Why taxi tracking trumps tracking passengers with apps in planning for the electrification of Africa’s paratransit

A.J. Rix, C.J. Abraham, M.J. Booysen
rix@sun.ac.za (A.J.R.)
24540463@sun.ac.za (C.J.A.)
mjbooysen@sun.ac.za (M.J.B.)

Highlights
Energy forecasting for paratransit electrification planning of seven African cities

Energy needs per city vary wildly, with the reliability of the city’s GTFS data

Discrepancies highlight the problem with using unreliable transport data

Public transport data (GTFS) is not suitable for electrification planning

Vehicle-based GPS data should be used for electrification planning of paratransit
Why taxi tracking trumps tracking passengers with apps in planning for the electrification of Africa’s paratransit

A.J. Rix,1,* C.J. Abraham,1,* and M.J. Booysen1,2,3,*

SUMMARY
Decarbonisation of Africa’s informal paratransit through electrification requires adequate data captured correctly. Field workers getting on-board as passengers with tracked phones are extensively used to measure flow rates and volumes of passengers and vehicles on sections of roads in transport planning applications. Although this method is acceptable for transport planning, it is inadequate for planning for electrification. Combustion engine vehicles have long ranges and refill fast. Drivers and fuel outlets have existed in a symbiotic relationship without the bondage of needing detailed mobility information and planning. With electrification, battery-powered vehicles have become inextricably coupled to roadside infrastructure through their mobility patterns. We compare the current state of public transport data with vehicle tracking data for forecasting the electrification of Africa’s paratransit. Discrepancies between them highlight the problem with using incomplete and/or unreliable data to estimate a city’s peak load, pointing to a need for vehicle-based data acquisition.

INTRODUCTION

“In light of the climate crisis, transport systems globally need to be decarbonized. This is particularly challenging in Sub-Saharan Africa (SSA) where transport systems are poorly characterized due to a lack of data, which contributes to hindering investment. We call for a more systematic approach to data collection to support the sustainable transition to electric vehicles in SSA.”—Collett and Hirmer (2021) in Nature Sustainability.

State of paratransit

Over the years, paratransit has become one of the most dominant forms of transport in Africa, providing countless jobs and cheap transport for millions of people every day. Paratransit is a mode of informal public transport, owned by small operators, and which does not follow fixed schedules or routes, but rather adapts to passenger demand. Paratransit is characteristic of African cities and accounts for over 70% of the modal share (excluding non-motorized modes of transport) (Behrens et al., 2016).

Although paratransit in the region takes various forms, such as minibus taxis, motorcycle taxis, and tuk-tuks (Ehebrecht et al., 2018), minibus taxis, in particular have become ubiquitous throughout Africa. They are able to carry many passengers (9–16, depending on the vehicle model) (Behrens et al., 2016) and, moreover, they are available at a low cost as second-hand imports from developed countries (Cervigni et al., 2013, p. 348). Unfortunately, this has led to the system being fraught with old, under-maintained vehicles, causing serious environmental concerns. One study done in 12 African cities found that the average age of minibus taxis was 15 years (Kumar et al., 2008). Powered by internal combustion engines, these vehicles contribute substantially to the emission of greenhouse gases and a general decline of air quality in African cities (Collett and Hirmer, 2021).

Part of the reason that Africa’s public transport is in this state is that there has been very little regulation of the paratransit industry. The paratransit industry originated from the collapse of state-owned public transport throughout the continent, during the restructuring of World Bank policies in the 1990s (Kumar, 2011; Kumar et al., 2008; Cervero and Golub, 2007). As a result, paratransit grew organically to fill this vacuum. By
the time governments sought to control the transition, the system had already grown to a size that made it resistant to change (Behrens et al., 2016; Jennings and Behrens, 2017).

Although there is a lack of regulation, it does create the advantage that the system has low barriers to entry, encouraging small businesses to enter the market. However, taxi operators also misuse this freedom to cut costs in taxi maintenance to squeeze profits from this relatively flooded market. At the same time, the urban poor have resigned to the poor standards of transport because they cannot afford any better.

Some countries have made various interventions to try to improve the quality, safety, and environmental footprint of these vehicles. Although some of these countries have attempted to establish control through regulation, this has often failed because of poor enforcement or resistance from taxi operators. Kumar et al. (2008, Sec. 3) give an overview of failed regulation attempts in various African cities. In other cases, countries have tried to financially subsidize the sector to upgrade their fleets. Kumar et al. (2008, Table 2.2) provide a few such examples. This has been successful in relatively economically advanced countries such as South Africa, which incentivized taxi owners with around USD 7500 to scrap their old taxis and replace them with newer and safer vehicles (Behrens et al., 2016). However, in most African countries paratransit subsidies do not exist and would be unaffordable by the government (Kumar et al., 2008).

Some countries have also tried to override the paratransit industry by competing against it with more formal modes of mass transit, such as metro rail, bus rapid transit, and the like. However, these have been fraught with poor government spending, poor integration with existing paratransit, and general incompatibility with the dynamic transport needs of the urban poor. All this has led to formal mass transit being a burden on the taxpayer (Stead and Pojani, 2017).

State of electric vehicles in Africa

The market penetration of electric vehicles has been extremely low in Africa. For example, South Africa had fewer than 1200 electric vehicles as recorded in 2019 (IEA, 2020). It can be seen that the electric vehicle revolution has largely been a trend in developed countries. However, this trend will inevitably find its way to Africa, because of Africa’s high reliance on vehicle imports from developed countries (IEA, 2021, p. 45). (60% of light-duty vehicles in the continent are second-hand imports from developed countries.) In 2021, although electric vehicle sales in Africa were extremely low, it increased by 90% with respect to 2020 values (IEA, 2022). It is only a matter of time before the African transport landscape is transformed.

In order for the successful migration of Africa’s transport to electric vehicle technology, it is important that the methods are not copied from developed countries. Rather the blueprint needs to be adapted and reworked in order to fit the context of Africa. The context of Africa is severely different from developed countries because of at least three reasons (Collett et al., 2021; Booysen et al., 2022a):

1. Vastly different mobility/traffic patterns and modal share.
2. Low availability of electric utility.
3. Limited capital availability, both on the utility end and the customer end.

However, Collett et al. (2021) do outline a number of advantages that electric vehicles would have in Africa on various stakeholder levels. In addition, the expectation for electric vehicles is set, as evidenced by the extensive database for electric vehicle policies for 52 African countries by the United Nations Environment Programme (2018).

