Transportation Research Part D 118 (2023) 103728

Contents lists available at ScienceDirect

Transportation Research Part D

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

Simulating mobility to plan for electric minibus taxis in Sub-Saharan Africa's paratransit

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

[a] Department of E&E Engineering, Stellenbosch University, South Africa
[b] Energy and Power Group at the Engineering Science Department, University of Oxford, United Kingdom

ARTICLE INFO

Keywords:
Paratransit
Electric vehicle
Minibus taxi
Global positioning system data
Energy expenditure
Efficiency

ABSTRACT

Sub-Saharan Africa (SSA) is faced with the challenge to integrate e-mobility into its paratransit (its informal mass transit). Old, unsafe, fuel-inefficient, polluting minibus taxis are the cornerstone of daily commuting for millions in the region. Planning for electrification requires accurate high-frequency mobility data, which is currently unavailable. We analyse and improve on existing simulation models, which predict the energy usage with a micro-traffic simulation, SUMO, that up-samples low-frequency mobility data. We show that, compared to using measured mobility data, the current simulation approach overestimates energy expenditure. Results show a mean energy expenditure per distance overestimation of 14%, and mean energy per trip overestimation of 46%. We identify and virtualisation errors in the current driver and infrastructure models, and identify shortcomings of imposing a virtual road network with simulation software in the SSA context. We recommend and demonstrate virtualisation improvements for accurate electro-mobility planning in future.

1. Introduction

In light of the global drive towards decarbonising transportation and electrifying mobility, high-frequency mobility data is required to assess the impact of electrifying minibus taxis — the backbone of mobility in sub-Saharan Africa (SSA). In the absence of such data, this paper uses the only measured dataset to investigate how mobility simulation models, which are used to estimate energy expenditure in electric vehicles (EVs) on a large scale, can be improved to offer more accurate forecasts for energy system planning in the region.

In this study, the term 'mobility data' is defined as timestamped geo-location (latitude, longitude, altitude, and velocity) data gathered from an on-board tracking device.

On a global scale, the transport sector generates 23% of all energy-related greenhouse gas (GHG) emissions (Sims et al., 2014). As a result, the electrification of transport in the Global North has been growing exponentially over the past decade as new regulations aim to reduce carbon emissions. This directly relates to UN Sustainable Development Goals (SDGs) 7 (affordable and clean energy), 11 (sustainable cities and communities) and 13 (climate action) (Zinkernagel et al., 2018; United Nations, 2015). Internal combustion engine (ICE) vehicles are being phased out, with various European original equipment manufacturers (OEMs) already confirming termination of engine development and announcing full electrical line-ups from as early as 2035 (Motavalli, 2021; Geospatial Commission, 2021; Sunday Times Driving, 2022). The Global South, and Sub-Saharan Africa in particular, has had to play catch-up

* Corresponding authors.
E-mail addresses: christopher.hull@eng.ox.ac.uk (C. Hull), mjbooysen@sun.ac.za (M.J. Booysen).

https://doi.org/10.1016/j.trd.2023.103728
Received 28 October 2022; Received in revised form 23 January 2023; Accepted 28 March 2023
Available online 4 April 2023

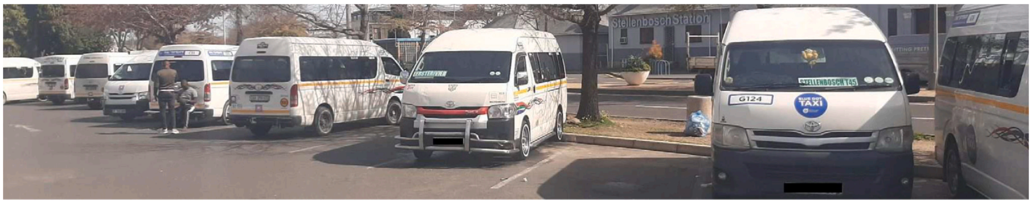


Fig. 1. The Toyota Quantum Ses'fikile, generally used in the South African minibus industry.

once again. To do so, it must plan for large-scale integration of EVs into its massive, informal public transport system, formally known as 'paratransit', which dominates daily commutes in the region. Such planning requires high-frequency vehicle mobility data that can be relied upon to construct high-fidelity estimates of vehicle energy requirements, which are necessary to understand appropriate vehicle and infrastructure design and deployment. As there are no electric vehicles operating in the SSA paratransit industry, no such data is available, and attempts on planning for the electrical transition depends on valid and realistic modelling.

Paratransit is an informal system that differs from how it is known in developed countries. In SSA, it consists of various vehicle modes, such as the minibus taxis in South Africa, Kenya, Nigeria, Uganda and Ghana, tricycle taxis in Nairobi and single passenger motorcycle taxis in the North Eastern part of SSA (Booysen et al., 2013; Diaz Olvera et al., 2019; Mutiso and Behrens, 2011). Their service evolved organically, is demand based, and operates somewhere between private and public transport (Ndibatya and Booysen, 2021; Neumann and Joubert, 2016). Paratransit is used by approximately 98%, 91%, 90% and 70% of daily commuters in Dar es Salaam (Tanzania), Kampala (Uganda), Lagos (Nigeria) and Johannesburg (South Africa), respectively (Behrens et al., 2015; Evans et al., 2018). The most common paratransit vehicle is the minibus taxi, shown in Fig. 1, which completes 83% of the sector's trips and carries from 9 to 16 passengers, depending on the vehicle model (Dorothy et al., 2016; Evans et al., 2018; KCCA, 2016; Bruun et al., 2016). For the aforementioned cities, approximately 72% of daily commuters makes use of the notorious minibus taxi (Behrens et al., 2015). Three regions where minibus taxis are extensively used, include Nigeria (460 000 minibus taxis Cervigni, 2013), South Africa (300 000 minibus taxis (Galuszka et al., 2021)) and Zimbabwe, with 100 000 minibus taxis in their capital, Harare, alone (Mbara, 2002).

These minibus taxis raise serious environmental concerns because the majority of the vehicles are old, under-maintained and fuel inefficient (Smith et al., 2001). The old age of the minibus taxis, often exceeding 20 years, results in higher emissions than are seen from modern ICE vehicles (van Benthem, 2021). The lack of emission regulations for old vehicles means that this industry makes a substantial contribution to the general decline in air quality across African cities (Collett and Hirmer, 2021). On a global scale, SSA was only responsible for 2.3% of carbon dioxide emissions, of which approximately 12% were from the transport industry (Ritche and Roser, 2018; SLOCAT, 2021). However, transport emissions grew by 84% from 2010 to 2016 (SLOCAT, 2021). This will undoubtedly increase due to urban migration, population growth and an expanding middle class (Collett and Hirmer, 2021). This trend is classified as a major threat to human health by the World Health Organisation, who estimated 110,0100 deaths in Africa were as a consequence of air pollution in 2016 alone (World Health Organization et al., 2018; United Nations, 2019). The call for the decarbonisation of the SSA paratransit industry comes from both a human- and planet-health perspective.

In view of the size and significance of paratransit in SSA, and its consequential emissions, the electrification of this sector rates amongst the most important transitions in modern SSA in the fight against climate change. Consequently, this sector now has to make the difficult shift to an electrical energy source (Bruun et al., 2016; Jennings and Behrens, 2017; Pojani and Stead, 2017; Ehebrecht et al., 2018). This difficulty stems from the fact that, despite their prevalence, the energy demand of future electric minibus taxis is still a wildcard.

Meeting the future electrical demand of vehicles will raise complex economic, environmental, societal and energy transition dynamics. Thus, it is crucial to understand how much electricity is needed to satisfy the operations of an electric minibus taxis. These energy demands determine vehicle design and the nature of vehicle interaction with the electricity grid, which can have a resounding impact in an energy-constrained context. Therefore, a robust understanding of the energy demands of electric equivalents of minibus taxis is critical to effectively plan the roll-out and integration of electric minibus taxis in cities. However, the challenge is compounded by the region's scarcity of mobility and operational data (Collett and Hirmer, 2021).

To complicate matters further, energy estimation results from kinetic models using simulated mobility data as input differ substantially from those using measured mobility data. Therefore, there is a need to quantify the electrical charging energy demand, so that plans can be developed to meet the future electrical energy needs of the transport system. Hull et al. (2022b) showed that high-frequency mobility data on the order of one sample per second is needed for high-fidelity estimates of electric minibus taxi energy expenditure. Given the scarcity of such data, and the relative abundance of low-frequency mobility data that only captures one sample per minute (Ebot Eno Akpa et al., 2016), it would be invaluable to have simulation models that can simulate high-frequency mobility data from low-frequency data well enough to obtain high-fidelity estimates of energy expenditure. However, underlying assumptions in models and simulation setups may affect the output. Models use different approaches to simulate and represent mobility in the real world. The main components of these approaches are how they model the physical road infrastructure, how they model the driver and their behaviour, and how they model the vehicle's kinetic and electric energy. This process is illustrated in Fig. 2.

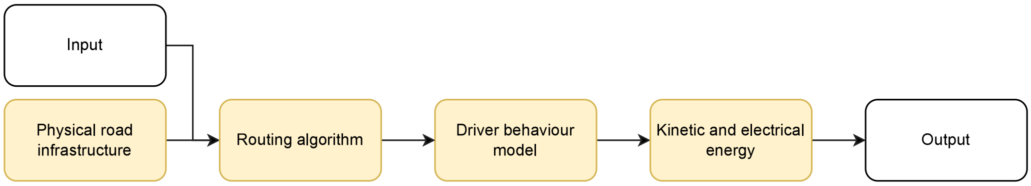


Fig. 2. Main components of mobility and electro-kinetic simulation models.

