Data in Brief 55 (2024) 110732

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

High-fidelity tracking data gathered on minibus taxis in Stellenbosch, South Africa

Christopher Hull [a], J.H. Giliomee [b], Katherine A. Collett [a], Malcolm McCulloch [a], M.J. Booysen [b,c,*]

[a] *Engineering Science Department, University of Oxford, United Kingdom*
[b] *Department of E&E Engineering, Stellenbosch University, South Africa*
[c] *Department of Industrial Engineering, Stellenbosch University, South Africa*

A R T I C L E I N F O

Article history:
Received 26 November 2023
Revised 10 June 2024
Accepted 5 July 2024
Available online 15 July 2024

Dataset link: 1Hz GPS Tracking Data (Original data)

Keywords:
Minibus taxis
Mobility
Electric mobility
Per-second kinetic data
Paratransit routes
Energy expenditure
Vehicle efficiency

A B S T R A C T

Minibus taxis, a form of informal shared mobility that carries up to 16 passengers, is the main mode of public transport in sub-Saharan Africa, and given global trends, a large-scale shift to electric paratransit is imminent in the coming decades. Modeling the energy consumption (kWh/km) of electric vehicle (EV) fleets is a pre-requisite for planning for fleet deployment, especially in energy-constrained contexts. Given the paucity of EVs in sub-Saharan Africa, ground-truth data on the energy consumption of electric paratransit does not exist for many developing contexts. Consequently, GPS tracking data on internal combustion engine (ICE) versions of these vehicles is often used to estimate the energy consumption of an electric equivalent. To date, only per-minute GPS tracking data has been captured on these vehicles and used for energy consumption estimates. But this sampling frequency is insufficient for accurate energy consumption estimates, especially given the unique micro-mobility patterns of minibus taxis that are characterized by many rapid acceleration/deceleration events in quick succession. Although simulators can be used to interpolate between the dataset, they have been shown to be inaccurate in the regional context. This article presents a dataset of high-fidelity micromobility data captured on minibus taxis in transit on four typical route types: inter-city, intra-city, uphill, and downhill. The

* Corresponding author.
E-mail address: mjbooysen@sun.ac.za (M.J. Booysen).

https://doi.org/10.1016/j.dib.2024.110732

main objective was to estimate energy requirements for the eventual electrification of these vehicles, the data was thus processed according to an electro-kinetic model. This high-fidelity mobility data was captured by "standardised passengers" with bespoke GPS-location logging devices sampling at 1 Hz. Trips on the four route types were recorded and saved in six folders – three routes, each in two directions, with one route being uphill in one direction and downhill in another. Each of the six folders have subfolders for time of day – morning, afternoon, and evening. In total 62 trips were recorded with varying durations, depending on the traffic and route length.

Specifications Table

Subject	Planning and development
Specific subject area	Paratransit mobility in sub-Saharan Africa. This data can be used to analyze energy requirements and mobility patterns of paratransit in the region.
Data format	Filtered
Type of data	Data Table (CSV)
Data collection	To capture the micro-mobility behavior of the minibus taxis, six tracking devices were used to record GPS data to an SD card at a frequency of 1 Hz. The six recording devices are based on the Arduino platform and powered from alkaline battery packs. The device can therefore operate independently of any other device during tests. Field workers, appointed to take trips on the taxis as standardized passengers, initiated data capture with the press of a button, which would be terminated once the vehicle had reached the destination. Each recorded trip creates an isolated file. This allows for different routes to be separately investigated and compared to other recordings made on the same route. Data was recorded on trips on three types of routes, in both directions, for a total of six distinct routes. Each route had trips recorded in the morning (before 11:30AM), afternoon (11:30AM-4PM) and evening (after 4PM). The three types of routes investigated were hilly, inter-city, and urban routes.
Data source location	Institution: Stellenbosch University / Oxford University City/Town/Region: Stellenbosch Country: South Africa The data was collected within the geographical region defined by the following coordinates: Latitude: −34.0846 to −33.8899 Longitude: 18.81974 to 18.96128 All GPS coordinates available in data.
Data accessibility	The raw data is hosted in a Mendeley repository, details below: Repository name: Survey data on the socio-economic characteristics and electric vehicle perceptions of paratransit owners and drivers Data identification number: 10.17632/xt69cnwh56.3 Direct URL to data: https://data.mendeley.com/datasets/xt69cnwh56/3
Related research article	Hull, Christopher, J. H. Giliomee, Katherine A. Collett, Malcolm D. McCulloch, and M. J. Booysen. "High fidelity estimates of paratransit energy consumption from per-second GPS tracking data." Transportation Research Part D: Transport and Environment 118 (2023): 103695. https://doi.org/10.1016/j.trd.2023.103695

