Master's Thesis in Geographical Information Science nr 108

Automating synthetic trip data generation for an agent-based simulation of urban mobility

Paul Gabriel dos Santos

2019

Department of
Physical Geography and Ecosystem Science
Centre for Geographical Information Systems
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Automating synthetic trip data generation for an agent-based simulation of urban mobility

---

Paul Gabriel dos Santos

Master's degree thesis, 30 credits in Geographical Information Systems

Department of Physical Geography and Ecosystem Science, Lund University

Supervisor
Abdulghani Hasan
Lund University

Paul Gabriel dos Santos

Abstract

This paper explores the use of an automated pipeline to construct synthetic (artificially derived) trip data from aggregate socio-demographic sources to build a simulation of individual vehicles interacting with one another. The study shows that quality data sources are required in order to do this effectively and accurately. It is shown that aspects of typical patterns and behaviours may still not be represented within the final simulation. It is often a complex, expensive and impractical exercise to obtain in situ measurements across an entire city to build simulation scenarios to help with effective planning and understanding of emissions at a fine resolution. Since road traffic is a major source of harmful pollutant emissions, we explore the use of the simulation to generate emission outputs at a per vehicle level through simulation time. The paper concludes that although a valid simulation scenario can be constructed from the derived synthetic dataset, new techniques need to be developed in order to obtain an equilibrium in the simulation to allow it to not only behaves as a valid urban mobility scenario but can also be calibrated to align to represent reality.

Keywords: Geography, GIS, Agent-based Modelling, Simulation Science, Network Modelling, Simulation of Urban Mobility, Emissions, Environmental Modelling

Advisor: **Abdulghani Hasan**

Master's degree thesis, 30 credits in Geographical Information Systems, 2019

Department of Physical Geography and Ecosystem Science, Lund University

Thesis nr 108

Table of Contents

Tables

Figures

1. Introduction

Using a computer simulation of individual vehicles interacting with one another within the City of Cape Town, South Africa we will attempt to create a calibrated representation of reality in an automated way from trips synthetically derived from aggregated demand statistics.

Using this simulated environment, we will calculate at a fine scale (per vehicle per second of the simulation) a range of emissions that these vehicles pump into the environment (CO_2, CO, HC, NOx, PMx) as well as fuel consumption to illustrate the utility of such a simulation.

Modelling this type of problem at such a fine resolution allows us to understand not only the dynamic interaction of agents acting in a system, but also which specific characteristics about the relationships between these agents have the largest impact on the simulated scenario. This project is unique in the sense that it will be the first piece of work at this level of modelling detail done for a sub-Saharan African city.

We wanted to find out if a simulation could be built without in-situ measurements that closely approximates real conditions using synthetic data derived from aggregate population-level statistics. If this is possible, there are many benefits to being able to construct simulations of urban mobility that are not dependent on installation of complicated equipment or the coordination of complex field surveys. The model that is produced is evaluated against historical aggregates to measure how closely it approximates reality.

The focus of this paper is on an automated pipeline that was developed to generate the relevant simulation data as well as the techniques that can be used to evaluate the simulation's outputs and behaviours in a novel way.

2. Hypothesis & Aims

2.1 Hypothesis

By using a detailed model of vehicles acting on one another in a micro-simulation resultant from aggregate-level demand data used to generate trips per simulated vehicle, details of the simulated environment (and scenario) can be modelled with a high degree of precision.

2.2 Aim

Automate the development of a microscopic simulation of urban mobility using synthetic trip data that can be leveraged to explore aspects of the generated traffic scenario throughout the simulated time and environment.

2.3 Sub-Aims

2.3.1 Develop an automated pipeline (process flow) for preparing the simulation input datasets from high level statistics describing the study area.

2.3.2 Evaluate how closely the simulation approximates expected historic traffic conditions.

2.3.3 Use the model to calculate emission estimations for the simulated scenario at a per vehicle level as an example of how such a simulation could be leveraged in practice.

3 Study Area

Transportation in South Africa is dominated by road travel –over 15 million commuters use 200 000 mini-bus taxis every day and this number continues to climb. The average age of these taxis is over 9 years old which poses safety risks to commuters – these vehicles are often poorly maintained and lack the emission efficiencies of newer counterparts. This informal sector of the economy accounts for approximately 45 000 direct and 600 000 indirect jobs. (Transactioncapital.co.za, 2017)

In 2017 Cape Town was named the most congested city in South Africa, according to the TomTom Traffic Index which measured congestion on the road of 390 countries around the world using data from 2016. According to this index, residents of Cape Town spent on average 35% extra time travelling (a 5% increase on the previous year) which amounts to 42 minutes per day or 163 hours per year per motorist. (TomTom Traffic Index, 2016).

The City of Cape Town announced plans in April 2018 to combat traffic congestion in Cape Town by investing in a R481 million (approximately 34.38 million United States dollars) traffic-relief construction project. (Reporter, 2018). Starting July 1, 2018, six projects are reportedly under construction and there are an additional 16 projects that are in the planning and design phase.

As seen in figure 1, the central business district of Cape Town is an interesting challenge to model as found at approximately 33.91S and 18.43E. The city centre faces unique challenges as the city 'bowl' is built against Table Mountain on one side with the Atlantic and Indian Ocean on the other. As urbanisation increases, the city needs to be deliberate in its efforts to upgrade road infrastructure to ensure these upgrades have the maximum impact.