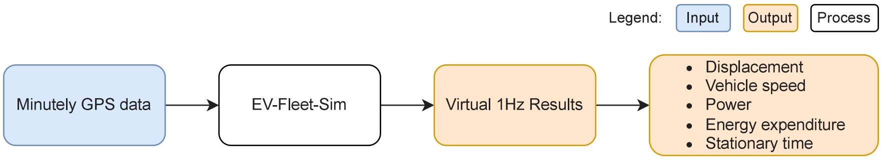


Fig. 3. How EV-Fleet-Sim was used by Abraham et al. (2021).

Results from literature on the quantification of the energy requirements for the electrification of the sub-Saharan Africa paratransit industry thus still vastly differs. This is partly due to different approaches taken, but all comes down to the characterisation of paratransit mobility not being defined. As simulation tools are used across the engineering industry, it is important that the simulated results reflect that of reality.

1.1. Contribution

Here, for the first time, we close the gap between simulation and measurement as far as the micro-level mobility of paratransit is concerned. Because a substantial difference is evident between the energy expenditure estimated by the measured mobility and simulated mobility, we systematically assess the virtualisation of each aspect of in simulation model. We propose and demonstrate improvements to the simulation to align these results, making the simulators accurate and useful in data-scarce environments. Improving micro-level mobility simulations and subsequent energy expenditure estimates will lead to optimal planning for EV transportation systems and infrastructure.

Specifically, we critically investigate the simulation methodology employed by Abraham et al. (2021) - which uses low-frequency (one sample per minute) mobility data, a mobility simulator (SUMO), and an underlying map representation (Open Street Maps, OSM) - to generate high-fidelity mobility data and subsequently estimate energy expenditure. This SUMO-based suite of simulation software used by Abraham et al. (2021) is titled EV-Fleet-Sim (Abraham et al., 2022).

Fundamentally, we assess the virtual (modelled) and actual (real-world) manifestations of each of the following: (1) Physical infrastructure, (2) the following of and routing between waypoints, and (3) the driver and driving style. The subsequent energy expenditure estimates, calculated using the electro-kinetic model by Hull et al. (2022b), are compared using two different inputs: the simulated data from EV-Fleet-Sim and the measured data. The accuracy of the SUMO model used by EV-Fleet-Sim can be assessed on a mobility level by comparing mobility results to the measured 1 Hz dataset; or the utility of estimating vehicle energy expenditure from SUMO based on micro-level mobility simulations can be assessed by comparing the results from a electro-kinetic model.

2. Literature review

In this section, we look at two different energy analyses approaches taken by Abraham et al. (2021) and Hull et al. (2022b). The key difference of their respective methods entails the way in which high-frequency mobility data was acquired for energy analysis. Abraham et al. (2021) used a simulation tool to upsample low-frequency data, whilst Hull et al. (2022b) used measured high-frequency data.

2.1. Simulated micro-level mobility

Abraham et al. (2021) investigated the energy expenditure of minibus taxis using low-frequency mobility data as input to a simulator. They made use of a publicly available software that uses a routing algorithm (SUMO, 2022e), mobility model (SUMO, 2022a) and energy model (Kurczveil et al., 2014) within SUMO to automate energy calculation steps. The software, called EV-Fleet-Sim (Abraham et al., 2022) can be found at https://gitlab.com/eputs/ev-fleet-sim.

The process of how EV-Fleet-Sim was used by Abraham et al. (2021) is shown on a high level in Fig. 3.

For the dataset used by Abraham et al. (2021) an average energy expenditure of 0.93 kWh/km was reported.

EV-Fleet-Sim requires three input files from the user:

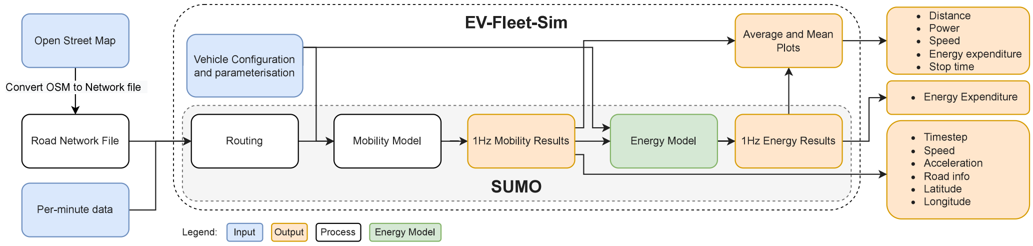


Fig. 4. A flow diagram of EV-Fleet-Sim input and outputs.

Table 1
Default parameters of the driver behaviour model parameters by SUMO as used in EV-Fleet-Sim.

Parameter	SUMO Default
Maximum speed	Road's posted speed limit up to a maximum of 100 km/h
Maximum acceleration	9.3 km/h/s (2.6 m/s^2)
Maximum deceleration	14.4 km/h/s (4.0 m/s^2)
Acceleration profile	Instantaneous step changes in acceleration (Erdmann, 2022)

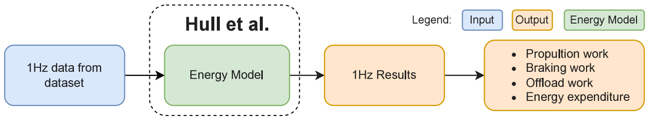


Fig. 5. How energy expenditure estimations are done by Hull et al. (2022b).

1. Per-minute GPS data: Time, speed, latitude and longitude data are required at a minimum. However, various other data fields can also be added (Abraham et al., 2022). The input to EV-Fleet-Sim can be given at any frequency, although Abraham et al. (2021) used one sample per minute for specific route accuracy.

2. Network File: A road network file is used with the GPS data to generate the vehicle routing. The network file can be downloaded from Open Street Map (OSM) (OpenStreetMap Contributors, 2022) for the applicable region of interest. The OSM file contains road network and terrain data. The file can be cropped to a specific geographic area with *OSMConvert* (Weber and Paleino, 2018). Finally, the cropped OSM file is converted to a simulation-ready network file with *Netconvert* (Lopez et al., 2018).

3. Vehicle Configuration and Parameterisation: This file is automatically generated when initialising a scenario in EV-Fleet-Sim, no outside input is needed. The file can however easily be edited to represent a chosen vehicle.

EV-Fleet-Sim uses pre-determined SUMO packages for the routing, mobility model and electric model. However, as stated in Section 3.2, the electro-kinetic model from Hull et al. (2022b) is used in this paper. Thus, SUMO Electric is replaced by Hull et al. (2022b)'s model and used in EV-Fleet-Sim for this paper to provide a fair comparison.

A full breakdown of all the inputs and outputs associated with EV-Fleet-Sim is shown in Fig. 4. In addition, internal automated processes of EV-Fleet-Sim are also shown.

The specific SUMO routing algorithm used in EV-Fleet-Sim is based on distance. It aims to find the shortest path between two data points, as described in by Rix et al. (2022).

Generated outputs from EV-Fleet-Sim are manifold. This is due to SUMO generating its own 1 Hz mobility results (SUMO, 2022d) and EV-Fleet-Sim using those results to generate various plots. Although the end goal remains on the resulting energy expenditure, driving style outputs such as speed and acceleration will assist in assessing the accuracy of the mobility model.

Although SUMO allows various driving style parameters to be defined, such as maximum acceleration, a speed factor of the road's speed limit, and the harshness of the accelerations; EV-Fleet-Sim currently only makes use of default parameters set by SUMO. These driving style parameters are summarised in Table 1.

It stands to reason that the driving style of the virtual vehicle should also be affected by the presence of traffic, as in real life. A car following model, which forms part of the mobility model, is implemented by default in SUMO. However, currently no other vehicles than the simulated electric minibus taxi are present in the simulation.

2.2. Measured micro mobility

Hull et al. (2022b) estimates the energy expenditure of an electric minibus taxi by using 1 Hz measured tracking data as input into their electro-kinetic model. The limited variety of different data types needed for their input, makes it quite appealing. All the required information is captured by a GPS receiver.

A high level of the process to construct energy expenditure estimations is shown in Fig. 5.

Although micro-level mobility data proves to be an advantage for accuracy, the scarcity of 1 Hz data in the minibus taxi industry of SSA, and the cost of generating and storing such data, means that their energy model cannot be widely used without intermediate data processing. In cases where data at a lower frequency, or only O+D data is available, a routing algorithm and driver model is needed to compute estimations of 1 Hz mobility.

2.3. Simulated inaccuracies

When comparing the simulated micro-level mobility to the measured micro-level mobility data, and the subsequent energy analysis results, it is clear that the simulated mobility data does not accurately represent that of real life mobility of the paratransit industry. Simulated mobility incorporates various aspects, as expanded on in Section 3.1. These aspects are investigated and quantified in terms of energy expenditure error, by comparing results by Abraham et al. (2021) and Hull et al. (2022b) using the same dataset. Through this, we aim to close the gap between their results to improve on existing simulation tools.

3. Method

We systematically analyse the simulator by using as input a per-minute downsampled per-minute version of the 1 Hz GPS data by Hull et al. (2022a). We compare the simulated 1 Hz output to the measured mobility data. The high-frequency simulated and measured mobility data are separately used as inputs to the electro-kinetic model by Hull et al. (2022b) to evaluate the difference in energy expenditure between the real-world and virtual mobility data.

The assessment separately evaluates the accuracy of the virtualisation of (1) the infrastructure, (2) the vehicle and (3) the driver, thereby encapsulating the mobility aspects discussed in Section 3.1; and is compared to the measured dataset.

For each of the aspects we identified as problematic, we evaluate the impact of modifying the simulation suite accordingly, and compare each individual modification's results to the measured mobility where applicable. The impact is quantified in terms of energy expenditure per distance by using the electro-kinetic model by Hull et al. (2022b).