1. Value of the Data

Planning for electric mobility requires an understanding of the energy requirements to provide the mobility. These energy requirements will inform the choice or design of the vehicle (e.g. drivetrain, battery capacity), and will also influence the charging requirements. In sub-Saharan Africa, where electric minibus taxis have not yet arrived, these estimations are currently based on analysing mobility patterns of ICE vehicles and estimating the energy requirements of EV equivalents. There are different ways of estimating electric energy requirements at the vehicular level, all linking some form (a) mobility estimation with some form of (b) efficiency of converting electric energy to mobility (and vice-versa for regenerative breaking).

The high frequency mobility (kinetic) measurements provided in this submission, are needed to present with a high time and spatial resolution what micro-mobility patterns are evident. The kinetic and potential energy inherent in the mobility samples can be combined with vehicle parameters, to provide an energy balance for each pair of samples, which can be used to estimate electrical energy requirements. An example of how this data can be used for energy estimation and planning for electrification can be found in Lacock et al. [5].

These measurements can therefore be used with an electro-kinetic model to:

1. Validate the average efficiencies of the vehicles in the sector for various types of routes.
2. Validate the driving simulation model for a given type of route and given vehicle.
3. Provide input for a machine learning model to potentially replace the driving simulation for a type of route and vehicle.

Given that the data acquisition can be done in a few days, this method is easily reproducible in other cities in the global south.

2. Data Description

Data was captured with handheld GPS devices taken on 62 minibus taxi trips around Stellenbosch South Africa. Data was recorded on three routes, in both directions, for a total of six distinct routes. Each route had trips recorded in the morning (before 11:30 a.m.), afternoon (11:30 a.m.-4 p.m.) and evening (after 4 p.m.). The three types of routes investigated were hilly, intercity, and urban routes.

The data was collected by students who used the 16-seater taxis like normal passengers. They had with them the GPS logger that was mounted into a food container. The logger recorded GPS location, speed and heading, and was inconspicuously placed on the lap of the student, with the axis aligned with the vehicle's direction of forward movement (not important for the presented data).

The data captured by the GPS module used for energy consumption estimates was location (lat/long), velocity (km/h), altitude (m), and time (hh:mm:ss). Details of these estimations and the model used to prepare the data for kinetic analysis can be found in Hull et al. [3], and subsequent analysis in Abraham et al. [1].

The raw and processed GPS data are located in the folders 'Raw Data' and 'Processed Data' respectively at the Mendeley Data repo in this link. In addition to the raw data, processed data files include the displacement between observations, calculated using Geopy's geodesic package, and the estimated energy provided by the vehicle's battery for propulsion, braking, and offload work.

All data is stored in comma separated variable (CSV) files. Each recorded trip is saved in a separate file. The data in these CSV files are stored with the following fields:

- Date [dd/mm/yyyy]
- Time [hh:mm:ss]
- Latitude [degrees]
- Longitude [degrees]
- Altitude [m]
- Speed [km/h]
- Heading (degrees from True North)
- Signal quality
- Number of connected satellites
- X-Axis acceleration [m/s^2]
- Y-Axis acceleration [m/s^2]

- Z-Axis acceleration [m/s^2]

The processed data has the following additional variables, which were calculated from the raw data and the kinetic model presented by Hull et al. [3]:

- Slope angle [rad]
- Displacement [m]
- Propultion Work [J]
- Braking Work [J]
- Offload Work [J]
- Energy Consumption [kWh].

The folders and csv files within ‘Raw Data’ and ‘Processed Data’ are structured in a route/time of day hierarchy. Each trip csv file is contained within the route/time of day folders that it was recorded in. The files, which are available in an n Mendelay Data repository at Hull et al. [4], are structured as follows, where `Kayamandi-Stellenbosch` is the folder for a specific route (in this case, Kayamandi to Stellenbosch), and the subfolders contain the Evening, Afternoon and Morning datasets. Each file contains the data captured during one trip on the route.

3. Experimental Design, Materials and Methods

The main objective of the data capture is to measure the micromobility apparent in minibus taxi mobility, to estimate energy requirements for eventual electrification. This micromobility measurement was done for minibus taxis in Stellenbosch, South Africa. Although the method of acquisition is generic, the data is specific to the vehicles used in this mode of paratransit, namely minibus taxis. Care was taken to set up the experiment to capture the main different driving conditions that could impact on micromobility, namely: urban (inside a city), inter-city, uphill, and downhill.