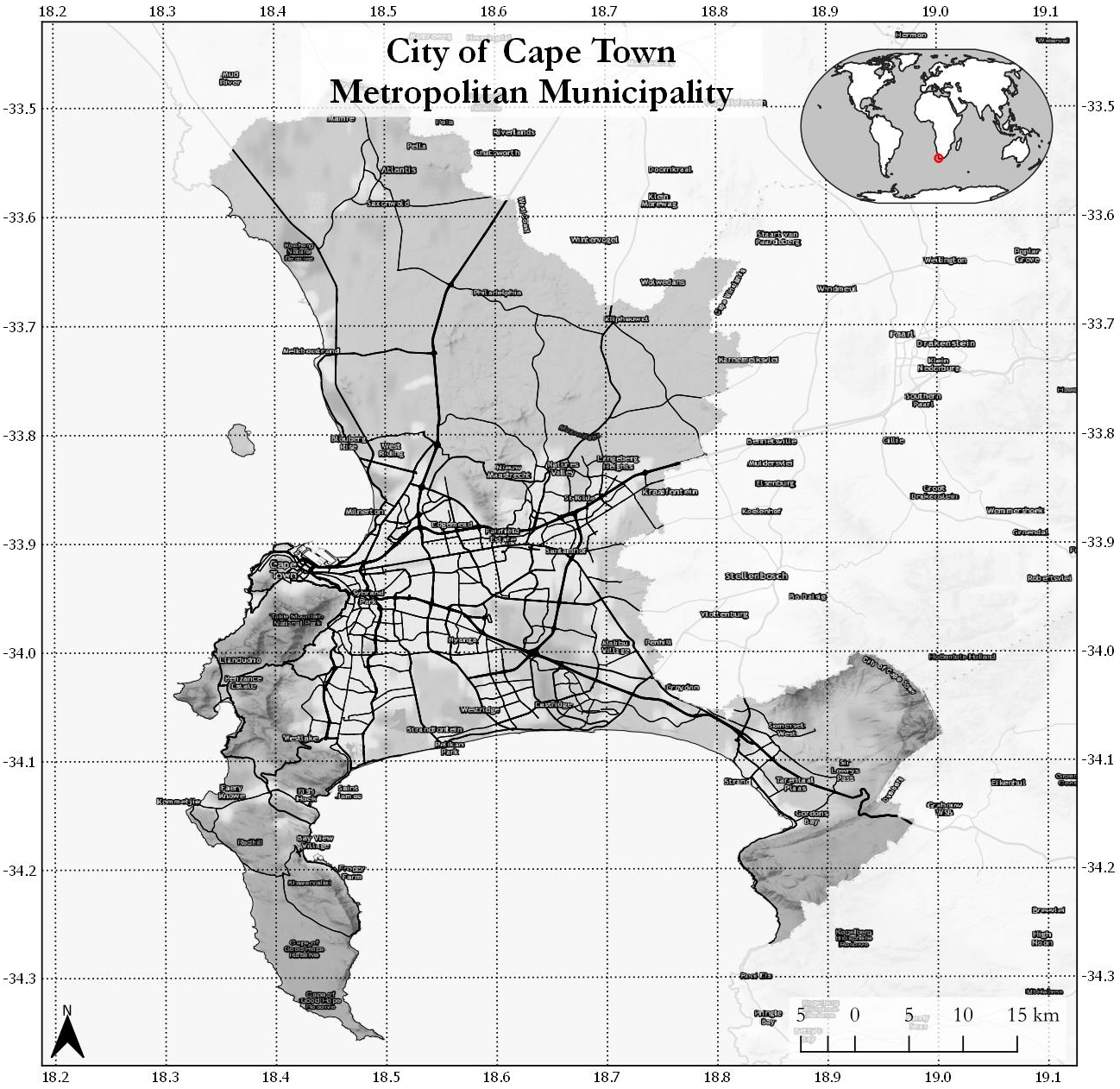

Figure 1 Study Area: City of Cape Town

4 Literature Review

4.1 Background

Road traffic is a significant contributor to global atmospheric pollutants. In rapidly urbanising countries traffic is increasing leading to overcrowded roads and increased pollutants being released into the atmosphere as a result (Ragab, Hashim and Asar, 2017). An Environmental Protection Agency (EPA) report in 2005 showed that road traffic contributes heavily to several types of emissions on a large scale – 58.8% of Carbon Monoxide (CO), 35.5% of Nitrogen Oxides (NOX), and 25.8% of Hydrocarbons (HC) emissions can be attributed to vehicle emissions (Environmental Protection Agency, 2005).

To accurately measure the effectiveness of proposed traffic management measures, microscopic traffic simulations; a class of scientific models of vehicular traffic dynamics, can play an important role (Jeihani et al., 2015). In contrast to macroscopic models, microscopic traffic flow models simulate single vehicle-driver units; the dynamic variables of the models represent microscopic properties like the position, velocity and emissions of individual vehicles.

Much research has been done in this space outside the context of Africa. There are numerous examples of such studies being done across the globe – Eastern countries such as China (Kun and Lei, 2007) as well as European countries such as Germany (Neunhauserer and Diegmann, 2010) and Austria (Zallinger et al., 2010) feature in papers examining the utility microscopic simulations show when assessing the impact of congested road networks on harmful emission outputs. Recently, Egypt (Ragab, Hashim and Asar, 2017) has received some academic attention in this space, however no literature can be found for Sub-Saharan African countries. Many South African cities are faced with the burden of increasing urbanisation. The City of Cape Town is the most congested city in Africa (TomTom Traffic Index, 2016).

4.2 The Evolution of Agent-based Simulation

Digital computers appeared in the late 1940s, by the mid-1950s scientific applications involving spatial as well as temporal problems were being explored. Mathematical theories these are based on were slowly developing prior to the invention of the computer. In 1955, the first models of traffic flow were implemented in a digital environment in the Chicago Area Transportation Study (Heppenstall et al, 2014).

Many of the first models and the ones that followed are typically referred to as Land Use Transportation interaction (LUTI) models due to the equilibrium-seeking nature – they aggregate at the population level to help with planning in the context of providing new transport capacity when urban areas develop driven by economic growth and are based in social physics and urban economics. Early models were very specific in design, but as our understanding of computer science grew, generic frameworks were developed that

paved the way for techniques such as cellular automata (CA) and agent-based models. (Heppenstall et al, 2014).

Cellular Automata Models are the simplest form of urban model because populations and environments are represented using the same 'cell' object which can effectively be turned on or off. In CA models, only the physically adjacent neighbours can influence each other, and this action is modelled to act uniformly in all directions, leveraging fractal patterns or using oriented diffusion based on certain developmental criteria. CA models are criticised for their supply-side modelling of physical constraints. Demand and the kinds of interaction reflected in transport cannot be modelled and CA models tend to simulate smooth spatial dynamics as a result. (Heppenstall et al., 2014).

Heppenstall, et al. (2014) state that microsimulations are temporally dynamic with individuals represented in terms of their behaviour which is intrinsically dynamic. The population in the simulation is described in terms of a distribution of characteristics which means that if a random sample of agents were drawn from the model you would be able to describe the larger problem space. Microsimulation models are essential tools for sampling large-scale populations where it is impossible to represent all the individuals explicitly and where some sense of the heterogeneity of the population needs to be represented in the model.

Bazzan and Klügl (2013) explain that agent-based modelling in the context of traffic and transportation are adaptive and robust due to their self-organisation capability. Autonomous agents are appropriate for modelling heterogenous systems since every entity is described in terms of its own architecture, state and behaviour. There is an intuitive level of interaction between users and the system through high-level abstractions that allow modellers to visually interpret microscopic properties of a given system. Problem solving phases that an agent needs to consider in the context and environment of the network being modelled means that dynamic relationships can be controlled, disconnected and established from a local point of view allowing the modeller to describe entities over their life span in a simulation in a coherent way.

4.3 Selecting a Modelling Environment

There are several agent-based simulation packages that have evolved to model different aspects of transportation networks. Saidallah, El Fergougui and Elalaoui (2016) conducted an in-depth evaluation of eleven different microscopic traffic simulation packages most commonly used. Of these, 4 were found to be open source with the remaining 7 being commercial in nature. For this study we will be focusing on the open-source variants – refer to table 1.

Table 1 Open Source Agent-based Road Network Simulation Environment Comparison

	Vers.	Release Date	Platform	Openness	License	Documentation
TRANSIMS	7.1	Jul-18	Windows	Source available on Google Code for version 4. SourceForge holds binaries for version 7.	NASA Open Source Agreement V1.3	Poor. Broken wiki and old mailing lists.
MITSIMLab	Not Listed	Aug-09	Linux - Redhat 7.3	Distributed as binaries, no public code repository	MITSIMLab Open Source License	Four-page flyer.
MATSim	0.10.1	Aug-18	All (Java Source)	Source available on GitHub	GNU General Public License	Dedicated documentation with a 600-page user guide at its core. Active community.
Sumo	1.0.1	Sep-18	All (Java Source)	Source available on GitHub	Eclipse Public License Version 2	Dedicated documentation with a manual, tutorial and examples. Active community.

Of these, MATSim and SUMO seem to be the most complete software packages at this stage. They have excellent documentation, are actively developed, are open in both use and code availability and have an extension/plugin framework allowing users to extend the base software past its initial design scope.