3.1. Mobility

Mobility has a direct bearing on the energy conversion in a vehicle from electric to kinetic and vice versa. Insight into the energy requirements for paratransit therefore necessitates an understanding of its unique mobility patterns on a high-frequency level. Fundamentally, the question of mobility is one of space and time. We analyse the mobility on two levels: **Routing:** Where these taxis go to and from, and which route they take, and **Driving style:** The way in which they move along the route from origin to destination.

3.1.1. Routing

Unlike the scheduled transit in the developed world, minibus taxis follow demand-driven schedules — often departing when full, stopping at ad-hoc locations to drop off and collect passengers on the way to the destination (Ndibatya and Booysen, 2021; Zeeman and Booysen, 2014). In some conditions, mostly in cities, minibus taxis are issued with permits to follow only certain routes. However, they often veer from regular and permitted routes to, for example, avoid traffic and to drop off passengers along the way (Neumann and Joubert, 2016). Ndibatya and Booysen (2021) shows their trajectories tend to follow a heavy-tailed power-law distribution, similar to a "Lévy walk", in their seemingly random search for passengers. Since commuters predominantly travel to and from work in urban areas, traffic peaks are apparent from 06:00 to 09:00 and from 15:00 to 18:00 (Abraham et al., 2021).

To accurately simulate the mobility patterns of minibus taxis, and thus obtain high-fidelity estimates of energy expenditure, comprehensive and representative spatio-temporal data is needed. Since the drivers mostly go where they please and do so unscheduled, this data has to be captured by tracking the vehicles. Although minutely sampled mobility data tends to be stored as part of some research projects (e.g., by Abraham et al., 2021 and Ndibatya et al., 2016), the norm in vehicle tracking applications to only store the Origin and Destination (O+D) data to reduce the cost of transmission and storage. To get routing information from O+D data, in the absence of GPS waypoints, requires a routing algorithm. In this sense, a waypoint refers to a single sample of input data. This could be done with simulation software such as the routing algorithm provided with SUMO, a multi-modal traffic simulator (SUMO, 2022c).

3.1.2. Driving style

A key driver of an EV's energy expenditure, and therefore efficiency, is the way in which the vehicle is driven. An aggressive driving style may result in rapid changes in location, speed and/or acceleration, resulting in specific high-frequency movement characteristics of the vehicle (Al-Wreikat et al., 2021; Faria et al., 2019; Murphey et al., 2009; Ebot Eno Akpa et al., 2019). Both intense accelerating and braking result in the loss of effective energy (e.g. heat dissipation). Given these losses, each instance of accelerating and braking results in inefficiencies. Moreover the movement of the vehicle through the natural environment also affects the energy balance between the battery and its mobility. For example, driving downhill vs. uphill, or with the direction of the wind.

To record the nature of these rapid changes in kinetic energy and adequately estimate the resulting energy losses, a correspondingly quick rate of data capture is required (Hull et al., 2022b; Shannon, 1949). This quick rate of capture is especially important in the context of minibus taxis, which notoriously engage in aggressive driving that involves rapid and quickly changing acceleration and deceleration events (Zeeman and Booysen, 2014). Hull et al. (2022b) have shown that only 56% of the energy inherent in their mobility patterns is captured when using per-minute rather than per-second mobility measurements (Hull et al., 2022b).

To obtain high-fidelity estimates of energy usage from GPS locations requires high-frequency data (Hull et al., 2022b). However, this requires a lot of data to be transmitted and stored, such as the 1 Hz GPS data used by Hull et al. (2022b), making high-frequency data measurement impractical on a large scale. In the absence of such measured data, a micro-traffic simulator on a virtual map, for example SUMO on Open Street Map, can be used to interpolate high-frequency data in between waypoints — an approach used by Abraham et al. (2021).

3.2. Electro-kinetic energy models to link mobility to energy in EVs

Electro-kinetic models use the movement characteristics of a vehicle through space and time to estimate the energy required to perform those movements. For EV modelling they therefore act as the bridge between kinetic energy and electrical energy stored in the battery. These models use mobility data with parameterised models of the propulsion and braking systems and virtual representations of the environment to estimate the efficiencies of converting energy from the battery into displacement and from movement into the battery when braking.

In addition to micro-level mobility information, vehicle parameterisation plays an important part in energy estimation. These parameters include vehicle mass, dimensions, powertrain efficiencies and resistance coefficients amongst other things. The specific integration of each vehicle parameter depends on the configuration of the kinetic energy model. The fundamental result obtained from an energy model is the energy required per movement, i.e. energy per distance. In this paper, the unit of energy per distance is chosen as kWh/km.

Various energy models exist in the literature that can be used to estimate the energy expenditure of a vehicle (Wu et al., 2015; Iora and Tribioli, 2019; Kurczveil et al., 2014; Hull et al., 2022b). Models differ in terms of the required input data and the sampling frequency thereof. Each model uses a unique set of physics-based formulas that determines the required energy to move between two data location samples in the sample period. Some models only determine the kinetic energy required at the wheels of the vehicle, while others transcend the kinetic domain into the internal power electronics of the vehicle. This is dependent on the type of vehicle used for data collection and sensors fitted to said vehicle.

In this paper we use the simple physics-based model proposed by Hull et al. (2022b).

3.3. Input data

A dataset from Hull et al. (2022a) is used as input to EV-Fleet-Sim. It consists of 1 Hz data captured by six tracking devices used by passengers on pre-selected routes in and around Stellenbosch, South Africa.

These routes were chosen to represent urban, uphill and downhill, and inter-city travel, having speed limits of 60 km/h, 80 km/h, and 100 km/h respectively. Trip distances ranged between 2.6 km and 23 km, with trip time ranging from 4 min to 27 min. As further described by Hull et al. (2022b), the bulk of the data was collected on a single day and it does not account for variation that may occur in the driving patterns of minibus taxis over

In total, 62 trips were recorded across the different routes at different times of the day (Hull et al., 2022b).

Fig. 6 compares the results yielded by different methodologies used for estimating paratransit energy. The 1 Hz GPS data from Hull et al. (2022a), and 1 Hz simulated mobility output data is used as input to the electro-kinetic model presented by Hull et al. (2022b). As described in Section 2.1, the data from the dataset is downsampled to one sample per minute when used as input to EV-Fleet-Sim, which in turns generates simulated 1 Hz mobility data. Other energy expenditure figures from Collett et al. (2021), Miri et al. (2020) and Abraham et al. (2021) are also included, along with the manufacturer stated energy consumption of several similar real world vehicles. The constant values from Collett et al. (2021) and Miri et al. (2020) are not generated from the input dataset, but rather are gathered from the mean energy expenditures suggested in their respective studies.

Comparing the results yielded from using simulated data input to measured data input, it seems simulated mobility data leads to an overestimation of energy expenditure. This is emphasised by the even larger overestimation by Abraham et al. (2021), who also made use of EV-Fleet-Sim; with the only difference being that they used the electro-kinetic model by Kurczveil et al. (2014). The differences in energy expenditure estimations caused by the different electro-kinetic models is scope for future work.

The difference in energy expenditure between using simulated and measured data is also shown to fluctuate according to route type. This renders making use of a constant ratio to translate energy expenditure from simulated data to measured data insufficient; and thus an in depth analysis is needed.

The question is thus raised, how much of the energy expenditure difference is due to the difference in physical vs virtual road infrastructure, the routing and waypoint progression of SUMO or the virtual driver process. Furthermore, can these differences be eradicated to generate high-frequency mobility data that is accurate enough to generate high-fidelity energy expenditure estimates in line with Hull et al. (2022b)?

3.4. Systematic analysis and improvement of the simulation model

We analyse the accuracy of the virtualisation of SUMO as used by EV-Fleet-Sim based on the three aspects stated in Section 1.1: transport infrastructure, waypoint following and driving style.

Firstly, the downloaded Open Street Map (OSM) file is investigated. Although the OSM file is downloaded from a publicly available data source (OpenStreetMap contributors, 2017), it is modifiable. Once the file is converted to a network file format, it can be opened with SUMO NetEdit, a graphical user interface to view network files (SUMO, 2022b). Using this programme,

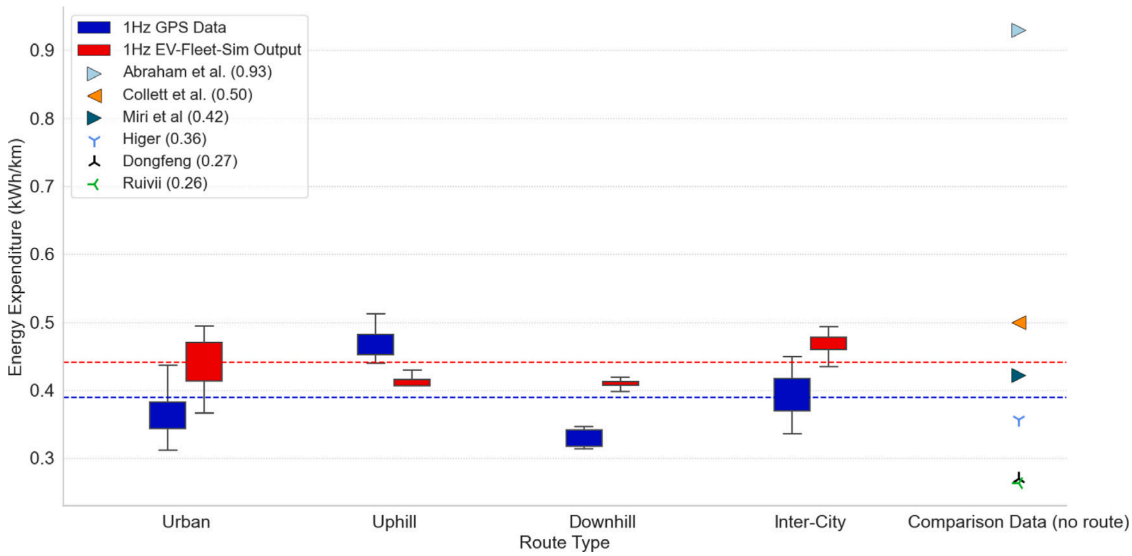


Fig. 6. Distributions of energy expenditure (kWh/km), compared to state-of-the-art from literature (Miri et al., 2020; Collett et al., 2021; Abraham et al., 2021) and real world vehicle specs (Ruivii, 2019; Dongfeng, 2021; Higer, 2020), and dis-aggregated by route type. The blue and red dashed lines represent the mean of the estimates constructed using measured data (0.39 kWh/km) and SUMO mobility output (0.44 kWh/km) as inputs into Hull et al.'s (2022b) model respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

allows information such as the completeness of the road network, road parameters, GPS information and traffic control features to be viewed and edited. The *Open Street Map* and *Road Network File* blocks, as presented in Fig. 4, are analysed and modified in this investigation.