To get the high-fidelity mobility data, there are two main methods currently used:

1. The first method uses origin-destination travel distances (either from processed GPS tracking data or from route analysis/planning). This method then does one of:
   a. Assumes a constant electric vehicle efficiency (km/kWh, e.g. 2.5 km/kWh) that has not been validated in local driving and environmental conditions, or
   b. Uses simulation to estimate the micromobility between the origins and destination. In addition to being reliant on accurate electro-kinetic vehicle modelling, this method is highly reliant on accurate mapping information, accurate routing models/simulators, accurate traffic modelling, accurate driver modelling. As shown by Giliomee et al. [2], these assumptions cannot easily be inferred for the sub-Saharan African environment and should be validated and adapted with high fidelity tracking information. For example, reverse geolocation (to get street addresses from GPS location samples) with inherent data’s GPS location errors, could easily lead to substantial routing errors.
2. The third uses minutely (or similar) GPS data as mobility input. Like the origin-destination data method, it then does one of:
   a. Assumes unvalidated constant efficiencies, which again uses unvalidated efficiencies.
   b. Uses simulation to interpolate micromobility between the minutely samples. Due to the GPS waypoints between the origin and destination, this method is not as susceptible to reverse geolocation errors, but it still depends heavily on the map and driver virtualization.
   c. Uses the GPS waypoints (with speed data) directly as in an electro-kinetic model to convert the kinetic mobility into electric energy usage. This is likely to underestimate the energy requirements, as we show in Hull et al. [3], by effectively under-sampling the mobility changes, partially due to the way these vehicles drive.

For all these methods, the output from this simulation is then used in an electro-kinetic model to estimate the electric energy flows at a high time resolution.

Therefore, the high frequency mobility (kinetic) measurements provided in this submission, are needed to present with a high time and spatial resolution what micro-mobility patterns are evident.

It is likely that finer grain analysis could be done for the different driving conditions (e.g. analysing driving in a road with traffic lights vs. stop signs, analysing a multi-lane dual carriageway vs a single track unpaved road), but the intention with this dataset is to establish a first level of detail, and to demonstrate a method of characterizing micromobility for minibus taxis. To be sure, the intention is to capture how minibuses move in normal operations. This “move” was eventually reduced to high frequency measurement and calculation of

- Location at each sample, from which elevation can be derived.
- Distance travelled between samples.
- Speed travelled at for each sample.
- Acceleration from one sample to the next.

Data collection was done through assigned field workers taking mini-bus taxis on predetermined routes, while equipped with a GPS recording device. The field workers merely entered the taxi as passengers and hit the recording button once the taxi started moving.

Five third-year engineering students at Stellenbosch University was allocated to drive predetermined routes in and around Stellenbosch. They were instructed to start the recording when they left the taxi rank and terminate the recording once the destination has been reached. Each of the three routes was to be taken once during each time of the day. In total 90 recorded were expected. Due to loss of GPS signal at times, user error and some external factors (e.g. motion sickness), 62 successful recordings were used after filtering.

In total, six devices were assembled and used for data collection. The GPS device was designed and built at the University of Stellenbosch. It was based on an Arduino Nano platform. Sampling was set to be done on the next full second of the Arduino’s internal clock. However, it can take up to 0.6 s for the GPS module to retrieve a clean NMEA sentence to be processed. Depending on the time delay due to the GPS waiting for a clean sentence, every few samples could be logged at a time difference of 2 s. Acceleration data is measured every 100 ms and averaged out and logged according to the time difference between the samples.

The Arduino code can be found at https://github.com/22046364/GPS_Tracker.git.

Each device was equipped with a dedicated alkaline battery pack, supplying 12 V directly to the Arduino, shown in Fig. 1. All components are enclosed in a reused food container.

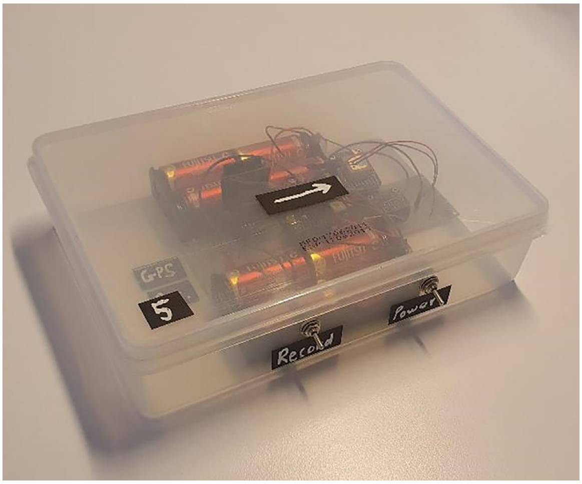

Fig. 1. Example instance of the bespoke data loggers developed for data capture.

