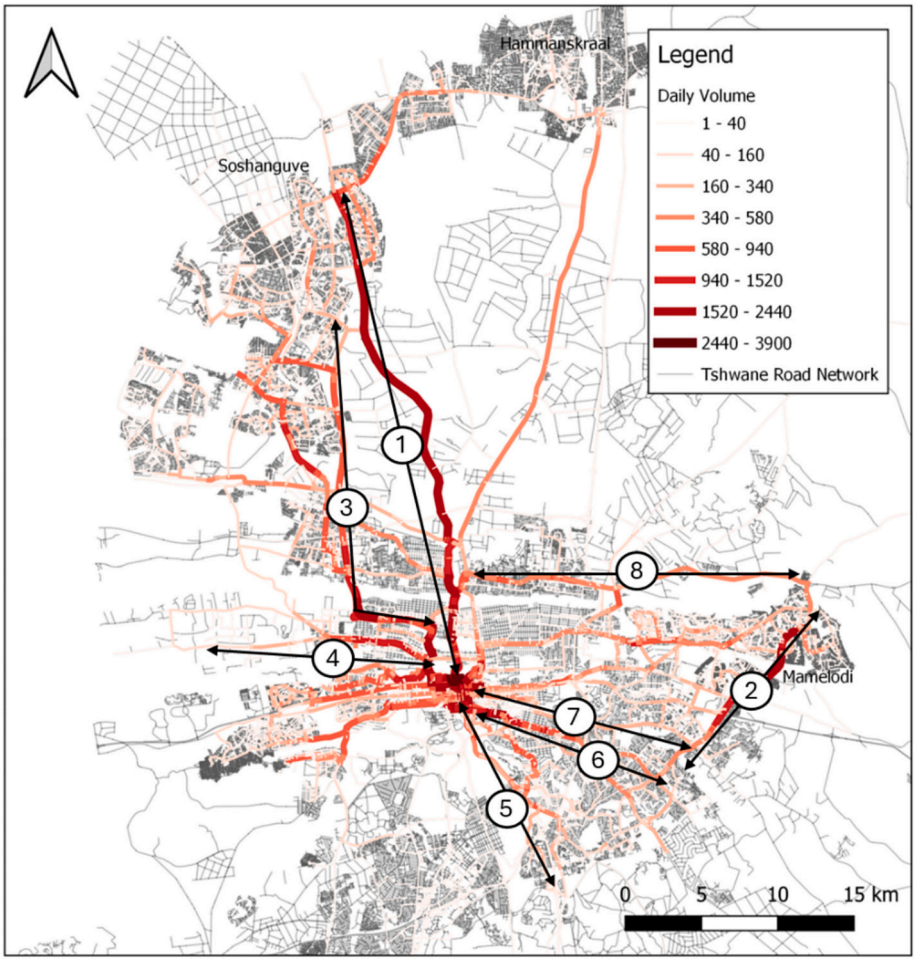



**Fig. 15.** Major minibus taxi corridors, average daily volumes shown in vehicles/day.

**Table 3**
Start and end locations for selected minibus taxi corridors with corresponding daily volumes.

| | Start location | End location | Average daily volume (veh/day) |
|---|---|---|---|
| Corridor 1 | Klippan | Pretoria CBD | 1840 |
| Corridor 2 | Mamelodi | Garsfontein | 2300 |
| Corridor 3 | Soshanguve | Pretoria CBD | 1780 |
| Corridor 4 | Magaliesmoot | Pretoria CBD | 1080 |
| Corridor 5 | Rietvalleipark | Pretoria CBD | 1060 |
| Corridor 6 | Garsfontein | Pretoria CBD | 1080 |
| Corridor 7 | Wapadrand | Pretoria CBD | 540 |
| Corridor 8 | Mamelodi | Wonderboom | 820 |

not available with typical current approaches for processing GPS data for informal public transport, as they typically do not include frequency information on maps. These insights could enable the identification of:

1) High-volume corridors: High-traffic corridors are identified through GPS data, validating the need for infrastructure upgrades to support efficient mobility as well as prioritising transport interventions and infrastructure investments along critical routes.
2) Key intersections: Frequently used intersections along minibus taxi routes are pinpointed as priority areas for improvements, offering benefits for both public transport operators and users.
3) Main public transport nodes: Areas of convergence of multiple high-volume routes provide suitable locations for place-based interventions like terminals, ranks, and walkability improvements.