Waypoint progression done in SUMO is analysed based on the virtual progression along the waypoints versus the actual measured route advancement. It is investigated by observing the virtual vehicle's movement in the SUMO graphical user interface and comparing it to the recorded movement in the dataset. The *Routing* block presented in Fig. 4 is analysed in this investigation.

Finally, we assess mobility observed in the virtual vehicle of SUMO. The method of this examination consists of manually editing the *Vehicle Configuration and parameterisation* and *Road Network File* files which affects the *Mobility Model* process. Here we do a mobility deep dive to assess some inaccuracies from what is observed from the measured data. The configuration and parameterisation of the vehicle is defined in a XML file, which can easily be edited. Speed and acceleration limits can be set here, along with other parameters which influences mobility. Various SUMO packages are used to form the base of the driver model (SUMO, 2022a). The aim is to see a more accurate representation of the driving style which is observed in the measured dataset.

4. Discussion of results

The results are presented according to the virtualisation of the physical infrastructure, virtualisation of the waypoints, and the virtualisation of the driver.

Table 2 previews these results, and summarises the impact of each individual aspect. The baseline impact is assessed based on the mean change in energy expenditure per distance and trip, obtained from using simulated mobility data (SUMO, 2022d) as input to the electro-kinetic model by Hull et al. (2022b); and compared to that from using measured mobility data (Hull et al., 2022a).

The measured 1 Hz dataset is taken as ground truth for the high-frequency data source. As previously mentioned, the data used as input to EV-Fleet-Sim to generate simulated 1 Hz (SUMO, 2022d) data was generated by downsampling the 1 Hz measured dataset by Hull et al. (2022a) to one sample per minute, thereby providing a like-for-like comparison.

The modifications aim to align the energy expenditure estimates constructed from SUMO's simulated high-frequency data to that what Hull et al. (2022b) reports when using the measured data. We aim to identify the contributing factors to the difference between the ground truth dataset and the simulation, and to quantify each factors' contribution in terms of energy expenditure per distance [kWh/km]. Where feasible, the virtualisation will be modified for improved accuracy.

Where energy expenditure difference is not reported in the table, it had no impact on the mean figure. The reader is referred to the individual sections for the reasoning.

Waypoint progression and stop sign adherence error corrections are emitted from the aforementioned results as it is only estimated based on energy per distance, not energy per trip.

Table 2
Summary of modifications to EV-Fleet-Sim's mean energy expenditure.

Electro-kinetic model	HF data source	Energy expend./km [kWh/km]	Change in expend./km [kWh/km]	Change in expend./trip [%]
Hull et al. (2022b)	Measured (Hull et al., 2022a)	0.387	–	–
Hull et al. (2022b)	Simulated (SUMO, 2022d)	0.441	–	–
Impact of improving virtualisation (simulation)				
Physical infrastructure	† Elevation	0.458	+0.017 (+4%)	+7%
	‡ Missing roads[a]	–	–	−3%[b]
	‡ Network parameters	0.393	−0.048 (−11%)	−4%
	• Traffic control features	–	–	–
Waypoints	– Waypoint progression[c]	0.333	−0.108 (−24%)	–
	• Reverse geocoding	–	–	−35%[d]
Virtual driver	† Speed profile	0.405	−0.036 (−8%)	−1%
	† Acceleration profile	0.393	−0.048 (−11%)	−4%
	– Stop sign adherence[c]	0.442	+0.001 (+0%)	–

† Modification integrated with EV-Fleet-Sim.
‡ Manual intervention needed for modification.
– Impact on energy expenditure of modification estimated.
• No energy expenditure changes are made in these sections.
[a]As there is only one missing road in the area of the dataset used, the impact on the mean energy expenditure over all trips is negligible.
[b]Change only reported for the Urban Route 1 route dataset as it is the only route where missing roads were found.
[c]Changes in energy/trip are not applicable as the impact was only analysed on energy per distance base, not energy per trip.
[d]Change in energy/trip is only for a single fixed trip as discussed in Section 4.2.2.

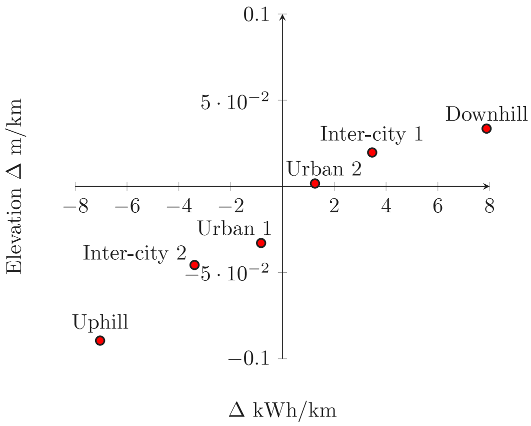


Fig. 7. Change in elevation per distance, versus change in energy expenditure per distance.

4.1. Physical infrastructure vs virtual infrastructure

In this section, various aspects of real life physical infrastructure are compared with the virtual infrastructure seen in the SUMO environment as defined by the OSM file. The effect of these observed differences on energy expenditure are quantified. Specifically, the effects of omitting elevation, having an incomplete road network, inaccurate road parameters, and inaccurate traffic control features on energy expenditure estimations are evaluated.

4.1.1. Elevation

Open Street Map (OSM) is a geographic database of the world used to create the network file in the simulation. However, elevation information is not standard information contained in an OSM dataset. Data nodes used in OSM are defined solely by an ID, latitude and longitude. This allows for a complete 2D picture of a chosen area. However, SUMO can also take elevation into account.

OSM data can be processed with OpenStreetMap (2022), a command line Java application. By using an *Osmosis* plugin, Nasa SRTM data (NASA, 2015) is used to overlay elevation information. The *Osmosis-srtm-plugin* (Graf et al., 2015) is used for this purpose. This will add a height tag to each note. The result of overlaying elevation data on the OSM file increased the mean energy expenditure per distance by 0.071 kWh/km to 0.458 kWh/km. The individual route impact of adding elevation in EV-Fleet-Sim is shown in Fig. 7.

(a) Actual route taken from Bassi to Setona Street (b) Road network file of Bassi and Setona Street

(c) As seen from Bassi Road (d) As seen from Setona Street

Fig. 8. (a) and (b) Missing road from OSM data used by the taxi. (c) and (d) Google street view of route taken from Bassi to Setona Street (Google, 2015).

The increase or decrease of mean energy expenditure correlates as expected for routes with a net elevation gain or loss; suggesting an improvement of the real world accuracy of SUMO's simulated data output. It is recommended that elevation is always included in EV-Fleet-Sim for more accurate physical infrastructure representation.

4.1.2. Missing roads

Movement in the SUMO model is restricted to road-only movements as defined in the OSM file. If a minibus taxi takes a shortcut over land which is not recognised on OSM as a road, it will severely impact the routing algorithm.

Even though the areas in which this study is conveyed has accurate coverage of the road network, it cannot account for all the routes taken by minibus taxis, especially in informal settlements. Editing the road network file, created from the OSM file, allows the virtual vehicle to access service and gravel roads in the SUMO simulation. Although this allows them more freedom of movement than a normal vehicle in the simulation, it still does not cover all the routine routes used by minibus taxis.

An example of this is seen along Urban Route 1. Between Bassi and Setona Street a throughway over an empty plot is used, as shown in Fig. 8.

From what is seen on Google Street View (Google, 2015) in Figs. 8c and 8d and the recorded trips, this throughway is used repetitively and unofficially part of the Urban Route 1 minibus taxi route.

Even though OSM data includes dirt and gravel roads, this throughway is not a real road. Thus, routing algorithms will have to determine alternative routes to complete routing between waypoints in Bassi and Setona Street, which is far removed from the route taken in reality.

The only solution for this problem is manual intervention, by visually investigating the route taken by a minibus taxi. Doing this with minutely sampled data could still lead to missing roads used by minibus taxis not being found; thus 1 Hz measured data is crucial to this investigation.

The aforementioned missing road on Urban Route 1 is the only shortcomings found in the area which covers the dataset used for this study. Thus, the impact on the mean energy expenditure is negligible over the 62 trips and 6 routes. However, manually adding the missing road in NetEdit reduced the mean result for trip distance of the Urban Route 1 dataset by 4% and subsequent energy expenditure by 3%.

Although only one road was unrepresented in the area which covers this dataset, the extent of this issue in the context of SSA remains unknown. Picking up on this error would require high-frequency measured data and manual inspection of all routes taken.

Although the energy expenditure per distance [kWh/km] is not affected by this shortcoming, the total distance and subsequent total energy requirement for minibus taxi operations will be overestimated due to missing roads. This is a flaw which simulations alone, including EV-Fleet-Sim, are not adequate for addressing.