Data for paratransit sustainability

To improve the state of African paratransit in future, a different approach needs to be taken. One paradigm through which to achieve sustainable mobility is the three-pronged “Avoid-Shift-Improve” approach (Transformative Urban Mobility Initiative, 2022; Galuszka et al., 2021). This paradigm follows three steps. The first is to avoid inefficiencies. For example, this can be done by reducing redundant taxi trips through computer optimization. The second step is to shift to alternative technologies (e.g., electric vehicles). The software presented in this article can be used to investigate the energy demands of various technologies to select the best alternative. Finally, the last step is to continuously improve the new technology through various efficiency upgrades. For example, an electric paratransit system can be enhanced with more
and/or faster charging stations, integration of solar energy charging, and more efficient vehicles. All three of these steps require adequate data to drive the process.

Unfortunately, instances of large-scale capture of paratransit data are scarce, leading to poor decision making (Collett and Hirmer, 2021). A few known attempts conducted in Africa are summarized in Table 1, and these are outlined in more detail in the following paragraphs.

**Passenger-based tracking data**

The paratransit data capturing approaches can be grouped into four categories: vehicle-based tracking, passenger-based tracking, roadside-based counting and household travel surveys.

---

**Table 1. Paratransit data-capturing methods**

| Data Collection Method | Vehicle-based Tracking | Passenger-based Tracking | Roadside-based Counting | Household Travel Survey |
|------------------------|------------------------|--------------------------|-------------------------|-------------------------|
| Strengths              | ✓                      |                          |                         |                         |
| Data is collected on a per-taxi basis. | ✓ | | | |
| This allows one to analyze the net energy that a taxi uses on any day of operation. | | | | |
| The times that the taxi makes stops between routes is known, allowing one to analyze how long the taxi has to charge on any day of operation. | | | | |
| The data is extremely robust since it is collected continuously and the taxi’s day plan is recorded repeatedly. Hence, the data is updated continuously. | | | | |
| Requires very little consultation with paratransit operators. | | | | |
| Minimal technical expertise required. | | | | |
| Can be cost-effective in countries where low-skilled labor is relatively cheap, if updating the dataset is not planned. | | | | |
| Weaknesses             | ✓                      |                          |                         |                         |
| GPS traces do not inform where and when a particular route stops and ends. Neither does it inform when a taxi stopped to deliver passengers (as opposed to stopping at a traffic light). | | | | |
| Further data analysis needs to be done on the collected data to make it useful. | | | | |
| Takes very long to capture data. | | | | |
| Data needs to be manually updated periodically. | | | | |
| No information on the exact itinerary of each of the taxis. | | | | |
| Data is has very little detail. | | | | |
| No information is known about the routes the vehicles take. | | | | |

¹Certain routes may take very long to traverse and need to be taken multiple times to obtain confidence in the data.
²A large number of respondents need to be manually interviewed.
We have found that out of these various paratransit data collection methods, the passenger-based tracking method has become the most prevalent. A list of many such datasets was presented by du Preez et al. (2019) in addition to the seven datasets used in this article (listed in Table 2). To compile this data into a consumable digital format, several standards could be considered. Most of the datasets made use of the General Transit Feed Specification (GTFS), which is an open format used for digitally specifying public transport routes and schedules. Additional datasets were collected by a company called GoMetro (du Preez et al., 2019) which uses their own custom specification. However documentation of their specification is not publicly available, and the GTFS standard is more widespread globally.

**GTFS data format specification**

GTFS was originally developed for formal public transport systems, which have predetermined timetables and routes. In the case of paratransit, schedules and routes are fluid and based on the demand. Hence, it becomes necessary to collect sufficient data to synthesize “typical” routes and schedules within a boundary of confidence. The GTFS standard has provisions to capture (1) the shapes of the various routes, (2) location, arrival time, and duration of stops along each route, and (3) what times of the day a taxi departs on each route.

The popular approach for capturing these data for paratransit systems is the one pioneered by The Digital Matatu Project (Williams et al., 2015). This approach employs data collectors equipped with mobile phones who ride on the taxis to collect data. As the data collector makes a trip on a particular route, the mobile phone records the path taken and the data collector will make a record when the taxi stops to drop off or pick up passengers. The advantage of this is that, ethical concerns notwithstanding, very little permission is required from the taxi association to collect data because nothing is installed on the vehicles. As a result, the capital costs of the data captured is minimized.

However, there are some major disadvantages with this approach. First, the data is only valid for as long as the paratransit system continues to follow the same movement patterns. However, in reality, paratransit movement patterns often adapt to the evolving demand of its customers. Therefore, data will need to be re-captured at regular intervals to keep the dataset up to date. Although unskilled labor is often cheap in Africa’s developing countries, the cost can quickly compound because of the number of person-hours required to collect the data. This could be the reason that none of the datasets we analyzed have been updated since their initial release.

**Digital transport for Africa dataset**

DigitalTransport4Africa (2021b) (DT4A) is a concerted effort to collect GTFS data for paratransit systems in African cities. The collection of datasets is hosted publically online, and can be used by anyone. Furthermore, it is easy to contribute to the collection, allowing for entities to add their own GTFS data and request corrections to existing datasets. This has facilitated rapid growth in the size of the collection. Currently, the site hosts GTFS data for the paratransit systems of eleven African cities (Abidjan, Accra, Addis Ababa, Cairo, Douala, Freetown, Harare, Kampala, Kumasi, Mali, and Nairobi). The quality of this data will be evaluated in this article, and the data will be used to estimate Africa’s readiness for EV taxis, while at the same time highlighting how the state of the art in paratransit data collection can be improved for future EV evaluations.

---

**Table 2. Data collection method, recency, and rigor**

| City    | Approach           | Months since last update | Number of Collectors | Number of Weeks | Reference | Effort |
|---------|--------------------|--------------------------|----------------------|-----------------|-----------|--------|
| Abidjan | Onboard + Static   | 2                        | 7                    | ~ 24            | (DigitalTransport4Africa, 2021a) | High   |
| Accra   | Onboard + Static   | 36                       | 6                    | ~ 5             | (Transitec, 2018) | Low    |
| Cairo   | Onboard            | 42                       | 19                   | –               | (TICD and TFC, 2017) | Low    |
| Freetown | Onboard            | 27                       | –                    | 2               | (Matekenya et al., 2021) | Low    |
| Harare  | Onboard            | 29                       | –                    | 10              | (Fried, 2019) | Low    |
| Kampala | Onboard            | 15                       | –                    | 10              | (Transport for Cairo, 2020) | Medium |
| Nairobi | Onboard            | 27                       | 5                    | ~ 24            | (Williams et al., 2015) | High   |
Aims and objectives
This article therefore makes use of the available GTFS data to:

- Forecast electric paratransit energy demands in the various paratransit systems.
- Evaluate the GTFS method’s suitability for electric paratransit forecasting.
- Validate if the quality of currently available GTFS paratransit data is acceptable.