## CRediT authorship contribution statement

**Lourens de Beer:** Investigation, Conceptualization, Data curation, Methodology, Formal analysis, Writing – original draft. **Christo Venter:** Supervision, Funding acquisition, Writing – review & editing, Project administration. **Lourens Snyman:** Visualization, Methodology, Software.

## Acknowledgements

The authors would like to thank the Volvo Research and Educational Foundation (VREF) via the BRT Centre of Excellence for funding and support. The City of Tshwane Metropolitan Municipality and iSAHA are also gratefully acknowledged for the provision of data for this research.

## Data availability

The authors do not have permission to share data.

## References

Ankunda, G., Venter, C., 2025. Studying transfers in informal transport networks using volunteered GPS data. Travel Behav. Soc. 39.
ArcGIS Pro, 2023. Data Classification Methods. Retrieved 8 May, 2023, from. https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/data-classification-methods.htm.
Asakura, Y., Hato, E., Maruyama, T., 2014. Behavioural data collection using mobile phones. In: Advances in Data Mining and Database Management, pp. 17–35.
Booysen, M.J., Ebot Eno Akpa, N.A., 2014. Minibus driving behaviour on the Cape Town to Mthatha route. In: Paper Presented at the Southern African Transport Conference, Pretoria, South Africa.

Bulbulia, T., 2023. Cape Town's Blue Dot Pilot Offers Lessons, Demonstrates Successes for Minibus Taxi Industry. Engineering News. Retrieved 19 October, 2023, from. https://www.engineeringnews.co.za/article/cape-towns-blue-dot-pilot-offers-lessons-demonstrates-successes-for-minibus-taxi-industry-2023-07-13.

Cervero, R., Golub, A., 2007. Informal transport: a global perspective. Transp. Policy 14 (6), 445–457.

City of Tshwane, 2015. Comprehensive Integrated Transport Plan, Pretoria, South Africa: City of Tshwane.

Coetzee, J., Krogscheepers, C., Spotten, J., 2018a. Mapping minibus-taxi operations at a metropolitan scale - methodologies for unprecedented data collection using a smartphone application and data management techniques. In: Paper Presented at the *Southern African Transport Conference*, Pretoria, South Africa.

Coetzee, J., Mulla, A., Oosthuizen, N., 2018b. Tools to assist in determining business values of individual minibus-taxi operations in Rustenburg, Northwest, South Africa. In: Paper Presented at the Southern African Transport Conference, Pretoria, South Africa.

du Preez, D., Zuidgeest, M., Behrens, R., 2019. A quantitative clustering analysis of paratransit route typology and operating attributes in Cape Town. J. Transp. Geogr. 80, 102493.

Gaibe, H., 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. In: Paper Presented at the Southern African Transport Conference, Pretoria, South Africa.

Giliomee, J.H., Hull, C., Collett, K.A., McCulloch, M., Booysen, M.J., 2023. Simulating mobility to plan for electric minibus taxis in Sub-Saharan Africa's paratransit. Transp. Res. Part D: Transp. Environ. 118, 103728.

Hull, C., Giliomee, J., Collett, K.A., McCulloch, M., Booysen, M.J., 2022. Using high resolution Gps data to plan the electrification of paratransit: a case study in South Africa. SSRN Electron. J. https://doi.org/10.2139/ssrn.4149228.

iSAHA, 2015. Tshwane Electronic On-board Survey on Minibus Taxis. Retrieved 30 March, 2022, from. http://www.isaha.co.za/portfolio_page/tshwane/.

Jia, W., Beukes, E., Coetzee, J., Van Ryneveld, P., 2022. Improving Paratransit in Maseru and Gaborone; Using Innovative Data Techniques in a Diagnostic Approach to Inform Strategy. The World Bank, Washington, DC. http://hdl.handle.net/10986/37301.

Kabashkin, I., Yatskiv, I., Prentkovskis, O., 2018. Reliability and statistics in transportation and communication. In: 18th International Conference on Reliability and Statistics in Transportation and Communication (RelStat), Riga, Latvia.