4.1.3. Network parameters

In SUMO, an edge represents a single lane connecting two intersections; or a collection of side-by-side lanes, travelling in the same direction between two intersections. Each edge has a set of parameters as defined in the OSM dataset. Edge parameters include:

1. **Direction:** Defined by identifying the two intersections (nodes) connected by the edge, and labelling them as 'to' and 'from'.
2. **Type:** Defines if the road is public or service, gravel or tarred, residential or tertiary etc.
3. **Speed Limit:** The legal speed limit of the edge.
4. **Allowed and Disallowed:** Specific vehicle types can be allowed or disallowed on the edge.
5. **Length:** The total distance which a vehicle will travel on the edge.

For the area covered by the dataset used in this study, it is found that the **Direction** and **Type** parameters are precisely accurate.

The **Speed limit** parameter is found to be mostly accurate. Some of the larger roads in Stellenbosch were found to overestimate the speed limit by 20 km/h. Other inaccuracies were found when a regional road crosses a traffic light intersection, or passes near a residential area. The actual speed limit reduction to 80 km/h or 60 km/h is not seen in the OSM dataset. In the dataset, the entire regional road has the highest speed limit of 100 km/h.

Energy expenditure per distance reduces to 0.393 kWh/km when fixing all speed limits of roads used in this dataset, as shown in Table 2. Note that this required manual intervention by searching through each edge's parameters in NetEdit and fixing the speed limit where needed. Thus, this modification is not uploaded to EV-Fleet-Sim and only shows the impact of incorrect speed limits on energy expenditure.

The type of vehicles that are **Allowed and Disallowed** on certain edges are found to be needing manual intervention too. In the context of a first world country, the road network is nearly entirely tarred. However, gravel roads forming part of daily commute is ordinary in SSA. Thus, gravel roads are manually set to *Allow* the taxi vehicle class using NetEdit. In addition, service roads are also set to *Allow* taxis, as the roads in the minibus taxi ranks are defined as such. This manual intervention is already part of the standard EV-Fleet-Sim and thus the impact of this modification is not analysed.

The accuracy of the **Length** parameter is not investigated in this study.

Intersection parameters dictate lane priority. Each lane has parameters in relation to other lanes in the intersection, which includes: lanes that it must yield for; lanes over which it has right of way; non-conflicting lanes; crossing lanes where traffic lights prevent vehicle collision.

The accuracy of the intersection's parameters are not analysed in this study.

4.1.4. Traffic control features

In addition to edges and intersections, various traffic control features are also present in the network file. This includes traffic signals, foot crossings and roundabouts.

The presence of pedestrian and vehicle traffic in the simulation dictate the effect of foot crossings and roundabouts on the virtual vehicle respectively. But as there are no traffic by default in the simulation model, the influence of these traffic control features are irrelevant for this paper.

Traffic signals are on a fixed 90 s cycle by default, irrespective of traffic. They operate on 90 s cycles (SUMO, 2022f).

One clear shortcoming in the virtual network is the omission of stop signs. The lack of stops signs in the virtual infrastructure is further investigated in Section 4.3.3, where the stop sign adherence of the actual vehicle in the measured dataset is analysed.

4.2. Waypoints in the real world vs. waypoints on a virtual map

In this section, waypoint progression and accuracy is analysed. It is compared to the measured mobility in the dataset and divergences are quantified in terms of energy expenditure and distance. Specifically, we look at how the vehicle advances from one waypoint to the following, and the accuracy of the waypoint on the virtual map versus the actual waypoint.

4.2.1. Waypoint progression

As previously described, a SUMO routing algorithm based on the shortest distance between waypoints is implemented in EV-Fleet-Sim. However, what happens after the virtual vehicle crossed each waypoint in SUMO, has not been assessed. The movement of the virtual vehicle can be observed in the SUMO graphical user interface, which allows the waypoint progression to be analysed as shown in Fig. 9.

To quantify the number of stops, the 1 Hz mobility output from SUMO is compared to the measured dataset (Hull et al., 2022a) and the number of stops counted. A complete stop is defined as an observed speed of less than 2.778 m/s or 10 km/h for more than one consecutive sample. An upper threshold of 15 km/h or 4.167 m/s is also implemented. Thus, a new stop can only be logged once this threshold is exceeded. These thresholds were determined with a control test by using a GPS receiver, where complete stops were performed and GPS speed analysed.

It is seen that the number of simulated stops greatly outnumber the measured stops on the Inter-city, Uphill and Downhill trips . The relatively high speed-limit on these routes, combined with the excessive number of stops is an indication of why energy

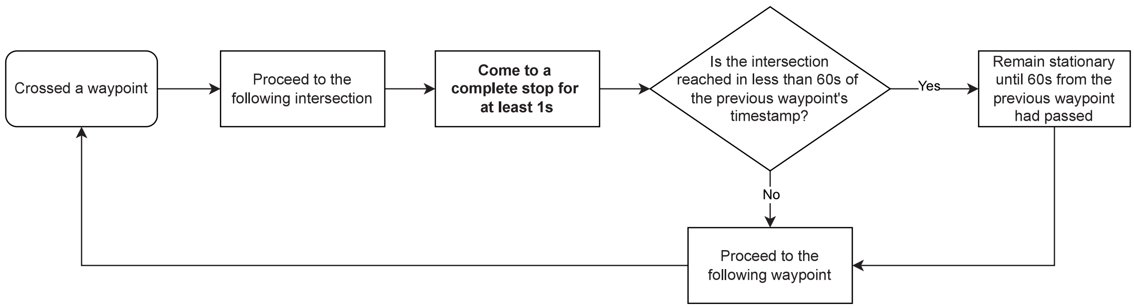


Fig. 9. Waypoint progression of the routing model used by EV-Fleet-Sim.

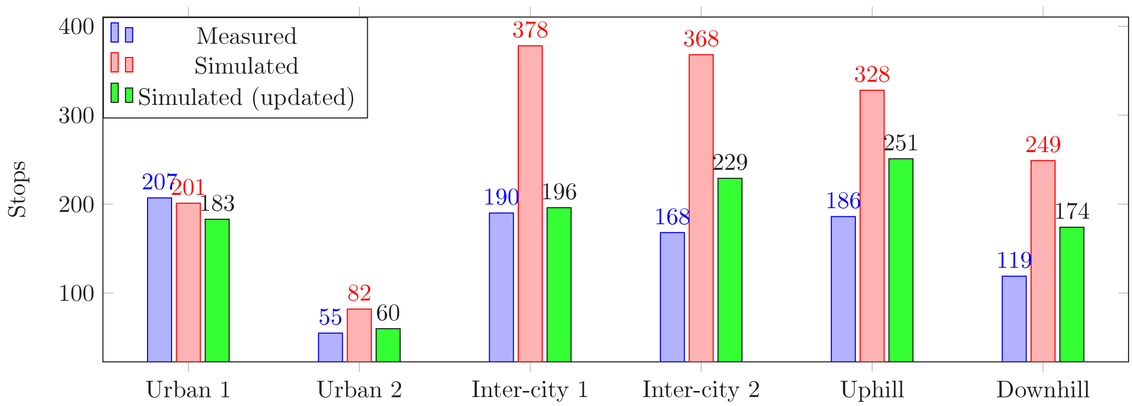


Fig. 10. Total stops observed from recorded dataset, Simulated and Simulated without the spurious high-speed stops that result from the waypoint progression method.

expenditure using simulated mobility data results in an overestimation. This is due to extreme energy losses, or low energy efficiency, associated with repetitive breaking and acceleration, rather than staying at a constant velocity.

On roads with a speed-limits above 50 km/h, traffic lights are the only traffic control feature that incite a vehicle to stop. All other stops observed in simulation from a speed of above 50 km/h is thus due to the waypoint progression of SUMO. To quantify the over stoppage due to waypoint progression, stops occurring from a speed higher than 50 km/h, excluding traffic light stoppages, are removed. The results are shown in Fig. 10 and are labelled as “Simulated (updated)”.

Much greater stoppage correlation is seen after high-speed stops due to waypoint progression is removed. To estimate the energy expenditure related to the over stoppage in the simulation model, the energy output from using the simulated mobility data is analysed.

Each trip is individually investigated where the removed high-speed stops are logged. The distance covered and energy expended during this stop are removed from the trip's results file. Through this, a new energy expenditure per distance for each trip is estimated as if these high-speed stops due to waypoint progression did not occur. This reduced the mean energy expenditure to 0.333 kWh/km, as shown in Table 2.