This article specifically focuses on minibus taxis because that particular mode of paratransit is the most prevalent across Africa (Behrens et al., 2016). Although the various GTFS datasets in the DT4A collection contained various modes of paratransit, the minibus taxi was the only mode present in all available GTFS datasets.

Summary
This article will use the state-of-the-art of paratransit data (DT4A datasets) to identify its suitability for projecting the energy of electric paratransit. The article first devises a few metrics to get a broader picture of the data quality, and then seeks to compute an energy analysis of this data. The energy analysis seeks to answer the following questions. First, from the perspective of the grid: How much additional energy would the system require, and what would the peak load be on the grid? Second, from the perspective of the vehicle: What battery size would be required, and what are ideal times for the taxi to stop and charge? Finally, the article will discuss the suitability and reliability of using this data to answer these questions. If it is found that the GTFS data format is not descriptive enough to answer some of our research questions, the article will suggest an alternative data format or suggest how the current data format can be improved.

Our custom-built software, which we call EV-Fleet-Sim, was used to obtain the results in this article. This software and its source code can be downloaded from the following website: https://gitlab.com/eputs/ev-fleet-sim. We publish the software for free use and modification, subject to the terms of the GPLv3 license. The data by DigitalTransport4Africa (2021b) is also publicly available, allowing for easy reproduction of the results: https://gitlab.com/digitaltransport/data.

RESULTS
In this section we discuss the results obtained from the various steps of the methodology. We first assess the reliability of the data and then evaluate what can be expected from the simulation. The simulation results are then presented related to the reliability. Finally, the simulation results are verified with the verification metrics discussed in the methodology section. Throughout this section we suggest reasons for data reliability issues and the consequences thereof.

Data reliability analysis
Table 2 captures a few metrics that indicate the method, recency, and rigor with which each dataset was collected.

Because four of the seven datasets were missing sufficient documentation, we were unable to populate some of the cells. Unfortunately, this reduced our ability to interpret the reliability of the data.

The table shows that the Abidjan dataset has a high effort rating. This was because they did more person-weeks of work (7 × 24 = 168 person – weeks) than any other dataset. We considered datasets with limited documentation also deficient in reliability. For each missing metric we assigned the worst value found in the other populated rows. As a result, Freetown and Harare were classified as low effort. As stated, this person-weeks value does not have much practical significance, but is useful for providing a starting point for classifying the data collection effort in the absence of more useful metrics/documentation.

Furthermore, all datasets provided paratransit data only for a “typical weekday”. In other words, the dataset did not express variation between days of the week, nor did it express the variation between seasons of the year. If the data was more descriptive, perhaps we could quantify how much less active taxis are on the weekend, when compared to the midweek, and how much later taxis start operating during winter, when compared to summer.
Although the datasets had these various issues, they were all relatively up to date. Only 1 dataset (Cairo) had not been updated in the last three years.

However, effort alone is insufficient as a proxy for reliability of a dataset. Table 3 compares the effort with the size of each dataset, using various metrics, and classified the datasets as small, medium or large.

Finally, we were able to classify the reliability of each of the datasets. For example, the Abidjan data collectors took a lot of effort in producing their dataset, although the measured paratransit size was small, instilling high confidence in the reliability of the dataset. On the other hand, with Harare, although the dataset was large, relatively little effort was exerted, resulting in an apparent low reliability. These classifications are shown in the last column of Table 3.

We can expect datasets with a high reliability to produce results that are more verifiable.

**EV simulation results**

The seven datasets were simulated with our EV-Fleet-Sim platform, which yielded the (aggregated vehicle) power and energy profiles per day, shown in Figure 1.

Four of the cities (Harare, Kampala, Nairobi, and Freetown) have a peak in their power profile in the mid-morning hours (around 9 AM). All of these except one (Kampala) also have an evening peak of power usage (at around 6 PM).

Cairo on the other hand displays a different power usage profile. Unlike the other cities, Cairo is north of the Sahara Desert. Its geographical differences may have different living and/or mobility patterns.

The remaining two cities (Abidjan and Accra), do not have any peaks in their power profiles. Rather, they appear flat with low power consumption profiles. This corresponds with the fact that the geographical spread of the cities is small as indicated by Table 3. It could also indicate that the cities use other modes of transport to satisfy the bulk of their mobility demands. These cities are geographically in closeness, proximity, in North-West Africa. Freetown, the remaining city from North-West Africa, also has a low overall energy usage like Abidjan and Accra.

The total energy demand of each of the simulation scenarios is summarized in Table 4. These values are be used to perform verification. We use data from an alternative source to approximate the energy requirements of each of the cities, and compare them to the results found in Table 4.

We also generated boxplots (Figure 2) which indicate the per-route energy usage of each paratransit system. If each route was apportioned a dedicated fleet of vehicles to service that route, the boxplots show how much energy the fleets would use. Consequently, if each route was apportioned a dedicated charging station for its fleet, the boxplots indicate how much energy the charging stations would require from the grid. Of course, such a scenario would be sub-optimal, but it was the only scenario we could think of that would allow the GTFS data to be useful on a disaggregated level.

**Table 3. Size and scope of the datasets**

| City     | Number of routes | Daily number of trips | Simulation area (km²)* | Dataset Size/Required Effort | Reliability |
|----------|------------------|-----------------------|-------------------------|------------------------------|-------------|
| Abidjan  | 73               | 8890                  | 1850                    | Small                        | High        |
| Accra    | 277              | 49214                 | 1356                    | Medium                       | Low         |
| Cairo    | 94               | 29000                 | 4150                    | Medium                       | Low         |
| Freetown | 69               | 16954                 | 530                     | Small                        | Medium      |
| Harare   | 486              | 86733                 | 4639                    | Large                        | Low         |
| Kampala  | 369              | 92262                 | 2478                    | Large                        | Low         |
| Nairobi  | 136              | 35640                 | 3097                    | Large                        | Medium      |

*Simulation size was determined by creating a rectangular box that bounded all routes in the dataset.
Results verification

To verify the results and see if they corresponded to what could be expected in reality, the taxi energy demands were categorized as high, medium, or low based on the city’s population size and the taxi modal split, as gathered from independent data sources. These are shown in Table 5. As shown in the table, Cairo is expected to have a high energy demand because it has an extremely large population that relies on taxis. Conversely, Freetown is expected to have a low energy demand because it has a relatively small population, and they don’t rely on taxis much, as indicated by the low modal split.