Klopp, J., Williams, S., Wagacha, P.W., Ochieng, D.O., White, A., 2015. Leveraging cellphones for wayfinding and journey planning in semi-formal bus systems: Lessons from digital Matatus in Nairobi. In: Planning Support Systems and Smart Cities. Springer, pp. 227–241.

Klopp, J.M., Delattre, F., Chevre, A., 2019. Open Data for Inclusive Urban Public Transport Globally. Agence Française de Développement, Paris, France.

Kumar, A., Zimmerman, S., Arroyo-Arroyo, F., 2021. Myths and Realities of "Informal" Public Transport in Developing Countries: Approaches for Improving the Sector. The World Bank, Washington, DC. Retrieved from. https://www.ssatp.org/sites/default/files/publication/SSATP_Informal_v_final_double_compressed.pdf.

Lloyd, C., 2010. Spatial Data Analysis: An Introduction for GIS Users. Oxford University Press, Oxford, UK.

Merry, K., Bettinger, P., 2019. Smartphone GPS accuracy study in an urban environment. PLoS One 14 (7). https://doi.org/10.1371/journal.pone.0219890.

Mungadze, S., 2019. Telematics to monitor minibus taxi driver behaviour. In: ITWeb. Retrieved 19 October, 2023, from. https://www.itweb.co.za/content/6GxRKqY81JNMb3Wj.

Ndibatya, I., Coetzee, J., Booysen, M.J., 2016. Mapping the informal public transport network in Kampala with smartphones: making sense of an organically evolved chaotic system in an Emerging City in sub-Saharan Africa. In: Paper Presented at the Southern African Transport Conference, Pretoria, South Africa.

Randall, L., Brugulat-Panés, A., Woodcock, J., Ware, L.J., Pley, C., Abdool Karim, S., Micklesfield, L., Mukoma, G., Tatah, L., Dambisya, P.M., Matina, S.S., Hambleton, I., Okello, G., Assah, F., Anil, M., Kwan, H., Awinja, A.C., Pujol-Busquets Guillén, G., Foley, L., 2023. Active travel and paratransit use in African cities: mixed-method systematic review and meta-ethnography. J. Transp. Health 28, 101558.

Rix, A.J., Abraham, C.J., Booysen, M.J., 2022. Why taxi tracking trumps tracking passengers with apps in planning for the electrification of Africa's paratransit. iScience 25 (9), 104943.

Saddier, S., Johnson, A., 2018. Understanding the operational characteristics of paratransit services in Accra, Ghana: A case study. In: Paper Presented at the Southern African Transport Conference, Pretoria, South Africa.

Saddier, S., Patterson, Z., Johnson, A., Wiseman, N., 2017. Fickle or flexible? Assessing paratransit reliability with smartphones in Accra, Ghana. Transp. Res. Rec. 2650 (1), 9–17.

Slocum, T.A., 2009. Thematic Cartography and Geovisualization, 3rd ed. Pearson Prentice Hall, Upper Saddle River, NJ.

Ukam, G., Adebanji, C.A.A., Ackaah, W., 2024. Factors affecting paratransit travel time at route and segment levels. Int. J. Transp. Sci. Technol. 14, 276–288.

van Zyl, J., Labuschagne, K., 2008. Attractive methods for tracking minibus taxis in South Africa for public transport regulatory purposes. In: Paper Presented at the Southern African Transport Conference, Pretoria, South Africa.

Vergel-Tovar, C.E., Leape, J., Carrasquilla, M.V., Arana, M.C.P., Gonzalez, D.T., Rubiano, L.C., Barón, E.S., Martinez, P., 2022. Mapping the transit network of greater Cartagena with mobile phones: coverage, accessibility, and informality. J. Transp. Geogr. 105, 103484.

Zeeman, A.S., Booysen, M.J., 2014. Public transport sector driver behaviour: Measuring recklessness using speed and acceleration. In: Paper Presented at the Southern African Transport Conference, Pretoria, South Africa.

Zegras, P.C., Eros, E., Butts, K., Resor, E., Kennedy, S., Ching, A., Mamun, M., 2015. Tracing a path to knowledge? Indicative user impacts of introducing a public transport map in Dhaka, Bangladesh. Camb. J. Reg. Econ. Soc. 8 (1), 113–129.