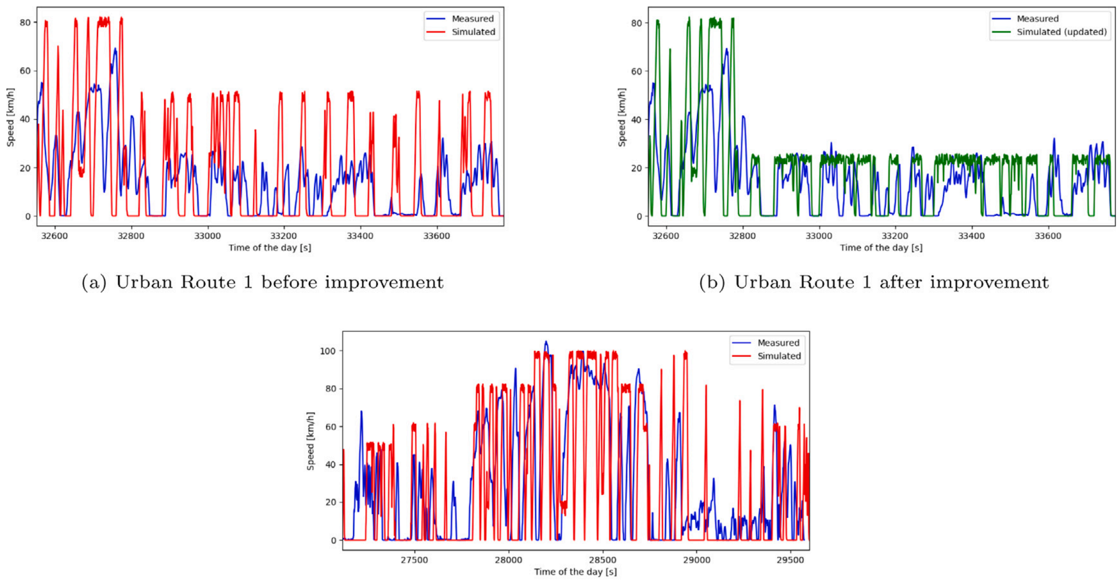
4.2.2. Reverse geocoding

Current GPS technology allows for about 4.9 m of accuracy, which decreases when travelling at high speeds (Hanacek, 2021; Rahemi and Mosavi, 2021). Inaccurate readings of only 4.9 m already causes great trouble when using a routing algorithm.

When waypoints are given to a routing algorithm, they are latched to the closest edge. Inaccurate waypoints could thus easily latch onto a wrong edge, as no other input data is used when creating a route structure. As the waypoints are routed according to their timestamp, one inaccurate waypoint will lead to an incorrect route and result in an over-estimation of route distance and thus total energy.

One such example is seen on a trip from the Inter-city 2 dataset. Here, an inaccurate GPS reading resulted in a total trip distance of 30.5 km. The measured distance for the same trip totals 20.6 km. The number of possible inaccurate GPS readings is directly proportional to the number of input waypoints. To put this into context, using only origin and destination waypoints for routing of this same trip results in a trip distance of 19 km. This gives reassurance that the routing done by SUMO is in fact accurate and not the origin of this problem.

After manually editing the faulty waypoint, the total trip distance reduced to 20.9 km and trip energy by 35%, as shown in Table 2. However, correcting faulty waypoints only changed the energy expenditure per distance by 0.4%. Thus, the reverse



(a) Urban Route 1 before improvement

(b) Urban Route 1 after improvement

(c) Inter-city 1

Fig. 11. Example of speed profiles before and after improvement in speed profile for the urban environment.

geocoding error only impacted the total energy per trip substantially, and not the energy per distance. Other comparisons made in this paper based on kWh/km were not impacted by this error.

In addition, travelling on narrow roads often requires the minibus taxi to drive on the other side of the road. Even accurate GPS readings will then lead to inaccurate routing, as the edge upon which the waypoint latches has a parameter dictating the traffic flow direction.

One could argue that origin and destination data could in this case solve the issue. However, as discussed in Section 3.1.1, it is shown that minibus taxis do not always follow the shortest distance between origin and destination (Neumann and Joubert, 2016; Ndibatya and Booysen, 2021), thereby avoiding traffic and passenger demands are prioritised.

Thus, the one sample per minute approach used by Abraham et al. (2021) is seen as the best compromise to capture the actual route. However, this is still a substantial methodological error, as mean trip distance and subsequent mean energy expenditure per route over the entire dataset is over-estimated by 27% and 46% respectively.

4.3. Virtual driver

In this section, the behaviour of the virtual driver is assessed. Specifically, we compare the speed profile, acceleration profile and stop sign adherence to the measured dataset. The impact of virtual driver inaccuracies is quantified in terms of energy expenditure and compared to Hull et al. (2022b) using measured data in Table 2.

4.3.1. Speed profile

The speed of the virtual vehicle is determined by the SUMO driver model and the speed limit of the edge on which it travels. The simulation model uses the measured speed of the actual vehicle only as zero or non-zero. Thus, the input speed is used solely to determine whether the vehicle is stationary or not. The exact speed of the measured waypoint is not considered in the virtual vehicle mobility.

Upon viewing the virtual vehicle's mobility in the SUMO GUI, the following observations are made: The vehicle will remain at the speed limit all the time. This includes crossing any intersection apart from when turning, encountering a red light or stopping after crossing a waypoint (as described in Section 4.2.1). When entering an edge with a lower speed limit, the virtual vehicle's speed will reduce before entering the reduced speed zone area.

To assess the accuracy of the virtual vehicle's speed in comparison to the actual speed, the speed profiles of several trips are plotted against each other.

The speed profile of a randomly chosen trip from the Urban Route 1 dataset is plotted in Fig. 11(a).

A great difference in vehicle speed is seen throughout the entire trip in Fig. 11(a). The low correlation in speed is emphasised in the latter part of the trip, which takes place on residential roads. Even though the speed limit in this part of this route is accurate, the simulation clearly does not accurately represent the actual speed of the minibus taxi. The beginning section of the trip, where the simulated top speed is seen as roughly 80 km/h, features an incorrect edge speed limit. The accuracy of this edge parameter

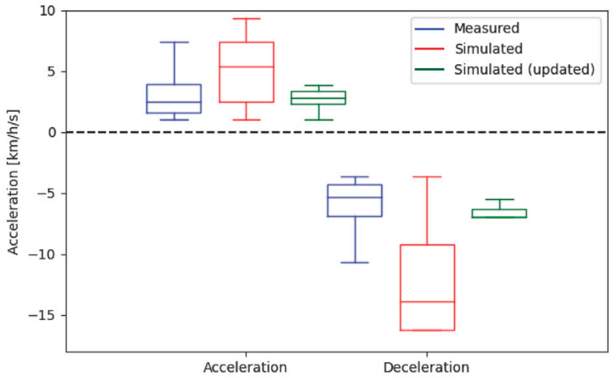


Fig. 12. Acceleration and deceleration observed with updated simulation parameters.

is fixed as part of Section 4.1.3. To investigate whether this is just the case for the urban route, the speed profile of a randomly chosen trip from the Inter-city 1 dataset is plotted in Fig. 11(c).

Apart from the excessive number of virtual vehicle stops and assumed real world traffic in the latter part of the trip, there is a strong correlation between the speed profiles. This reinforces some confidence in the speed limits defined in the OSM dataset.

From Fig. 11(a) it can be concluded that the speed limit of roads in residential areas are not an accurate representation of the actual vehicle speed. Inter-city roads however show a more accurate representation. The lack of the presence of traffic in SUMO undoubtedly also plays a role in the virtual vehicle's speed.

With the NetEdit application, residential roads can be selected with their *type* parameter: *highway.residential*. The speed limit thereof can be manually modified. According to the measured dataset, an average top speed of 25 km/h is seen on residential roads. After implementing this speed limit change on residential roads on the network file, the new speed profile of the same trip as in Fig. 11(a) is shown in Fig. 11(b).

Mean energy expenditure reduced to 0.405 kWh/km when setting the speed limit of all residential roads to 25 km/h, as shown in Table 2.

4.3.2. Acceleration profile

Acceleration plays a significant role in the energy consumption of any vehicle. High acceleration and deceleration results in a reduction of the efficiency of the vehicle, thus increasing the energy expenditure.

The SUMO Driver Model used in EV-Fleet-Sim shows to be a much more aggressive driving style than what is measured, with mean values for acceleration and deceleration of 5.1 km/h/s and 12.2 km/h/s respectively. Comparing this to the measured mean acceleration of 3.2 km/h/s and mean deceleration of 6.0 km/h/s, it is clear that this is a big element in the overestimation of energy expenditure in the simulation. In addition, alternating infinite and zero jerk (change of acceleration) are observed in the simulated data.

As part of the input vehicle configuration, as shown in Fig. 4, a maximum acceleration and deceleration can be set. By default, the maximum values are set to 9.36 km/h/s for acceleration and 14.4 km/h/s for deceleration. For this modification, the limits in the simulation parameters are set according to the measured upper Q3 acceleration and lower Q1 deceleration: 3.9 km/h/s and 6.9 km/h/s.

The effect of modifying the acceleration and deceleration parameters is shown in the box plot of Fig. 12, excluding outliers. To eliminate low acceleration observed at constant speeds, a threshold of +1 km/h/s and -3.6 km/h/s is implemented.

The mean acceleration and deceleration in the simulation reduced to 2.82 km/h/s and 6.5 km/h/s respectively. This reduced the mean energy expenditure to 0.393 kWh/km, as shown in Table 2.

4.3.3. Stop sign adherence

As stated in Section 4.1.4, the network file used in the simulation model does not support stop signs. To investigate the extent of the impact of this shortcoming in the virtual infrastructure, we need to understand the stop sign adherence of a real-world minibus taxi.

To investigate the stop adherence of the actual vehicle in the measured dataset, Table 3 is drawn up. It was analysed in terms of stop sign ratio; which indicates the percentage of stop signs they obey. A dead stop is again defined as a speed of less than 10 km/h for two consecutive samples; with an upper threshold of 15 km/h. Stop sign and traffic light crossings are individually prescribed to every recorded trip. GPS error of 12.5 m is accounted for at a stop sign and traffic lights account for 50 m of backlog traffic, or queue length. These parameters were defined according to a control test.

As shown, the measured minibus taxis only adhered to roughly 30% of stop signs, totalling 925 stops. With a total of 245 stop signs crossed, a 100% stop sign adherence would increase the total stops to 1098. As stated in Section 4.2.1 the total number of

Table 3
Stop analysis.