For the most part, the energy usage from the simulations (in Table 4) corresponds with the expected energy demands (in Table 5). For example, Cairo has the highest simulated energy usage, and it also has a high expected energy demand. Abidjan has the lowest simulated energy usage, and it also has a low expected energy demand. The only results that did not correspond well were for Harare and Kampala. Harare had a higher simulated energy usage than expected, whereas Kampala had a lower simulated...
energy usage than expected. This can be attributed to the fact that both of those datasets had a low reliability rating in Table 3.

We conclude that GTFS data can be used to give a rough estimate of the order of magnitude of a paratransit system’s energy demands. However, because of the current state of data quality, they are likely to be unreliable for quantifying the exact energy demands, as shown in the next section.

The original intention was to approximate the expected energy demand using the number of taxis in each paratransit system. However, this proved to be difficult because of unreliable numbers found for this metric because the transport authorities in most of the countries do not publish it. We therefore relied on independent sources, whose figures do not agree. For example, Harare’s transport authority did not publish any figure on the number of taxis, and the figure of 100,000 which we found from an independent source seems disproportionate when one takes into account the city’s population size. In addition, that number was the most recent we could find, although it is from 2002. This highlights, once again, the need for some level of systematic data collection.

Reliable statistics on the number of taxis would have also been useful for downscaling the power and energy profiles in Figure 1, to get the average taxi’s energy demand. In the following paragraphs, we will nevertheless downscale the Kampala profile by the number of taxis to compare its power profile to one we computed from independently gathered, reliable, vehicle-based paratransit data (Booysen et al., 2022b).

## DISCUSSION

### Passenger-based data shortcomings

Transport engineers typically obtain information by employing people in two outdated ways, either by standing next to the road to record inflows and outflows or by getting into vehicles to act as human carriers of tracking devices. These methods are particularly attractive in developing countries where operational traffic monitoring infrastructure is sparse and cheap labor is abundant. But, as mentioned in Table 1, this shortcut has many pitfalls.

Developments in vehicle tracking technology have changed the way in which data can be captured. Although the setup cost is more, tracking devices have the substantial advantage that they are not susceptible to human behavioral problems. For example, they do not wake up late, get tired and need eating breaks, and they do not sleep.

To illustrate the difference, in Figure 3 we show Kampala’s energy profile from the two vantage points. The overlay shows two plots. One is the passenger-based power profile from Figure 1C, which was downscaled by the number of taxis in Kampala (25,000, as shown in Table 5 (Spooner et al., 2020)). The second plot is a vehicle-based power profile of Kampala obtained in a small study by Booysen et al. (2022b). The differences are stark. First, the vehicle-based dataset reveals that the taxis started moving at around 4 AM, before the fieldworkers managed to get on board in the passenger-based dataset. Second, the passenger-based profile grinds to a halt just before dinner time. But the vehicle-based profile shows that the minibus taxis are mobile, even more so than in the morning, until after 9 PM. This makes passenger-based data unreliable for estimating the stop times (and hence the charging requirements) of the taxis.

| City      | Daily Energy Usage (MWh/day) |
|-----------|-----------------------------|
| Cairo     | 3778                        |
| Harare    | 3053                        |
| Kampala   | 2835                        |
| Nairobi   | 1612                        |
| Accra     | 1405                        |
| Freetown  | 943                         |
| Abidjan   | 250                         |

Table 4. Summary of simulation results
Using the passenger-based data to estimate the charging requirements is also made difficult by the fact that the data does not track the individual taxis. Hence, although we know that there was a lull of activity at around midday for some of the datasets, we don’t know which taxis were stopped at that time. Even if one is to assume that all vehicles were captured while mobile by tracking passengers, there is no way to tell exactly when each vehicle was actually stationary, how long the vehicle was stationary and where it was stationary. All of this information is required to quantify the charging requirements of the taxis on the grid, and is only reliably available through vehicle-based data capture.

Clearly, if we make assumptions about the power profile and energy requirements of electric minibus taxis from passenger-based data, we will be wrong by a substantial margin. This adds weight to the statement by Collett and Hirmer that we quoted earlier, stressing the need for a more systematic approach to data collection for the purpose of planning the decarbonization of paratransit (Collett and Hirmer, 2021).

Concluding remarks

Climate change is forcing the world to adapt to new, cleaner technologies, especially in the transport sector, which contributes a large share of carbon emissions. Developed countries have committed to ambitious plans to reduce their emissions in the short term, and are turning away from fossil fuel vehicles. Eventually, developing countries will be compelled to do the same, and must be prepared for such big changes.

This article prepares for this eventual migration by providing a method for modeling and simulating electric paratransit, Africa’s most used mode of motorized transport (Behrens et al., 2016). Using this method, this article has attempted to establish the energy requirements that paratransit would have on the electric utilities of several cities across Africa. Publicly available GTFS data was used for the simulations to evaluate how the state of the art can be improved for future evaluations, and whether the approach of paratransit data collection should be changed.

The first objective was to establish the net energy demand that the paratransit systems would have on each of the cities. This was plotted in Figure 1 and summarized in Table 4. Although the presented method was able to obtain the net energy demands of the paratransit systems, our evaluation of the data collection processes revealed a few quality issues, pointing to potentially unreliable data. Therefore, we propose more rigorous approaches in future paratransit data collection projects. Despite this, the results still give an indication of what the future energy demand would look like and what the taxis’ movement profiles would look like.

Four of the cities (Harare, Kampala, Nairobi, and Freetown) indicate that the highest energy demand of the taxis is at around 9 AM. This would indicate that taxis have a period of high activity in the so-called “morning rush”. It is clear therefore that taxis need to be adequately charged overnight to sustain this high-activity period. Three of these cities also indicate a high-activity period in the evening. This would mean that the taxis may also need to charge around midday when there is a period of relative inactivity.
Although we know the periods of relative inactivity, we don’t know exactly how long the typical taxi stops. This is because the data does not capture the individual taxis’ itineraries. Rather, it gives the perspective from an infrastructure, passenger demand, and traffic modeling point of view. Specifically, it captures the itinerary and frequency of each route and assumes that taxis will be available to service the route. Therefore, because we do not know when the taxis stop, it is not possible to know exactly when and how much the peak strain on the electric grid would be due to the charging of the vehicles. To do this, tracking data of the individual taxis will be required. At best, with this data, we have an indication that the peak charging strain on the grid would be approximately the peak power usage from the electric vehicles, but a few hours later.