4.4 Emission Calculations

With regards to emission calculations for scenarios, in MATSim, you are dependent on an extension called 'emissions'. This extension uses the road segment enter and exit events, vehicle speed, stop & go & free flow sections of the link to look up emissions against the HBEFA 3.1 database for arbitrary vehicle types after emissions characteristics have been selected for the vehicle types. In SUMO, emission models are part of the core software – linking you to several emission models HBEFA 2.1, HBEFA 3.1, PHEMlight and an Electric Vehicle Model. Sumo allows you to evaluate the emissions for individual agent trips and to visualise emissions per vehicle or per lane. All models allow for the measurement of CO_2, CO, HC, NOx, PMx and fuel consumption. The functionality available in SUMO appears to be more granular and is therefore most aligned with the research aims of this paper and thus was chosen as the modelling environment for this study.

4.5 Emission Models

The Handbook Emission Factors for Road Transportation (HBEFA) provides emission factors for existing vehicle categories such as PC, LDV, HGV, urban buses, coaches and motor cycles in g/km for a

variety of traffic situations. Both regulated and non-regulated pollutants are included - CO, HC, NOx, PM, several components of HC (CH4, NMHC, benzene, toluene, xylene), fuel consumption (gasoline, diesel), CO_2, NH3 and N2O, PN and PM. This is an open source reference. HBEFA was developed on behalf of the Environmental Protection Agencies of Germany, Switzerland and Austria (Hbefa.net, 2018)

The Passenger Car and Heavy-Duty Emission Model (PHEM) is a commercial, closed-source instantaneous vehicle emission model. A simplified version of this model was developed and embedded within SUMO to leverage the publicly available component of this model – this only covers two emissions classes which coverts passenger vehicles powered by Diesel and gasoline. Further vehicles classes need to be licensed to use them in SUMO (Sumo.dlr.de, 2018). This model, due to its closed source nature was not chosen for use in this study.

4.6 Calibration

An important calibration step in building an agent-based simulation of urban mobility depends on the way in which the network is loaded to ensure that people are travelling from one area to the next based on true demand scenarios.

In MATSim an agent is given a plan for a series of stops which the synthetic traveller needs to follow. Entities are treated individually but modelled with a low level of detail meaning that the model is more mesoscopic in nature – this allows MATSim to run very quickly – often able to simulate 24 hours with a 1% sample of millions of travellers in 10 minutes. The vehicle dynamics are loaded through a queue model and allows for other methods such as full kinematic waves. Public transit can be modelled with real schedules. (Matsim.atlassian.net, 2018)

Demand (the agent plans) and supply (the execution of said plans within the constraints of the system) are iteratively resolved allowing the agents to learn. Agents favour plans with good performance but are also able to modify plans or re-evaluate inferior plans.

Using real traffic count data Flötteröd, Chen and Nagel (2011) identified the importance of model calibration. Calibration can only happen for network segments that have real-world data – agents acting in other parts of the system, therefore, are implicitly changed through interactions with the adjusted agents in the network.

In SUMO there are several methods for loading demand – the ACTIVITYGEN module can turn aggregate population statistics into traffic demand, OD2TRIPS allows you to use origin-destination data available from transport authorities into trip definitions, data from induction loops (traffic counters) can be used along with several other methods. Depending on availability of data for a given study, this allows for more adaptability in data sources that can be used to calibrate the model correctly. (Sumo.dlr.de, 2018)

Flitsch et al (2018) conducted a study specifically on the calibration of traffic simulation in SUMO – using heterogenous data sources (phone data, induction loops and Bluetooth sensors). They found that a significant amount of data for the whole network is required to effectively calibrate the entire network and that the use of real-time traffic data to load the model both initially and while trying to model scenarios in real-time is both possible and effective. Test rides and traffic cameras vehicle counts are identified as good sources for plausibility checks to assess the traffic data quality.

4.7 Previous Experimentation

Agent-based simulation has far reaching applications in traffic control and management system architectures & platforms, roadway transportation, air-traffic control and management, railway transportation and multi-agent traffic modelling and simulation (Dias et al., 2013).

In 2009, Stevanovic et at. developed a traffic control optimisation to reduce fuel consumption and vehicular emission using an agent-based simulation at the level of components in the road network, namely, traffic signals. The authors encountered several challenges in their study based on the modelling techniques used. For example; if a vehicle is stopped and idling, it consumes less fuel than one that runs at 40 mph. So, although its fuel-per-mile consumption is higher when the vehicle is idling, its fuel-per-second consumption and emissions are lower. The authors final conclusions state that the number of stops, fuel consumption and CO_2 emissions are not reliable objective functions in the optimisation of signal timings – optimising for fuel consumption in their model only resulted in a reduction of 1%.

Leveraging SUMO, Dias et al. (2014) developed an inverted ant colony optimization algorithm to be used as a decentralised traffic management system where the agents act in the way an ant colony does to avoid congestion and distribute traffic through the network. Their findings show that traffic density can be decreased using this type of routing algorithm and that travel times can be reduced by up to 84%. This has a significant impact on emissions as a result. It was found that when 75% of the agents were using the inverted ant colony optimisation, there was a reduction in CO_2 emissions of 44% – with the future of autonomous cars not too far away, this type of insight will mean that self-driving cars can be designed in as optimal a fashion as possible.

4.8 Policy and Legislation

South Africa is planning on implementation a Carbon Tax in 2019 – the aim is to improve energy efficiency through switching to alternative energy, stimulate growth through new technologies and enterprises as well as to encourage the development of new products to stimulate investment and provide opportunity for the creation of additional jobs.

Due to the Paris Agreement (Pmg.org.za, 2018), a global initiative to reduce carbon emissions to try and ensure that temperatures to not increase past 2 degrees Celsius, most countries will have a form of carbon tax in the next few years. South Africa is one of the most carbon-intense economies in the world due to their tightly coupled dependence on coal which drives economic activity. The proposed pricing of R120(~$8.7)/$tCO_2e$ (in 2019) is being criticised as insufficient to properly institutionalise nation-wise change (Fakir, 2018).

The South African Department of Environmental Affairs released a report in 2014 titled the *Green House Gas National Inventory Report* which covered the time from 2000 to 2010 in line with the Intergovernmental Panel on Climate Change (IPCC) 2006 guidelines. The main challenged of this extensive piece of work is the availability of accuracy activity data to be able to calculate emissions across different sectors of the economy (GHG National Inventory Report South Africa 2000 - 2010. (2014)). No update to the findings of the report have been conducted since, which leaves 8 years of uncertainty in terms of emission estimates at a national level. All findings are reported at a national level with a heavy focus on industries. South Africa has an abundant supply of mineral resources – the economy was originally centred on natural resources and agriculture with mining as a major component of the Gross Domestic Product (GDP).

4.9 Reflection

Agent-based simulations have developed over time into a useful tool that allows us to model complex behaviours in an efficient manner. This allows us to explore system-wide implications of agent-modelled behaviours. Particularly in the context of transportation network optimisation and planning it provides us the necessary tools to simulate both changes to the network and events on the network to inform decision making.

Global policy is shifting to accommodate 'greener' thinking. Exploring emissions on a microscopic level thanks to technological advances can aid in how these policies evolve over time and can provide insight into how we impact our environment through daily commutes.

5 Method

5.1 Scope

The following components are included in the scope of this study and have a direct impact on model outputs.

5.1.1 Traffic Management

Arguably one of the most important control mechanisms in the dynamics of traffic modelling. Traffic lights are loaded directly from OSM using an actuated pattern. Actuated traffic lights switch phases based on time gaps between successive vehicles. This allows for better distribution of green-time among phases and affects cycle duration in response to dynamic traffic conditions.