Route	Stop signs encountered	Vehicle stopped[a,b]	Stop sign ratio (%)	Traffic light stopped	Other stopped[c]	Total stops
Urban 1	64	17	26.6	13	177	207
Urban 2	12	4	33.3	11	40	55
Inter-city 1	36	13	36.1	64	113	190
Inter-city 2	21	6	28.6	54	108	168
Uphill	63	19	30.2	33	134	186
Downhill	49	13	26.5	28	78	119
Total	245	72	29.4	203	650	925

[a]Vehicle is considered to have stopped if the GPS-recorded speed is below 10 km/h for two consecutive seconds or more.
[b]GPS error of 12.5 m is accounted for at stop sign.
[c]Other stops are presumably to drop off and pick up passengers.

stops in the simulation reduced to 1093 when removing high-speed stops. This shows near perfect correlation to the number of real stops if all stop signs were obeyed, 1098.

It can be concluded that the omission of stops signs in urban areas are made up by the way in which waypoint progression is done, assuming the minibus taxi obeyed all stop signs. However, it is shown that only approximately 30% are obeyed in reality. We thus aim to estimate the change in energy expenditure if the virtual vehicle also 'obeyed 30% of stop signs'.

An estimated average speed of between 25 km/h and 35 km/h is chosen to quantify the energy expenditure due to obeying a stop sign in SUMO. A mean energy expenditure of 0.03 kWh/stop is found, for a mean distance of 0.09 km. After the removal of 173 stops (the number of stop signs the real minibus taxi did not obey), a new mean energy expenditure of 0.442 kWh/km was found as shown in Table 2.

5. Conclusion

Due to labour requirements and high cost to generate accurate 1 Hz mobility data, micro-traffic simulators are essential for understanding the micro-level mobility of paratransit in Sub-Saharan Africa, an industry where minutely sampled mobility data is relatively common. Although there are tools used in industry to simulate mobility and subsequently calculate energy expenditure, we have shown that they do not adequately model and simulate paratransit in a SSA context. In this paper, various aspects of the upsampling of low-frequency mobility data is analysed and compared to measured high-frequency data and the subsequent energy expenditure results thereof. We have shown that there are various modifications that can be made to; and shortcomings still present in mobility simulators as a substantial difference is evident between the energy expenditure estimated by the measured mobility and simulated mobility. As part of this paper, we have rectified the omission of elevation data, and inaccurate speed and acceleration profiles.

An initial overestimation of 14% for energy per distance is found when using the simulated data from EV-Fleet-Sim as input to Hull et al. (2022b)'s model, with a total energy per trip over estimation of 46%. This suggests a smaller battery requirement for electric minibus taxis and a reduced impact on the electrical grid than was initially suggested when working with only simulated mobility data from EV-Fleet-Sim.

In another use case of EV-Fleet-Sim, Abraham et al. (2021) suggested an energy expenditure of 0.93 kWh/km. In their paper, the built in electro-kinetic model by Kurczveil et al. (2014) is used. Apart from a different input dataset used, this is the only methodological difference in calculating energy expenditure to this paper. As previously mentioned, further investigation into this matter is suggested.

Additional areas found for further work entails that of waypoint progression and reverse geocoding. As explained in Section 4.2.1, the way in which waypoints are followed in the simulation results in a substantial over-estimation of energy usage. With the analysis and theoretical contribution provided on this matter, future work on the routing model within SUMO is suggested. Although reverse geocoding cannot be rectified due to the nature of how vehicles interact with the road, the theoretical contribution on the impact this has on the energy analyses provided in this paper serves as a guideline and caution for future simulation use.

Another shortcoming in the simulation, which was not investigated, is the omission of virtual traffic. As this effects the driving style and subsequent energy analysis, an accurate representation of real-world traffic is paramount for the eventual accuracy of simulated results. It is advised that modification on the simulation tool in this regard, and analyses on the implication of the absence of virtual traffic be done as part of future work.

Although various other methodological errors are pointed out when simulating high-frequency mobility, the resultant improved EV-Fleet-Sim mobility output assists in having accurate energy expenditure estimations in regions where micro-level mobility data is scarce. It is thus clear that the simulation should first be improved, whereafter managerial and policy recommendations to Government, paratransit operators and other stake-holders can be made. This benefits all parties in the electrification of the minibus taxi industry; from designing the optimal electric minibus taxi, to quantifying the impact on the electrical grid of an operating electric minibus taxi fleet.

In addition, having accurate micro-level mobility data lays the groundwork for future work to do drive cycle analysis of the industry. This will allow energy expenditure to be spatially estimated for various regions. Adequate charging infrastructure can subsequently be planned according to these results and assist in the e-mobility transition of paratransit in Sub-Saharan Africa as a whole.

Data availability

All our data is public and linked in the paper.

References

Abraham, C., Giliomee, J., Fußl, L., Ritchie, M., 2022. EV-fleet-sim. URL: https://eputs.gitlab.io/ev-fleet-sim/docs.html.

Abraham, C.J., Rix, A.J., Ndibatya, I., Booysen, M.J., 2021. Ray of hope for sub-Saharan Africa's paratransit: Solar charging of urban electric minibus taxis in South Africa. Energy Sustain. Dev. 64, 118–127. http://dx.doi.org/10.1016/j.esd.2021.08.003.

Al-Wreikat, Y., Serrano, C., Sodré, J.R., 2021. Driving behaviour and trip condition effects on the energy consumption of an electric vehicle under real-world driving. Appl. Energy 297, 117096. http://dx.doi.org/10.1016/j.apenergy.2021.117096, URL: https://www.sciencedirect.com/science/article/pii/S0306261921005444.

Behrens, R., McCormick, D., Mfinanga, D., 2015. In: Behrens, R., McCormick, D., Mfinanga, D. (Eds.), Paratransit in African Cities: Operations, Regulation and Reform, first ed. Routledge, London, pp. 1–311. http://dx.doi.org/10.4324/9781315849515.

Booysen, M.J., Andersen, S.J., Zeeman, A.S., 2013. Informal public transport in Sub-Saharan Africa as a vessel for novel intelligent transport systems. In: 16th International IEEE Conference on Intelligent Transportation Systems. ITSC 2013, pp. 767–772. http://dx.doi.org/10.1109/ITSC.2013.6728324.

Bruun, E., Mistro, R.D., Venter, Y., Mfinanga, D., 2016. The state of public transport systems in three Sub-Saharan African cities. In: Behrens, R., McCormick, D., Mfinanga, D. (Eds.), Paratransit in African Cities: Operations, Regulation and Reform, first ed. Routledge, New York, http://dx.doi.org/10.4324/9781315849515.

Cervigni, R., 2013. Assessing Low-Carbon Development in Nigeria: An Analysis of Four Sectors. World Bank Publications.

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

Collett, K.A., Hirmer, S.A., Dalkmann, H., Crozier, C., Mulugetta, Y., McCulloch, M.D., 2021. Can electric vehicles be good for Sub-Saharan Africa? Energy Strategy Rev. 38, 100722. http://dx.doi.org/10.1016/j.esr.2021.100722.

Diaz Olvera, L., Plat, D., Pochet, P., 2019. Looking for the obvious: Motorcycle taxi services in Sub-Saharan African cities. J. Transp. Geogr. 102476. http://dx.doi.org/10.1016/J.JTRANGEO.2019.102476, URL: https://www.sciencedirect.com/science/article/pii/S0966692317306269.

Dongfeng, T., 2021. Dongfeng electric 15 seater mini van (RHD). URL: https://tara-dongfeng.co.uk/listings/dongfeng-ev-15seat-2-2/.

Dorothy, M., Herrie, S., David, M., 2016. The nature of paratransit operations. In: Roger, B., Dorothy, M., David, M. (Eds.), Paratransit in African Cities: Operations, Regulation and Reform, first ed. Routledge, New York, pp. 59–78.

Ebot Eno Akpa, N.A., Booysen, M.J., Sinclair, M., 2016. Publicly available annotated dataset of tracked taxis. URL: http://staff.ee.sun.ac.za/mjbooysen/TaxiData/. Archived at: https://gitlab.com/eputs/data/fcd-stellenbosch.

Ebot Eno Akpa, N.A., Booysen, M.J., Sinclair, M., 2019. Fuel savings as an incentive for speed compliance in the informal public transport industry in South Africa. In: 2019 IEEE Intelligent Transportation Systems Conference. ITSC, pp. 492–496. http://dx.doi.org/10.1109/ITSC.2019.8917491.

Ehebrecht, D., Heinrichs, D., Lenz, B., 2018. Motorcycle-taxis in sub-Saharan Africa: Current knowledge, implications for the debate on "informal" transport and research needs. J. Transp. Geogr. 69, 242–256. http://dx.doi.org/10.1016/j.jtrangeo.2018.05.006, URL: http://www.sciencedirect.com/science/article/pii/S0966692317303344.

Erdmann, J., 2022. How to make accelerations change smoothly? URL: https://github.com/eclipse/sumo/issues/10639.

Evans, J., O'Brien, J., Ch Ng, B., 2018. Towards a geography of informal transport: Mobility, infrastructure and urban sustainability from the back of a motorbike. Trans. Inst. Br. Geogr. 43 (4), 674–688. http://dx.doi.org/10.1111/tran.12239, URL: https://rgs-ibg.onlinelibrary.wiley.com/doi/abs/10.1111/tran.12239.

Faria, M.V., Duarte, G.O., Varella, R.A., Farias, T.L., Baptista, P.C., 2019. Driving for decarbonization: Assessing the energy, environmental, and economic benefits of less aggressive driving in Lisbon, Portugal. Energy Res. Soc. Sci. 47, 113–127. http://dx.doi.org/10.1016/j.erss.2018.09.006.

Galuszka, J., Martin, E., Nkurunziza, A., Achieng' Oginga, J., Senyagwa, J., Teko, E., Lah, O., 2021. East Africa's policy and stakeholder integration of informal operators in electric mobility transitions—Kigali, Nairobi, Kisumu and Dar es Salaam. Sustainability 13 (4), http://dx.doi.org/10.3390/su13041703, URL: https://www.mdpi.com/2071-1050/13/4/1703.