If there were tracking data of the individual taxis, we would know exactly which routes they serve during the day, and how long their breaks are between routes. We would then be able to profile the individual taxis’ daily energy requirements which would be useful for calculating the battery size required for the taxis. This would have also been possible to find this metric through the GTFS data, if we knew the number of taxis operating the paratransit systems. However, statistics on the number of taxis operating in the cities are unreliable.

To evaluate the renewable energy generation capacity that is needed to supplement the grid to charge the taxis, we would need to know the stopping times of the taxis. Because most renewable energy resources are time-variant, the exact stopping times and durations are important. From the GTFS data, we have identified that, in some of the cities, the taxis have a dip in activity at midday, which presents an opportunity for charging from solar energy. However, to quantify exactly how much solar resource is required, the stopping times need to be known.

In conclusion, the GTFS data that seems to be the state of the art in paratransit data capture is useful in giving us an idea of the total energy requirements of a paratransit system, in giving indications of the supply-side energy requirements, in estimating the distributed battery resource available to the electric network, and in identifying opportunities for charging the taxis through renewable energy generation. However, the data quality needs to be improved to establish enough confidence in the data to quantify these energy requirements and opportunities with a scientific level of accuracy. More rigor needs to be applied when collecting the data, possibly by increasing the number of field workers and the number of times data is collected on each route. The results clearly indicate that it is not possible to identify the per-vehicle energy requirements from GTFS data. Although an attempt was made to scale down the power profile by the number of taxis, we found that the result did not correspond satisfactorily with independently gathered GPS tracking data. This has highlighted the issues of human data collection, calling for a more systematic, autonomous approach to capture data. With all this in mind, the authors recommend that
future data collection efforts employ the vehicle-based strategy (e.g., through GPS trackers). Through higher quality data driving the process of decision making, we can finally compute our way toward sustainable mobility in Africa.

Limitations of the study

Publicly available GTFS data was used for to evaluate the energy requirements and whether this approach of paratransit data collection should be changed. Our evaluation of the data collection processes revealed a few quality issues, pointing to potentially unreliable data and the energy requirement modeling therefore would not be absolutely accurate, although the comparative results still have merit and gives an indication of what the future energy demand would look like.

The periods of relative fleet inactivity is known but we don’t know exactly how long the typical taxi stops, therefore it is not possible to know exactly when and how much the peak strain on the electric grid would be because of the charging of the vehicles. With the lack of available tracking data we could not accurately calculate the required battery size for the taxis or evaluate the renewable energy generation capacity that would be needed to supplement the grid.

The GTFS data quality needs to be improved to establish enough confidence in the data to quantify the energy requirements and opportunities with a scientific level of accuracy. For electrification planning a vehicle-based data collection strategy (e.g., through GPS trackers) will assist the process of decision making and allow us to compute our way toward sustainable mobility in Africa.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Data retrieval and verification
  - Data reliability analysis
  - Minibus taxi mobility modeling from data
  - The taxi EV model

Figure 3. Comparison of Kampala per-vehicle power profiles derived from passenger-based and vehicle-based data
ACKNOWLEDGMENTS

We acknowledge MTN South Africa (Contract 5003061), Eskom (Tertiary Education Support Program) and the National Research Foundation of South Africa (Reference number: MND200609529659) for funding this research through the MTN Mobile Intelligence Lab.

AUTHOR CONTRIBUTIONS

A.J.R.: Conceptual design, analysis and interpretation, editing, and supervision. C.J.A.: Writing original text, implementation of simulation and visualization. M.J.B.: Conceptual design, data acquisition, analysis and interpretation, editing, supervision, funding.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Abane, A.M. (2011). Travel behaviour in Ghana: empirical observations from four metropolitan areas. J. Transport Geogr. 19, 313–322. https://www.sciencedirect.com/science/article/pii/ S0966692310000347?via%3Dihub.

Abraham, C.J., Rix, A.J., Nidibatya, I., and Booyse, M.J. (2021). Ray of hope for sub-Saharan Africa’s paratransit: Solar charging of urban minibuses in South Africa. Energy Sustain. Dev. 64, 118–127. https://doi.org/10.1016/j.esd.2021.08.003.

Addelkhoum, A., and Bonnafous, A. (2001). Urban transport microenterprises in abidjan. https://openknowledge.worldbank.org/bitstream/handle/10986/9905/28537hingend18.pdf?sequence=1.

Akpa, N E.E., Booyse, M., and Sinclair, M. (2016). Publicly available annotated dataset of tracked taxis. https://gitlab.com/eputs/data/fcd-stellenbosch.

Andrade, P.R., Pereira, R.H.M., Vieira, J.P.B., and d’Hersz, Saraiva, M. (2021). ipeagit/gtfs2gps: v1.5-4. https://doi.org/10.5281/zenodo.5501855.

Behrens, R., McCormick, D., and Mfinanga, D. (2016). Paratransit in African Cities: Operations, Regulation and Reform, First edition (Routledge).

Booyse, M.J., Abraham, C.J., Rix, A.J., and Nidibatya, I. (2022a). Walking on sunshine: pairing electric vehicles with solar energy for sustainable informal public transport in Uganda. Energy Res. Soc. Sci. 85, 102403. https://doi.org/10.1016/j.ersc.2021.102403.

Bruun, E., Mistro, R.D., Venter, Y., and Mfinanga, D. (2016). The state of public transport systems in three Sub-Saharan African cities. In Paratransit in African Cities: Operations, Regulation and Reform, First edition, R. Behrens, D. McCormick, and D. Mfinanga, eds. (Routledge). chapter 2. https://doi.org/10.4324/9781315849515.

Cervo, R., and Golub, A. (2007). Informal transport: a global perspective. Transport Pol. 14, 445–457. https://doi.org/10.1016/J.TRANPOL.2007.04.011.

Cervigni, R., Rogers, J.A., and Dvorak, I. (2013). Assessing Low-Carbon Development in Nigeria: An Analysis of Four Sectors (World Bank). https://openknowledge.worldbank.org/handle/10986/15797.

Collett, K.A., and Himrer, S.A. (2021). Data needed to decarbonize paratransit in Sub-Saharan Africa. Nat. Sustain. 4, 562–564. https://doi.org/10.1038/s41893-021-00721-7. https://rdcu.be/cKcez.