Traffic lights are coordinated and grouped by joining nearby junctions. The default 4-arm intersection layout will be used, this includes; a straight phase, a right-turning phase (for a dedicated lane), a straight phase orthogonal to the first one and a right-turning phase orthogonal to the first one (for a dedicated lane).

Bus and taxi routes are explicitly modelled including bus and taxis stop at their dedicated locations.

Re-routing devices are a mechanism in SUMO that allow us to model on-the-fly rerouting of vehicles based on current conditions within the simulation.

5.1.2 Driver Behaviour

The car-following model default in SUMO is called carFollowing-Krauss (Krauß, 1998). This is a modification of the model defined by Stefan Krauß in his paper Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics. Vehicles can drive as fast as possible while maintaining perfect safety; always being able to avoid a collision if the leader starts braking within the acceleration bounds of both the leader and follower. A Euler-position update rule is used at each simulation step (a method solving the first order first degree differential equation with a given initial value). Drivers can change lanes while driving. A driver's impatience is measured and grows whenever a driver has to stop unintentionally. This measure is used to represent a driver's willingness to impede vehicles with higher priority – this effectively means that drivers with high impatience will use any gap that is safe while avoiding collisions even if other vehicles have to break as hard as they can.

5.1.3 Mesoscopic Model

The MESO model is a configuration of SUMO that runs up to 100 times faster than the microsimulation but still accepts all the same inputs. The decision to use this configuration in the simulation was made in order to ensure that the study area could cover a larger area while still having reasonable processing time per iteration. Over and above this, the MESO model is more tolerant of network modelling errors that

may be present in the network configuration or the traffic light signalling which means the user of the modelling pipeline need not worry about significant edits to the underlying road network for the simulation to run correctly. This speed increase is possible since traffic is modelling as a series of queues along 100m segments of network based on edges rather than lanes.

Passenger road traffic is modelled explicitly. We do not explicitly model multi-modal routing apart from passengers taking buses or taxis and walking the first or last part of the trip. Due to the highly detailed nature of the model it is not practical to run a large-scale simulation as the output emissions files will be many 100s of Gigabytes. Over and above this it would take a considerable amount of time to calibrate the model equilibrium which requires iterative processing steps.

5.1.4 Data Collection

The bulk of the data used in this study is data freely available in the public domain. The only proprietary data sources leveraged include a POI dataset (to improve the quality and completeness of business locations loaded in the simulation) as well as a historical traffic patterns dataset which has been used to evaluate aspects of how closely the simulation behaves in accordance with real world measurements.

All datasets were used as provided with minimal transformation apart from clipping and saving to different formats suitable for the modelling exercise. The characteristics and sources for the data used in this study are detailed in table 2.

5.1.5 Data Sources

Table 2 Data sources required by the modelling pipeline (workflow) that was developed for this study

Dataset	Characteristics	Source
Road Network Data	High Accuracy Must include detailed descriptive attributes such as • Road class types • Road name • Road Type • Speed Limits • Travel Direction • Linked topology (through graph-based node/edge model) • Number of lanes • Lane categories	Open Street Map Overpass API
Emission Reference Data	HBEFA v3.1-based pollutants by vehicle type, including	HBEFA/SUMO

Dataset	Characteristics	Source
	• CO_2 • CO • HC • NO_x • PM_x • fuel consumption	
Vehicle Population Statistics	Reference statistics of the number of registered vehicles per class per province	eNaTiS
Population Demographic Information	Population statistics broken down by age-group and household.	Statistics South Africa Census 2011 & Community Survey 2016
Census 2011 Sub Places	Geographic data describing the physical extent of the aggregated census units of geography	Statistics South Africa
MyCity Bus Schedule	Descriptive characteristics of the bus schedule for the public transport system in Cape Town	MyCiti Web Site
Dwelling Framework	Point dataset of each dwelling across the study area. This can be used to help calibrate origin-aspects of the demand loadings.	Statistics South Africa Dwelling Framework December 2017
School Data	Detailed information about schools including: • type • number of children • geographic location	South African Department of Education 2018
Business Location Data	Points of Interest representing business locations including classification types.	Here Technologies Q3 2018
Historic Traffic Patterns	Aggregated historical GPS-based traffic conditions in 15-minute time slices for the past two years	Here Technologies Q3 2018
Bus Data	Routes and stops representing the public transport bus network in the study area	City of Cape Town Open Data Portal (2017)
Taxi Data	Routes representing the informal taxi network in the study area	City of Cape Town Open Data Portal (2017)

5.2 Pre-Pipeline Data Processing

The following steps were performed to prepare several datasets for the automated pipeline (processing workflow). This is intentionally minimal in nature as we wanted to allow the model to be able to handle as much of the data processing as possible.

The Dwelling Framework is a 3.8GB Esri File Geodatabase containing a point file representing each household with several attributes. Keeping only the location attributes, we exported this to a flat file for consumption by the model.

The Business POIs come from a 20.3GB Esri File Geodatabase of Africa and the Middle East. The POIs falling in the study area were selected and exported to a GeoPackage.

The taxi routes do not have accompanying stops or stations as these are very informally defined – in reality taxis stop and start where passengers require them to do so in South Africa. To simulate this behaviour, we generated points every 500m along each taxi route and excluded points along highways where this type of stopping is not possible. This output was then used as the taxi stop input.

Digitisation of Traffic Lights – it became apparent early on that having accurate traffic light information would be important in ensuring that the simulation behaved correctly. Unfortunately, most traffic lights were either not digitized or incorrectly digitised in Open Street Map for the City of Cape Town.

By extracting major intersections and pedestrian crossings from the road network and using a 5km x 5km grid of the study area we started with an exercise to digitize every traffic signal across the city using the OpenStreetMap iD editor.

Figure 2 shows a screenshot of the OSM iD editor window and figure 3 shows the point locations of all traffic lights that were digitised.

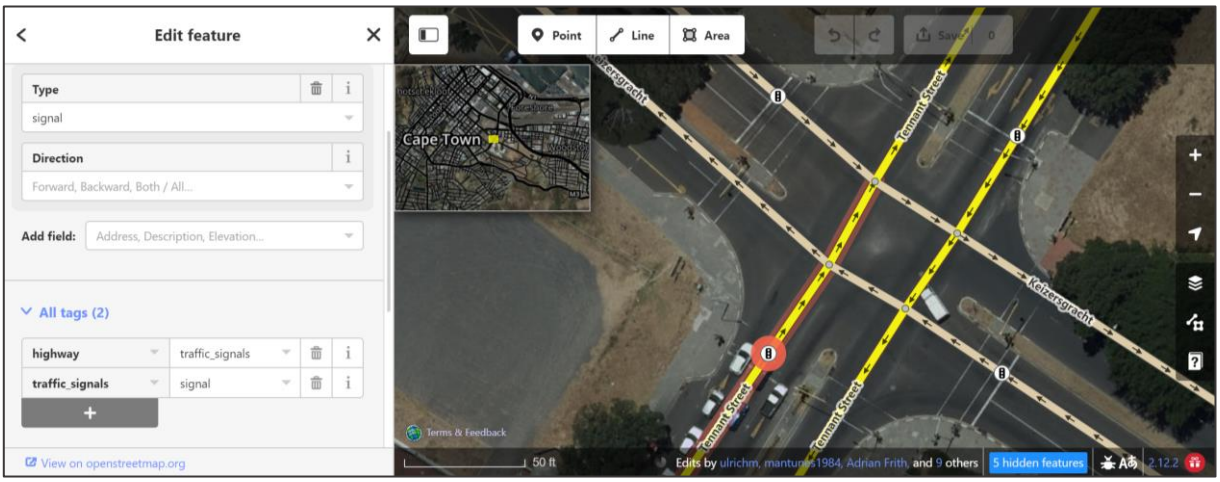

Figure 2 Screenshot of OSM iD Editor showing the feature editor capability used to digitise traffic signals