Three breakout modules are used in each device. To capture acceleration, an Adafruit ADXL345 – Triple Axis Accelerometer is used. It is interfaced with the Arduino through I2C communication. An Adafruit Ultimate GPS Breakout was used to log GPS locations. This breakout includes an antenna. Serial communication is used with this module. All of the data is stored to an onboard SD card, through a High Speed SD/Micro SD Card Reader by Robotdyn. This is done through a SPI interface. Low correlation in the logged altitude by the GPS model was found, thus altitude data was added with the use an online altitude tool, gpsvisualizer.com, and using the GPS location as input.

The user interface is kept very simple, with only 2 inputs and 2 outputs. For input, a power and recording switch is located on the front of the device. Two red LEDs to indicate GPS signal fix and active recording is located inside the device, which can be seen through the transparent lid. The LED located on the Arduino serves as an indication that the device is successfully powered on.

Limitations

Although the dataset is useful to provide improved estimates of energy requirements, it has many limitations, bulleted here.

- The dataset is limited to a specific vehicle type. The vehicles are Quantum minibus taxis, and do not necessarily reflect the same performance (torque, speed, weight) as those used in other regions in the global south.
- The dataset is limited to the drivers who happened to drive the vehicles on the 62 trips that were recorded. More aggressive or less aggressive driving will result in different mobility.
- The dataset is limited to the state of our road infrastructure. The road infrastructure in and around Stellenbosch is generally of a high quality, except in Kayamandi. In other developing cities, the road infrastructure could be completely different.
- The dataset is limited to the traffic conditions in and around Stellenbosch.
- The method assumes that drivers will drive in the same way with electric vehicles. This could be severely impacted by the eventual performance of electric equivalents.

Ethics Statement

N/A.

Data Availability

1Hz GPS Tracking Data (Original data) (Mendeley Data)

CRediT Author Statement

Christopher Hull: Methodology, Software, Formal analysis, Data curation, Writing – original draft, Visualization; **J.H. Giliomee:** Methodology, Software, Formal analysis, Investigation, Data curation, Writing – review & editing, Visualization; **Katherine A. Collett:** Writing – review & editing; **Malcolm McCulloch:** Conceptualization, Methodology, Writing – review & editing, Supervision; **M.J. Booysen:** Methodology, Formal analysis, Resources, Writing – review & editing, Visualization, Supervision, Project administration.

Acknowledgements

M.J. Booysen and J.H. Giliomee acknowledge financial support from MTN South Africa (Contract S003061) and Eskom (Tertiary Education Support Programme). All authors acknowledge financial support from World Bank project Paratransit Decarbonization in South Africa.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C.J. Abraham, A. Rix, M.J. Booysen, Aligned Simulation models for simulating Africa's electric minibus taxis, World Electr. Veh. J. 14 (2023) 230, doi:10.3390/wevj14080230.
[2] Giliomee, J.H. Christopher Hull, Katherine A. Collett, Malcolm McCulloch, M.J. Booysen, Simulating mobility to plan for electric minibus taxis in Sub-Saharan Africa's paratransit, *Transport. Res. Part D*, Volume 118, 2023, 103728, 10.1016/j.trd.2023.103728.
[3] C.R. Hull, J.H. Giliomee, K.A. Collett, M.D. McCulloch, M.J. Booysen, High fidelity estimates of paratransit energy consumption from per-second GPS tracking data, Transport. Res. Part D 118 (2023) 103695, doi:10.1016/j.trd.2023.103695.
[4] C.R. Hull, J.H. Giliomee, K.A. Collett, M.D. McCulloch, M.J. Booysen, 1Hz GPS tracking data on minibus taxi paratransit vehicles in South Africa, Mendeley Data (2022) V1, doi:10.17632/xt69cnwh56.1.
[5] S. Lacock, A.A. du Plessis, M.J Booysen, Using driving-cycle data to retrofit and electrify sub-saharan Africa's existing minibus taxis for a circular economy, World Electr. Veh. J. 14 (2023) 296, doi:10.3390/wevj14100296.