Collett, K.A., Himrer, S.A., Dalkmann, H., Crozier, C., Mulugetta, Y., and McCulloch, M.D. (2021). Can electric vehicles be good for sub-saharan africa? Energy Strategy Rev. 38, 100722. https://doi.org/10.1016/j.esr.2021.100722. https://www.sciencedirect.com/science/article/pii/S2211467X21001085.

DigitalTransport4Africa. (2021a). Abidjan transport. http://doc.digitaltransport.io/abidjantransport.

DigitalTransport4Africa. (2021b). DT4A data. https://github.com/digitaltransport/data.

Ehebrecht, D., Heinrichs, D., and Lenz, B. (2018). Motorcycle-taxis in sub-Saharan Africa: current knowledge, implications for the debate on “informal” transport and research needs. J. Transport Geogr. 69, 242–256. https://doi.org/10.1016/j.jtrangeo.2018.05.006.

El Esawey, M., and Ghareib, A. (2009). Analysis of mode choice behavior in greater Cairo region. Technical report. https://www.researchgate.net/profile/Mohamed-Elesawey/publication/282134483_Analysis_of_Mode_Choice_Behavior_in_Greater_Cairo_Region/links/594e304a0f7e9be7b2da7963/Analysis-of-Mode-Choice-Behavior-in-Greater-Cairo-Region.pdf.

Falchetta, G., Nousseman, M., and Hammad, A. (2021). Comparing paratransit in seven major african cities: an accessibility and network analysis. J. Transport Geogr. 94, 103131. https://doi.org/10.1016/j.jtrangeo.2021.103131. https://www.sciencedirect.com/science/article/pii/S0966692321001848.

Fridlund, J., and Wilen, O. (2020). Parameter Guidelines for Electric Vehicle Route Planning.

Fried, T. (2019). Readme. https://git.digitaltransport4africa.org/data/africa/harare/-/blob/178fe08537240b8a3fa8f2d72186b6b00fa1244/README.md.

Galusza, J., Martin, E., Nikununiza, A., Achieng’ Oginga, J., Senyagwa, J., Teko, E., and Lah, O. (2021). East Africa’s policy and stakeholder integration of informal operators in electric mobility transitions—Kigali, Nairobi, Kassu and dar es Salaam. Sustainability 13, 1703. https://doi.org/10.3390/su13041703.

GeoFabrik. (2022). OpenStreetMap data extracts. https://download.geofabrik.de.

Graeff, J. (2009). The organization and future of the Matatu Industry in Nairobi, Kenya. In Paper prepared by Center for Sustainable Urban development (earth Institute, Columbia University), pp. 1–3. http://www.vref.se/download/18.53e8780912f2dbb3e5b80002295/1377188299857/Organization+and+Future+of+the+Matatu+Industry.pdf.

IEA (2020). Global ev outlook 2020. https://www.iea.org/reports/global-ev-outlook-2020.

IEA (2021). Global ev outlook 2021. https://www.iea.org/reports/global-ev-outlook-2021.
iScience Article

IEA (2022). Global ev outlook 2022. https://www.iea.org/reports/global-ev-outlook-2022.

Jennings, G., and Behrens, R. (2017). The case for investing in paratransit: strategies for regulation and reform. https://trid.trb.org/view/1660898.

Johnstone, L., and Ratanavaraha, V. (2020). Public transport initiatives in West Africa, a case study of Abidjan. Int J Geomat 19, 197–204. https://www.geomaticjournal.com/user/download/1929/197-204-5529-Leonard-Aug-2020-72.pdf.

Koroma, B., Oviedo, D., Yusuf, Y., Macarthy, J., Cavoli, C., Jones, P., Careen, L., and Selli, S. (2021). City profile: freetown: base conditions of mobility, accessibility and land use. https://discovery.ucl.ac.uk/id/eprint/10122487/1/

Kumar, A., Kwakye, E., and Girma, Z. (2004). What works in private provision of bus transport services—case study of Accra and Addis Ababa. http://www.codatu.org/wp-content/uploads/What-works-in-private-provision-of-bus-transport-services-case-study-of-Accra-and-Addis-Ababa-A.-KUMAR-E.-KWAKYE-Z.-GIRMA.pdf.

Kumar, A., Kwakye, E., and Girma, Z. (2004). What works in private provision of bus transport services—case study of Accra and Addis Ababa. In 11th Conference of CODATU. http://www.geomatejournal.com/user/download/1929/197-204-5529-Leonard-Aug-2020-72.pdf.

Kumar, A., Kwakye, E., and Girma, Z. (2004). What works in private provision of bus transport services—case study of Accra and Addis Ababa. J. Transport Geogr. 37, 694–707. https://doi.org/10.1016/j.jtrangeo.2019.102853.

Mbara, T.C. (2002). Activity patterns, transport and policies for the urban poor in Harare, Zimbabwe. Final country report [En ligne]. www.transport-links.org/transport-links/filearea/documentstore/305_Zimbabwe%20Final%20Report.

Ndlubaya, I., and Booyens, M.J. (2020). Minibus taxis in Kampala’s paratransit system: operations, economics and efficiency. J. Transport Geogr. 88, 102853. https://doi.org/10.1016/j.jtrangeo.2020.102853.

OpenStreetMap Contributors (2022). OpenStreetMap data. https://www.openstreetmap.org.

da Preez, D., Zuidgeest, M., and Behrens, R. (2019). A quantitative clustering analysis of paratransit route typology and operating attributes in Cape Town. J. Transport Geogr. 80, 102493. https://doi.org/10.1016/j.jtrangeo.2019.10.005.

Spooner, D., Mwanka, J.M., Natamba, S., and Manga, E.O. (2020). Kampala bus rapid transit: understanding kampala’s paratransit market structure. https://www.researchgate.net/profile/Dave-Spooner/publication/342233740_Kampala_Bus_Rapid_Transit_Understanding_Kampalas_Paratransit_Market_Structure_/links/5e9250a6dcd73e5828b60/Kampala-Bus-Rapid-Transit-Understanding-Kampalas-Paratransit-Market-Structure.pdf.

Stead, D., and Pajoni, D. (2017). The urban transport crisis in emerging economies: a comparative overview. In The Urban Transport Crisis in Emerging Economies, D. Pajoni and D. Stead, eds. (Cham: Springer International Publishing), pp. 283–295. https://doi.org/10.1007/978-3-319-43851-1_14.

TICD, TFC (2017). How can transit mapping contribute to achieving adequate urban mobility? the case of Cairo, Egypt. https://giiddigitransport4africa.org/data/africa/cairo/-/blob/ae279a3097e0c1a1f99003acfd072410da947778/Research%20Policy%20Papers/TICD_TFC_How%20can%20Transi%20Mapping%20Contribute%20to%20Achieving%20AUM%202016-2017%20%20Resource%20Center.pdf.