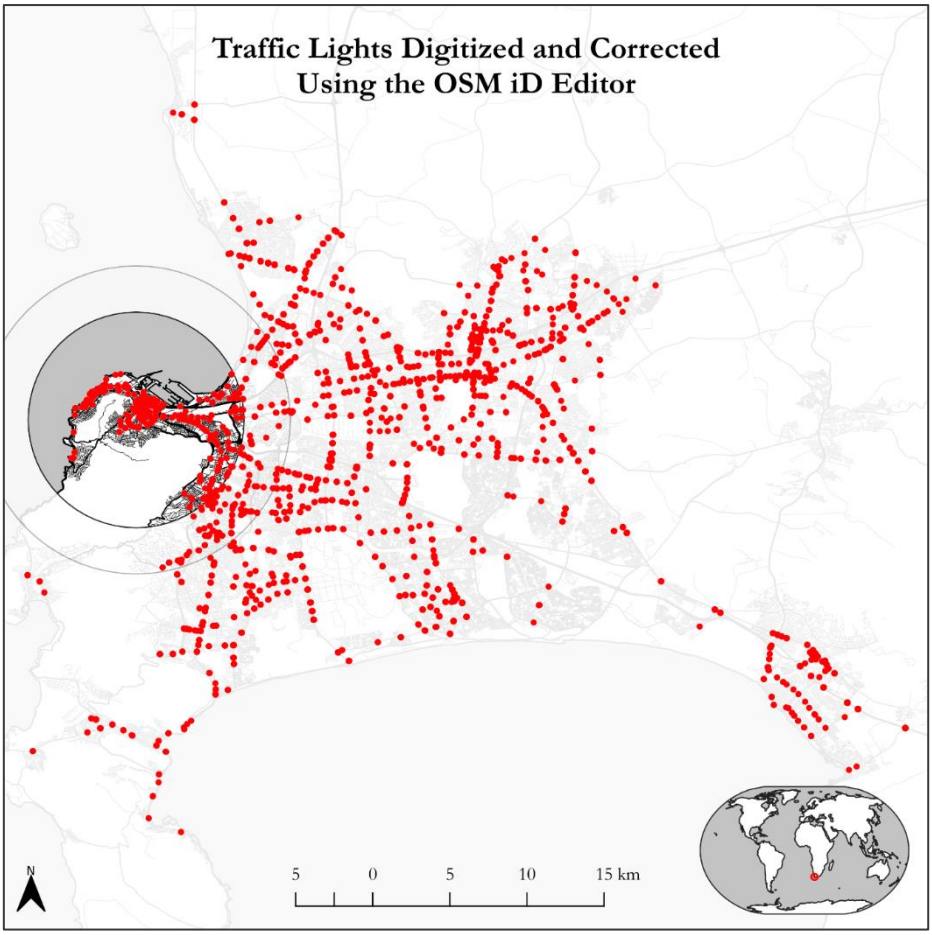

Figure 3 Locations of individual traffic signals digitised or corrected for simulation purposes

Vehicles are assigned to routes across a distribution per vehicle type. The following distributions were created for the model:

Default Type

As can be seen in figure 5, this is comprised of 6 petrol vehicles, 6 diesel vehicles & 1 alternative fuel type vehicle. Euro 6 is the latest emission standard available in HBEFA3. This represents the 2014 standard for light passenger and commercial vehicles. Since we were not able to attain a dataset that describes the relative age of registered road vehicles for the study area, the model shown in figure 4 was assumed based on year-of-manufacturing linked to the HBEFA classifications. The model was applied to both the petrol and diesel vehicles and a very small percentage of vehicles were given the probability that they run on an alternative fuel source (such as electricity). The distribution closely follows a second order polynomial with a bias towards the EURO6 standard. Vehicles are twice as likely to be assigned the EURO6 standard than the EURO3 standard of 2000.

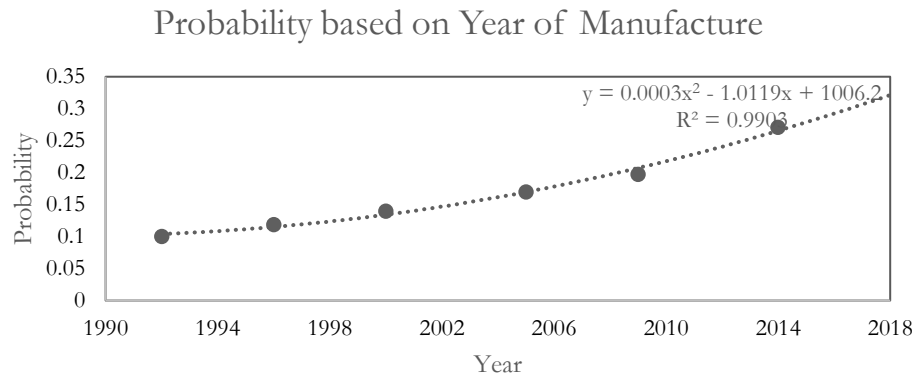

Figure 4 Year of manufacture probability of HBEFA class assignment for vehicles within the simulation

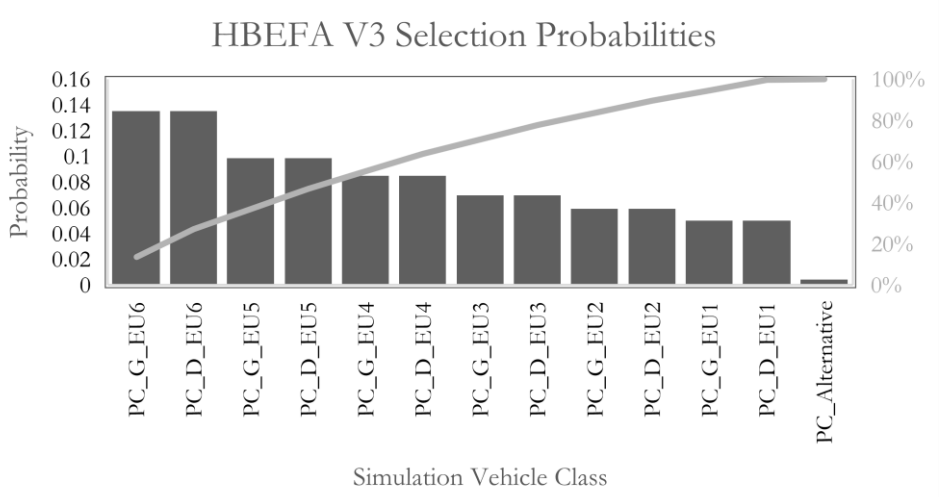

Figure 5 HBEFA selection probabilities by class for vehicles within the simulation

Random Type

This is a replica of the default vehicle type, but the vehicles are assigned unique IDs, so we know that they form part of the random set of generated trips which fall outside of the demand-derived trip set.

Bus Type

Only one bus type was defined using the generic HBEFA3 Bus type.

Taxi Type

Three taxi types are modelled using the light duty vehicle diesel standard: 20% of class EURO4, 60% of class EURO5, 20% of class EURO6.