Geospatial Commission, 2021. Positioning the UK in the fast lane. URL: https://www.gov.uk/government/news/the-geospatial-commission-announces-transport-innovation-competition-winners-and-outlines-how-to-position-the-uk-in-the-fast-lane.

Google, 2015. Google maps. URL: www.google.com/maps.

Graf, F., Whitesides, R., Mihai, B.C., Benno, K., 2015. SRTM plugin for OpenStreetMap's osmosis. URL: https://github.com/locked-fg/osmosis-srtm-plugin.

Hanacek, N., 2021. How do you measure your location using GPS? URL: https://www.nist.gov/how-do-you-measure-it/how-do-you-measure-your-location-using-gps.

Higer, 2020. Higer H5C EV specifications. URL: https://en.higer.com/NewEnergy/info.aspx?itemid=6851.

Hull, C., Giliomee, J., Collett, K.A., McCulloch, M.D., Booysen, M.J., 2022a. 1Hz GPS Tracking Data on Minibus Taxi Paratransit Vehicles in South Africa. Mendeley, http://dx.doi.org/10.17632/XT69CNWH56.1, URL: https://data.mendeley.com/datasets/xt69cnwh56/1.

Hull, C., Giliomee, J., Collett, K.A., McCulloch, M.D., Booysen, M.J., 2022b. High fidelity estimates of paratransit energy consumption from per-second gps tracking data. Transportation Research Part D http://dx.doi.org/10.1016/j.trd.2023.103695

Iora, P., Tribioli, L., 2019. Effect of ambient temperature on electric vehicles' energy consumption and range: Model definition and sensitivity analysis based on Nissan Leaf data. World Electr. Veh. J. 10 (1), http://dx.doi.org/10.3390/wevj10010002, URL: https://www.mdpi.com/2032-6653/10/1/2.

Jennings, G., Behrens, R., 2017. The Case for Investing in Paratransit: Strategies for Regulation and Reform.

KCCA, 2016. Multimodal Urban Transport Master Plan for Greater Kampala Metropolitan Area. Technical Report, Kampala Capital City Authority (KCCA), Kampala.

Kurczveil, T., Álvarez López, P., Schnieder, E., 2014. Implementation of an Energy Model and a Charging Infrastructure in SUMO. Springer Berlin Heidelberg, pp. 33–43. http://dx.doi.org/10.1007/978-3-662-45079-6_3.

Lopez, P.A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., Wießner, E., 2018. Microscopic traffic simulation using SUMO. In: The 21st IEEE International Conference on Intelligent Transportation Systems. IEEE, pp. 2575–2582, URL: https://elib.dlr.de/127994/.

Mbara, T.C., 2002. Activity patterns, transport and policies for the urban poor in Harare, Zimbabwe. Final ctry. Rep. [En Ligne.]. www.transport-links.org/transport_links/filearea/documentstore/305_Zimbabwe%20Final%20Report, URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1003.6243&rep=rep1&type=pdf.

Miri, I., Fotouhi, A., Ewin, N., 2020. Electric vehicle energy consumption modelling and estimation—A case study. Int. J. Energy Res. 45, http://dx.doi.org/10.1002/er.5700.

Motavalli, J., 2021. Every automaker's EV plans through 2035 and beyond. URL: https://www.forbes.com/wheels/news/automaker-ev-plans/.

Murphey, Y.L., Milton, R., Kiliaris, L., 2009. Driver's style classification using jerk analysis. In: 2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems. pp. 23–28. http://dx.doi.org/10.1109/CIVVS.2009.4938719.

Mutiso, W., Behrens, R., 2011. 'Boda Boda' bicycle taxis and their role in urban transport systems: case studies of Kisumu and Nakura, Kenya. In: 30th Annual Southern African Transport Conference 11-14 July 2011 "Africa on the Move", CSIR International Convention Centre. Pretoria, South Africa, pp. 430–444, URL: http://hdl.handle.net/2263/17308.

NASA, 2015. Enhanced shuttle land elevation data. URL: https://www2.jpl.nasa.gov/srtm/.

Ndibatya, I., Booysen, M.J., 2021. Characterizing the movement patterns of minibus taxis in Kampala's paratransit system. J. Transp. Geogr. 92, 103001. http://dx.doi.org/10.1016/j.jtrangeo.2021.103001, URL: https://www.sciencedirect.com/science/article/pii/S0966692321000545.

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: Proc. 35th Southern African Transport Conference. Pretoria, pp. 4–7.

Neumann, A., Joubert, J.W., 2016. The "minibus" contribution. In: Horni, A., Nagel, K., Axhausen, K.W. (Eds.), The Multi-Agent Transport Simulation MATSim. Ubiquity Press, London, http://dx.doi.org/10.5334/baw.17.

OpenStreetMap, 2022. Osmisis. URL: https://wiki.openstreetmap.org/wiki/Osmosis.

OpenStreetMap contributors, 2022. Planet dump retrieved from https://planet.osm.org. https://www.openstreetmap.org.

OpenStreetMap Contributors, 2022. OpenStreetMap data. https://www.openstreetmap.org.

Pojani, D., Stead, D., 2017. The urban transport crisis in emerging economies: An introduction. In: Pojani, D., Stead, D. (Eds.), The Urban Transport Crisis in Emerging Economies. Springer International Publishing, Cham, pp. 1–10. http://dx.doi.org/10.1007/978-3-319-43851-1_1.

Rahemi, N., Mosavi, M., 2021. Positioning accuracy improvement in high-speed GPS receivers using sequential extended Kalman filter. IET Signal Process. 15, http://dx.doi.org/10.1049/sil2.12027.

Ritche, H., Roser, M., 2018. CO2 and greenhouse gas emissions. Our world in data in.

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 http://dx.doi.org/10.1016/j.isci.2022.104943.

Ruivii, 2019. Toano 15 seats electric minibus. URL: https://www.ruviivehicle.com/products/minibus/toano-15-seats-electric-minibus/.

Shannon, C., 1949. Communication in the presence of noise. Proc. IRE 37 (1), 10–21. http://dx.doi.org/10.1109/jrproc.1949.232969.

Sims, R., Schaeffer, R., Creutzig, F., Cruz-Núñez, X., D'Agosto, M., Dimitriu, D., Meza, M.F., Fulton, L., Kobayashi, S., Lah, O., McKinnon, A., Newman, P., Ouyang, M., Schauer, J., Sperling, D., Tiwari, G., 2014. Transport. In: Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., von Stechow, C., Zwickel, T., Minx, J. (Eds.), Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III To the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p. 603,

SLOCAT, 2021. *Tracking Trends in a Time of Change: The Need for Radical Action Towards Sustainable Transport Decarbonisation* - Transport and Climate Change Global Status Report – 2nd edition. URL: www.tcc-gsr.com.

Smith, T.R., Kersey, V., Bidwell, T., 2001. The effect of engine age, engine oil age and drain interval on vehicle tailpipe emissions and fuel efficiency. SAE Trans. 110, 1838–1861, URL: http://www.jstor.org/stable/44742784.

SUMO, 2022a. Definition of vehicles, vehicle types, and routes. URL: https://sumo.dlr.de/docs/Definition_of_Vehicles%2C_Vehicle_Types%2C_and_Routes.html.

SUMO, 2022b. Netedit is a graphical network editor included in SUMO. URL: https://sumo.dlr.de/docs/Netedit/index.html.

SUMO, 2022c. Simulation of urban MObility. URL: https://www.eclipse.org/sumo/.

SUMO, 2022d. SUMO floating car data output. URL: https://sumo.dlr.de/docs/Simulation/Output/FCDOutput.html.

SUMO, 2022e. SUMO simulation routing. URL: https://sumo.dlr.de/docs/Simulation/Routing.html.

SUMO, 2022f. Traffic lights. URL: https://sumo.dlr.de/docs/Simulation/Traffic_Lights.html.

Sunday Times Driving, 2022. Car makers' electric vehicle plans for 2022 and beyond. URL: https://www.driving.co.uk/news/new-cars/current-upcoming-pure-electric-car-guide-updated/.

United Nations, 2015. Do you know all 17 SDGs? URL: https://sdgs.un.org/goals.

United Nations, 2019. World population prospects 2019, custom data acquired via website.

van Benthem, A., 2021. Arthur van Benthem on new cars, old cars, and air pollution today. URL: https://penntoday.upenn.edu/news/wharton-arthur-van-benthem-new-cars-old-cars-and-air-pollution-today#:~:text=While%20federal%20emissions%20standards%20have,of%20total%20pollution%2C%E2%80%9D%20says%20van.

Weber, M., Paleino, D., 2018. OSMtools. URL: https://gitlab.com/osm-c-tools/osmctools.

World Health Organization, et al., 2018. Indicator 3.9. 1: Mortality rate attributed to household and ambient air pollution (per 100 000 population).

Wu, X., Freese, D., Cabrera, A., Kitch, W.A., 2015. Electric vehicles' energy consumption measurement and estimation. Transp. Res. D 34, 52–67. http://dx.doi.org/10.1016/j.trd.2014.10.007, URL: https://www.sciencedirect.com/science/article/pii/S1361920914001485.

Zeeman, A., Booysen, M.J., 2014. Public transport sector driver behaviour : measuring recklessness using speed and acceleration. In: Southern African Transport Conference. URL: http://hdl.handle.net/2263/45581.

Zinkernagel, R., Evans, J., Neij, L., 2018. Applying the SDGs to cities: Business as usual or a new dawn? Sustainability 10 (9), http://dx.doi.org/10.3390/su10093201, URL: https://www.mdpi.com/2071-1050/10/9/3201.