Transformative Urban Mobility Initiative (2022). Sustainable urban transport: avoid-shift-improve (A-S-I). https://www.transformative-mobility.org/assets/publications/ASI_TUMI_SUTP_INUA_No_9_April-2019.pdf.

Transitec (2018). Recording, mapping and analyzing the performance of paratransit services. https://gitlab.com/digitaltransport/data/africa/accra/-/raw/master/papers/0844_170-gui-ssd-2-AM3_guide.pdf?inline=false.

Transport for Cairo (2020). Kampala and Paratransit street usage study. https://transportforcairo.com/2020/09/01/kampala.

United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019, custom data acquired via website. https://population.un.org/wpp/DataQuery.

United Nations Environment Programme (2018). The global electric vehicle policy database. https://www.unep.org/resources/publication/global-electric-vehicle-policy-database.

Watundu, S. (2015). Urban household road travel demand and transport mode choice: the case of kampala, Uganda. http://publication.aeracaffilibary.org/handle/123456/89/2077.

Williams, S., White, A., Waiganjo, P., Orwa, D., and Klopp, J. (2015). The digital matatu project: using cell phones to create an open source data for Nairobi’s semi-formal bus system. J. Transport Geogr. 49, 59–51. https://doi.org/10.1016/j.jtrangeo.2015.10.005.
STAR METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Deposited data      |        |            |
| General Transit Feed Specification (GTFS) data for paratransit systems in 11 African cities. | (DigitalTransport4Africa, 2021b) | https://gitlab.com/digitaltransport/data |
| GPS Tracking data of 10 minibus taxis in Stellenbosch, South Africa. | (Akpa et al., 2016) | https://gitlab.com/eputs/data/fcd-stellenbosch |
| GPS Tracking data of 9 minibus taxis in Kampala, Uganda. | (Ndibatya and Booysen, 2020) | https://doi.org/10.1016/j.jtrangeo.2020.102853 |
| Software and algorithms |        |            |
| EV-Fleet-Sim | (Abraham et al., 2021) | https://ev-fleet-sim.online |

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the Corresponding Author, Prof. Marthinus J. Booysen (mbooysen@sun.ac.za).

Materials availability
This study did not generate new materials.

Data and code availability
- Data: This article used three datasets, listed in the Key resources table. The GTFS dataset and the Stellenbosch GPS dataset are publicly available. The Kampala GPS dataset is obtainable by contacting the lead contact.
- Code: This article uses a software called EV-Fleet-Sim which was built by the authors. The source code and documentation is publicly available through the link provided in the Key resources table.
- Any additional information required to reanalyze the data reported in this article is available from the lead contact on request.

METHOD DETAILS

The method consisted of four steps. First, we evaluated whether the datasets were validly defined according to the GTFS standard. Second, we qualitatively evaluated the reliability of the datasets. This was done by taking into consideration the method used to capture the data, the age of the dataset, the number of data collectors, and the duration that the data collection took place. Third, we ran the datasets through our electric mobility simulation tool. And lastly, we verified the results by checking their correspondence to the input data.

Data retrieval and verification
GTFS datasets were obtained for eleven African cities from the publicly available paratransit data collection (DigitalTransport4Africa, 2021b). The data was then inspected to ensure that the datasets were fully defined according to the GTFS standard. Out of these datasets, four datasets were failed the validation process, as they were missing the datafile which describes the routes followed by the vehicles. The remaining seven datasets successfully described the schedules, routes, and stop locations of minibus taxis.

Data reliability analysis
Although the methods described in this article should be extendable to any other form of public transport that can be described in the GTFS specification, minibus taxis is unique in that it is highly unregulated. Because of this, the compilers of these GTFS datasets were required to do their own data collection to
compile their GTFS datasets. This meant that different approaches were used in this process, each with varying degrees of scientific rigor and reliability.

Therefore, the first step was to inspect the datasets and the methods used to compile them, to evaluate their reliability in a systematic way. As discussed in the introduction, a few methods of paratransit data collection exist. We analyzed publications on each dataset and contacted the dataset compilers to establish the details of the methods used to collect the seven datasets. All datasets were compiled using the approach of deploying human data collectors into the field. Because none existed, we identified quality metrics of how the data was collected and how much effort was put into collecting the data and keeping it up to date. These include:

- Whether data collection was passenger-based or roadside-based:
  - Passenger-based data collection means that the collectors boarded the minibuses as passengers, and recorded data by traveling the various routes.
  - Roadside-based data collection means that the collectors stood at the various terminals, recording the frequency at which taxis arrive and depart.
- Months since last update: Indicates whether the dataset is being kept up to date.
- Number of human data collectors: More data collectors would indicate more effort put into collecting the data.
- Number of weeks of data collection: A longer data collection period would indicate more effort put into collecting the data.

We tabulated these metrics for each of the seven cities for which data collection methodology was known to compare the rigor with which their data was captured. Furthermore, we classified the effort behind each dataset into three categories: high, medium and low. We classified datasets with less than 50 person-weeks as low effort, less than 100 person-weeks as medium effort, and above 100 person-weeks as high effort. Unfortunately, the actual values of the thresholds had to be rather arbitrarily chosen. The values of the person-weeks metric does not have much practical significance because of the fact that the actual data collection effort is dependent on substantially more factors than the number of data collectors and the number of weeks of data collection. However, the person-weeks metric is useful for providing a starting point for classifying the data collection effort, which will ultimately be a qualitative judgment call by the researcher. This is unfortunately a consequence to the fact that the datasets do not provide detailed, uniform information regarding the methods and rigor in which they were collected.

However, to make a fair comparison of the reliability of the datasets, it was also necessary to capture metrics that would illustrate the size of each dataset. A larger dataset would naturally require more effort to capture. Hence, another table was generated with the following metrics to describe the size of each dataset:

- Number of routes: The data capturers would need to make at least one trip on each route of the paratransit system. Hence, more routes there are, the more effort required to complete the data capturing project.
- Number of trips per day: Each day, multiple trips are taken on a particular route. This metric summarizes how many trips are taken per day across all routes of the transport system.
- Geographic spread of the transport system: This is the area in km² of the transport system, as derived from the dataset. This is calculated by creating a bounding box from the minimum and maximum coordinates that are recorded in the transport system’s route definitions, and calculating the area of that bounding box. A larger geographical spread would indicate that the routes are longer. A longer route would probably require more time to capture than a shorter route.