5.3 Pipeline

An automated pipeline was developed to be able to generate the synthetic trip data for anywhere in the study area. The rationale behind this model is that, although this study is being conducted for the area around the City of Cape Town CBD, in practice a user would be able to input a centre point for any area to build a complete simulation in an automated manner utilising the methodology documented in this

paper. This makes the work highly reusable with minimal reconfiguration – the main changes required would be to source datasets to ensure all the elements are properly covered within the new extent which makes it transferable to other region outside of the selected study area.

Figure 6 illustrates the various steps included in the processing model that was coded. The model is aligned to the requirements of both SUMO itself and the ACTIVITYGEN tool that allows a user to generate synthetic trip data for a given study area from aggregate population statistics. The pipeline produces most of the inputs required by SUMO and ACTIVITYGEN in a dynamic, programmatic manner.

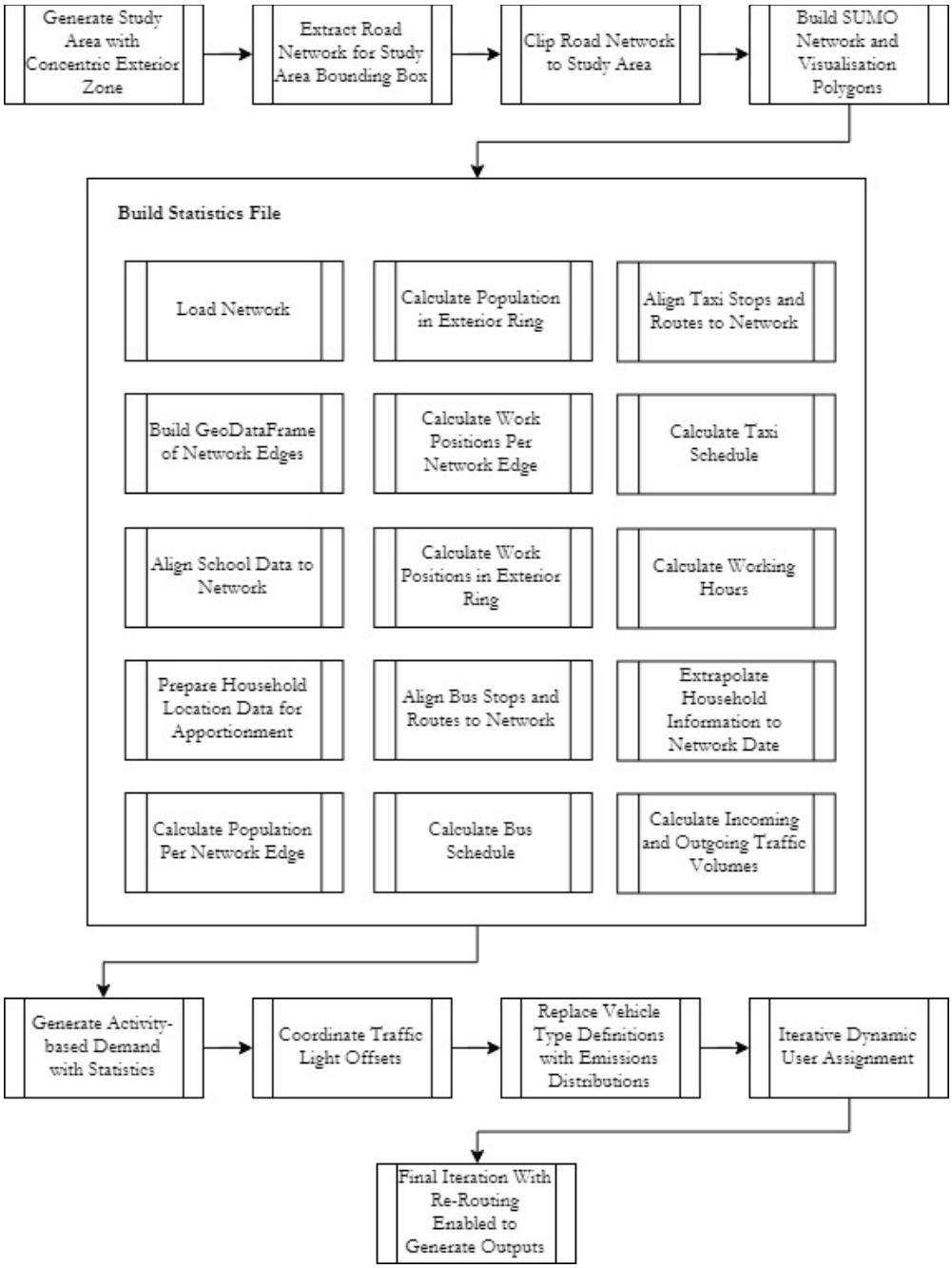

Figure 6 Steps inside the automated pipeline that generate the required input datasets for the simulation of urban mobility utilising derived synthetic trip data

Each of the steps in the methodology will be highlighted in further detail throughout section 5.3.

5.3.1 Generate Study Area with Concentric Exterior Zone

This function accepts a centre point (in WGS84 Geographic Coordinates), the study area radius (in meters) and a concentric boundary ring radius (in meters). The inner and outer rings are illustrated in figure 7. The point is projected to the Albers Equal Area Africa projection. The concentric buffers are created and the output re-projected back to WGS84 (EPSG:4326). The internal ring is written to a file that conforms to the OSM POLY forma which will be used later to clip data downloaded from OSM. The interior and exterior rings are then written to Shapely data types and stored within a GeoPandas GeoDataFrame for use elsewhere in the script as required. For this paper we are using a study area with a radius of 7km from the CBD with an exterior ring of 3km.

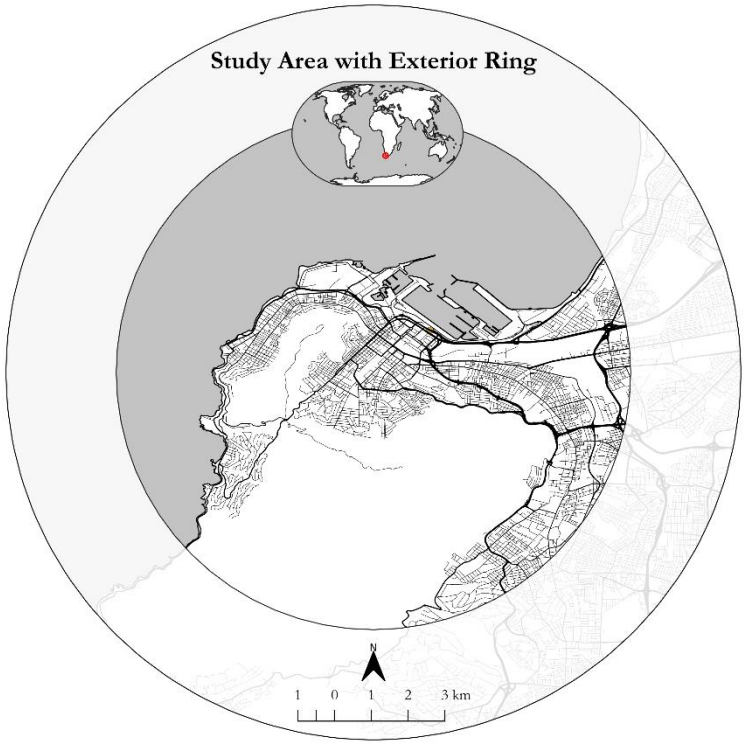

Figure 7 Generated study area and exterior ring showing the road network used in the simulation

5.3.2 Extract Road Network for Study Area Bounding Box

Data is extracted from Open Street Map using the Overpass API. The API accepts a bounding box for data extraction. The bounding box for the study area is calculated and passed to the API. The data that is returned is saved to the OSM *.osm.xml format.

5.3.3 Clip Road Network to Study Area

In order to clip the returned road network a utility called *OSM Convert* is used. The POLY file generated earlier is used to clip the network. Road segments that intersect the boundary are kept.

5.3.4 Build SUMO Network and Visualisation Polygons

The Open Street Map extract needs to be converted into a SUMO network. Using the SUMO NETCONVERT module within the *osmBuild.py* convenience script included with SUMO, the network is built.

5.3.5 Build Statistics File

SUMO includes a utility called ACTIVITYGEN that can generate trip definitions from certain input statistics to effectively generating a synthetic population from high level inputs.

These statistics include the following characteristics:

General

High level attributes about the city including total inhabitants, households, children age limit, retirement age limit, probability for an adult to own a car, the unemployment rate, the distance someone is willing to travel on foot, the total incoming traffic and outgoing traffic. Several of these are calculated in the pipeline, apart from the following: Car ownership probability – calculated from city-level eNatis statistics of total vehicle ownership in the city minus buses, trucks, motorcycles and 'other'. The max foot distance is set at 500m.

Parameters

These settings impact the behaviour of the population within the generated 'city'. Preference for someone to take public transport (set at a 0.6 likelihood given they have the option of taking a taxi or a bus within the foot distance limit). Mean Time per kilometre in the city (set at 360s). Free time activity use of personal vehicles (0.15). Proportion of uniform random traffic that should be generated relative to the total traffic demand (0.2). Departure variation to model human behaviours of departing against a fixed schedule (300s).

All other components of the Statistics file are generated in the pipeline.

These include Population age distribution; Work hours distribution; Number of people living on each road; Number of job positions on each road; City gates – roads at the edge of the network that generate incoming and outgoing traffic calculated from the outer ring; School positions and characteristics; Bus Lines, Stations and Schedules.

In order to include taxi data we loaded these as a subtype of the bus data type with specific identifiers that make them distinguishable as taxis in order to properly load their vehicles types in the simulation since ACTIVITYGEN does not cater for taxi trip generation.

Using this method for generating trips we are not explicitly covering certain types of traffic. Through traffic from outside the simulation to outside the simulation, i.e. people moving through the study area as part of a longer journey; Traffic for specific businesses, e.g. deliveries as well as tourist traffic.

However, since we have specified that 20% of the trips should be generated randomly, some of these traffic scenarios will implicitly be a part of the simulation and covered by this added 'randomness' added.

5.3.6 Load Network

The SUMO network is loaded using the sumolib Python library. This is used in several places later.

5.3.7 Build GeoDataFrame of Network Edges

This is one of the utility functions in the pipeline. The GeoDataFrame is used to assist in visualizing aspects of the network when generating the statistic file.

5.3.8 Align School Data to Network

The school data is loaded and clipped to the study area, see figure 8 for schools in the study area extent. Geometries are created from the included coordinates. Based on the school types, the age brackets per school are calculated. All schools are given an opening time of 6AM and a final closing time of 5PM. Using the loaded SUMO network, the schools are linked using linear referencing to the part of the road network that they belong to.

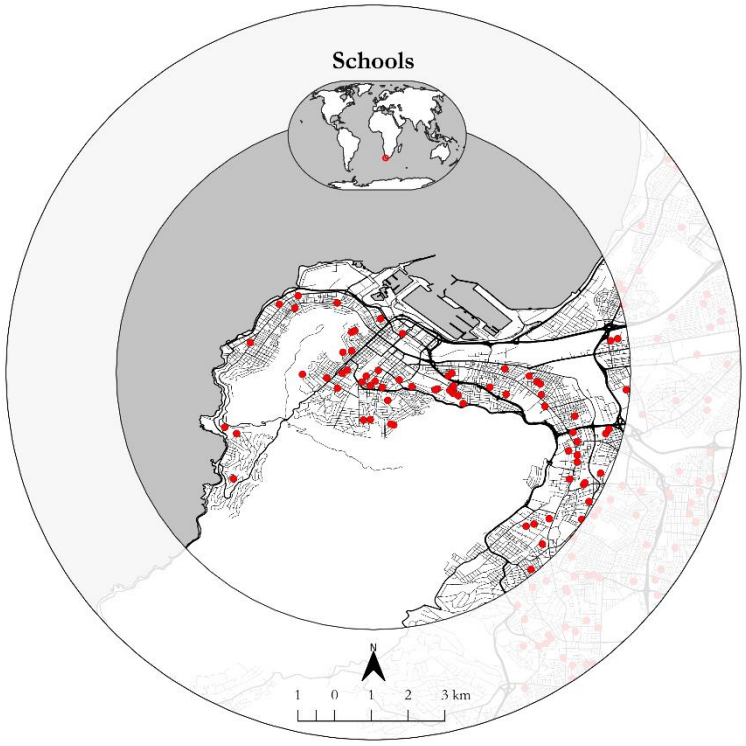

Figure 8 Schools modelled in the simulation

5.3.9 Prepare Household Location Data for Apportionment

The dwelling framework is loaded and clipped to the study area. For the study area, the households are linked to their road edges using linear referencing. For households in the exterior ring the raw household data is returned to be used later to be used in the calculation of incoming and outgoing traffic.

5.3.10 Calculate Population Per Network Edge

The study area dwelling framework linked to the road network is linked to StatsSA population statistics. A point-based apportionment is then applied to calculate the population per network edge. Since the population statistics are from 2011, we then apply a transformation based on the population growth at the city level between 2011 and the 2016 Community Survey and project to 2018. This is effectively a year-on-year increase of ~1.4%. The total population is then divided by the total length of the network edge since ACTIVITYGEN requires the per meter population.

5.3.11 Calculate Population in Exterior Ring

The exterior ring dwelling framework is linked to StatsSA population statistics. Since the population statistics are from 2011, we then apply a transformation based on the population growth at the city level between 2011 and the 2016 Community Survey and project to 2018. This is effectively a year-on-year increase of ~1.4%. The total population is return and will be used for incoming and outgoing traffic generation.

5.3.12 Calculate Work Positions Per Network Edge

The POIs are filtered to the study area. The POIs are then linked to the nearest network edge using linear referencing. Based on distribution of the POI categories a normalised value is created from 0 to 1 and multiplied by 50 in order to simulate the number of jobs available at each POI. At this stage the selection of 50 is somewhat arbitrary – it was decided based on the link between the number of working adults and number of jobs required to support them within this specific synthetic population being generated after several iterations. The total number of jobs is then divided by the total length of the network edge since ACTIVITYGEN requires the number of per meter work positions.

5.3.13 Calculate Work Positions in Exterior Ring

The POIs are filtered to the exterior ring. Based on distribution of the POI categories a normalised value is created from 0 to 1 and multiplied by 50 in order to simulate the number of jobs available at each POI. At this stage the selection of 50 is somewhat arbitrary – it was decided based on the link between the number of working adults and number of jobs required to support them within this specific synthetic population being generated after several iterations. The total number of jobs is then returned and will be used in the generation of incoming and outgoing traffic.

5.3.14 Align Bus Stops and Routes to Network

Bus stops and routes are loaded into the model, see figure 9 for the spatial extent of the bus routes in the study area. .Using linear referencing, the bus stops and routes are linked to the SUMO network The bus stops are linked to their associated routes.