We therefore classified the size of each dataset as small, medium or large. For example, a dataset with many routes and a large geographic area would be classified as “large”. We multiplied the routes with the area (in km²) for each of the datasets. If the resulting value was below 200,000, it was classified as small.
Values below 400,000 were classified as medium, and values above 400,000 were classified as large. Once again, the values of these thresholds are arbitrary.

From these classifications, it was possible to evaluate the reliability of each of the datasets by comparing the effort exerted to the dataset size. If the level of effort was less than the dataset size, then the dataset’s reliability was classified as “low”. If the effort was equivalent to the level of the dataset size, then its reliability was classified as “medium”. If the effort exceeded the level of the dataset size, then its reliability was classified as “high”. For example, a medium sized dataset would have its reliability classified as “low” if the effort was low, “medium” if the effort was medium, and “high” if the effort was high.

Minibus taxi mobility modeling from data

For each of the cities, the following method was used to create a simulation-ready model from the data:

The first step was to obtain geographical data that would describe the road network and terrain that the paratransit system operates in. We were able to utilize a free dataset developed by OpenStreetMap (OpenStreetMap Contributors, 2022). The download server stores the data as one file per country. Therefore, we were forced to download the data for the whole country in which the city lies (Geofabrik, 2022). For each of the routes defined in the GTFS file, we searched for the coordinates of the smallest possible rectangular bounding box which would enclose all routes of that city. We then cropped the geographical data to the bounding box, using the OSMConvert program (marqqs, 2022). This would extract only the appropriate section of the data for our study, greatly reducing simulation overhead. This geographical data was then converted to a simulation-ready road network using the Netconvert program (Lopez et al., 2018).

The software searched this road network to find the exact path (sequence of roads) that the simulated taxi must follow to go from one stop to the next. The location of these stops was extracted from the GTFS file (Andrade et al., 2021). The times and sequences of the stops were also extracted and combined with the paths generated to create route plans, files which direct the movement of the electric vehicle model during the simulation.

The Dijkstra algorithm was used for solving the paths (Lopez et al., 2018). This algorithm chooses between various optimization objectives when solving for a path: time, distance, energy usage, etc. We chose to use distance, as it is the computationally cheapest optimization (Lewis, 2020). However, for more realistic paths, the time objective should be used instead. With the distance objective, the algorithm might choose short paths that go along roads with low speed limits and possible congestion. For example, it might choose a path through the city rather than along the highway. With the time objective, the algorithm might choose the highway option, which is longer in distance but shorter in time. This would be more realistic because the taxi driver would prefer the quicker option.

For each route, the route plan was generated by traversing through the route’s sequence of stops. For each stop, the nearest road on the road network from the stop’s coordinates was found, and the shortest path from the road of the previous stop to the current road is calculated. This path is appended to the route plan being built. The route plan also specifies that the simulated vehicle should stop on the current road until the departure time of the current stop. This process is illustrated in Algorithm 1.

The taxi EV model

With the route plans established, the next step was to define an electric vehicle model that can follow the route plans. The model would need to be configurable and use the route’s distance, road inclination and curvature to evaluate the electric vehicle’s energy usage.

The SUMO software (Lopez et al., 2018) was used for this purpose. The electric vehicle model built into SUMO allows the user to set various parameters. Although the focus of this study is minibus taxis, a common minibus taxi, namely the Toyota Quantum, was used as a basis for the EV parameters. The following model parameters were chosen from the Quantum’s geometry:

Height: 2.3m, width: 1.9m, front-facing surface area: 4 m², weight: 2900 kg.
We approximated the rest of the parameters according to the recommendations by Fridlund and Wilen (2020). These include: Constant power intake: 100W, propulsion efficiency: 0.8, recuperation efficiency: 0.5, roll drag coefficient: 0.01 and radial drag coefficient: 0.5.

The software initialized the EV model for each route that was defined in the GTFS file. The EVs followed the route plans, obeying all speed limits, traffic signals, etc. as defined in the road network. For every second of simulation time, the simulator generated the energy consumption and speed of the EV as it progressed along its route.

The simulation resulted in energy and power usage profiles of the vehicle for each of the routes. The GTFS file defines a frequencies.txt file. This file indicates the frequency at which new trips commence on each route, for various periods of the day. For example, a new trip may commence on a particular route every 30 min from 6 AM to 8 AM, and every 10 minutes from 8 AM to 9 AM. Based on this frequency data, the results were replicated for the trips on each of the routes.

With an average of 45,000 daily minibus taxi trips per city, the volume of output data from the simulation was substantial. Our goal was to obtain useful metrics that would summarize this data. The first metric we considered was the total power usage profile of the whole electric minibus paratransit system. Such a profile would indicate how much power the system would require at various times of the day. This profile would be indispensable, for example, for estimating the expected peak load times and the aggregate daily energy needs of the electric fleet on the grid.

We calculated this profile as follows: First we retrieved the power-vs-time results of each trip, to generate power profiles of each trip. For each route, we aggregated power profiles of the trips done on that route, to get the total power profile of the route. Finally, we aggregated the power profiles of all routes in the taxi system, to get the total power profile of the city’s taxis. The profile was plotted with respect to time.

In addition to the power usage characteristics, the energy usage of the minibus taxi system was also of interest. The power profiles obtained in the previous step were integrated with respect to time, to obtain energy-vs-time profiles. The net energy profile of the minibus taxi system of each city would give an estimate of the total daily energy usage, how much energy was saved because of regenerative braking, and how much energy would be used between times of relative inactivity.

Finally, for each city, boxplots were created to show the median energy usage of the routes, as well as the spread of energy usage across the various routes of the city.
In order to verify the results, we generated a table of the characteristics of paratransit in the various cities. The simulation results were compared to these characteristics to see if they corresponded. We chose characteristics of the city that would indicate the magnitude of its transport needs. The following metrics were chosen for this purpose:

- Number of taxis: A higher number of taxis would indicate that there are probably more trips to be serviced, and hence more energy would be required.
- Number of inhabitants: A city with more inhabitants would have higher transport needs, and, hence, more energy usage could be expected.
- Modal split of minibus taxi in paratransit: If a higher percentage of trips are taken via minibus taxis, as opposed to other modes, the energy usage would be higher.

After further investigation, we found that some of the cities did not have reliable, up-to-date statistics on the number of taxis. We therefore used the remaining two metrics to classify the expected energy demand as high, medium or low. If the product of the population size and the modal split was below a threshold of 1.5 million, we classified the expected energy demand as low. If it was below 2.5 million it was classified as medium, and if was above 2.5 million it was classified as high.