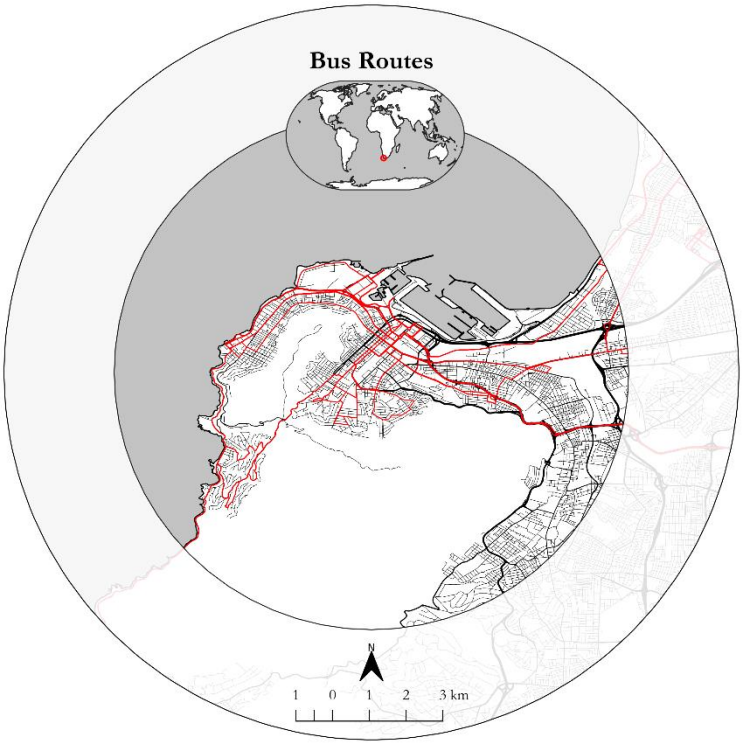

Figure 9 Bus routes modelled in the simulation

5.3.15 Calculate Bus Schedule

A default maximum end-to-end trip duration is set at 3000 seconds (50 minutes) as per the MyCity bus schedule. The rate of a bus returning to a stop is set to a default 20 minutes. All routes start at 6AM. All routes end at 10PM. This is somewhat of a crude schedule. Additional time in future iterations should be spent on more accurately modelling the schedule.

5.3.16 Align Taxi Stops and Routes to Network

Taxi stops are loaded into the model, see figure 10 for the taxi routes that traverse the study area. Using linear referencing, the taxi stops are linked to the SUMO network.

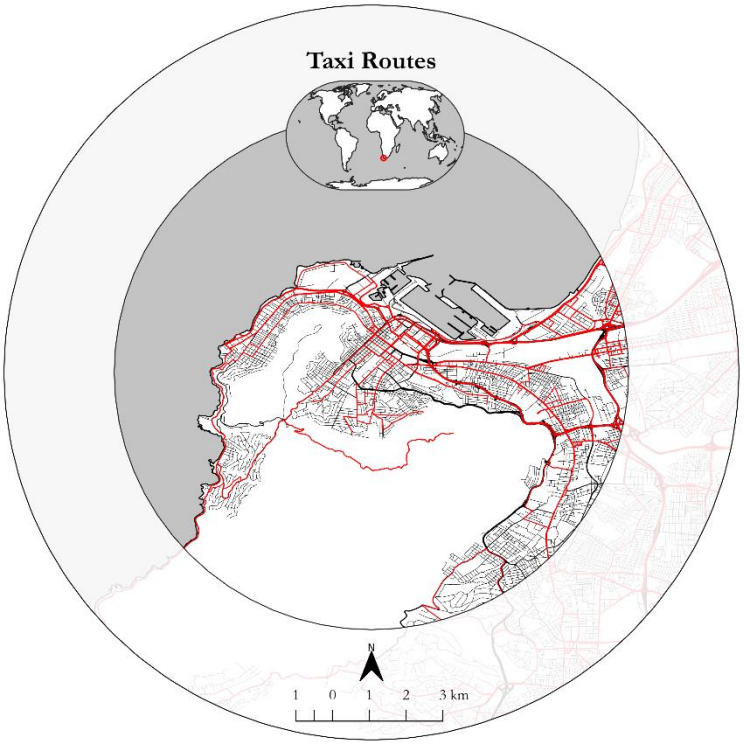

Figure 10 Taxi routes modelled in the simulation

5.3.17 Calculate Taxi Schedule

A default maximum end-to-end trip duration is set at 3000 seconds (50 minutes). The rate of a taxi returning to a stop is set to a default 30 minutes. All routes start at 6AM. All routes end at 10PM. This is somewhat of a crude schedule. Additional time in future iterations should be spent on more accurately modelling the schedule.

5.3.18 Calculate Working Hours

A normally distributed set of hours for arriving and leaving work are generated. Morning hours extend from 6:30 to 10:00. Afternoon hours extend from 15:30 to 19:00.

5.3.19 Extrapolate Household Information to Network Data

StatsSA 2011 household information is loaded and clipped to the study area. The total households are calculated and extrapolated to 2018 figures based on the growth between 2011 and 2016 which is effectively a rate of ~2.7% year-on-year.

5.3.20 Calculate Income and Outgoing Traffic Volumes

The incoming traffic is calculated as:

$$Abs(\frac{exterior\ ring\ jobs - \ study\ area\ jobs}{exterior\ ring\ job}) \times boundary\ population$$

The outgoing traffic is calculated as:

$$Abs(\frac{study\ area\ jobs - exterior\ ring\ jobs}{study\ area\ job}) \times boundary\ population$$

5.3.21 Generate Activity-based Demand with Statistics

Using the ACTIVITYGEN module trips are generated that describe the synthetic population using the Statistics file that is dynamically generated using the pipeline. The output of this is a file containing 758 534 trips. Some of these trips (particularly the uniform randomly generated trips) are invalid in that they may link parts of the network that are not connected in a way a vehicle may cover the trip and will be dropped in the simulation. Figure 11 shows the number of generated trips by hour of the day.

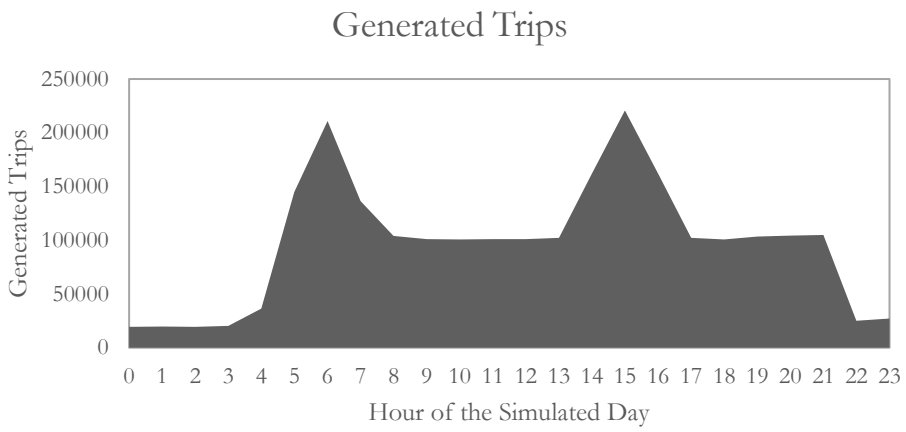

Figure 11 Trips generated by ACTIVITYGEN using the automated pipeline

5.3.22 Coordinate Traffic Light Offsets

Using the tlsCoordinator script supplied by SUMO, the traffic signals are coordinated. This script should be run every couple of iterations to ensure the coordination is in line with the simulation trip definitions. Figure 12 illustrates a traffic-controlled junction within the simulation.

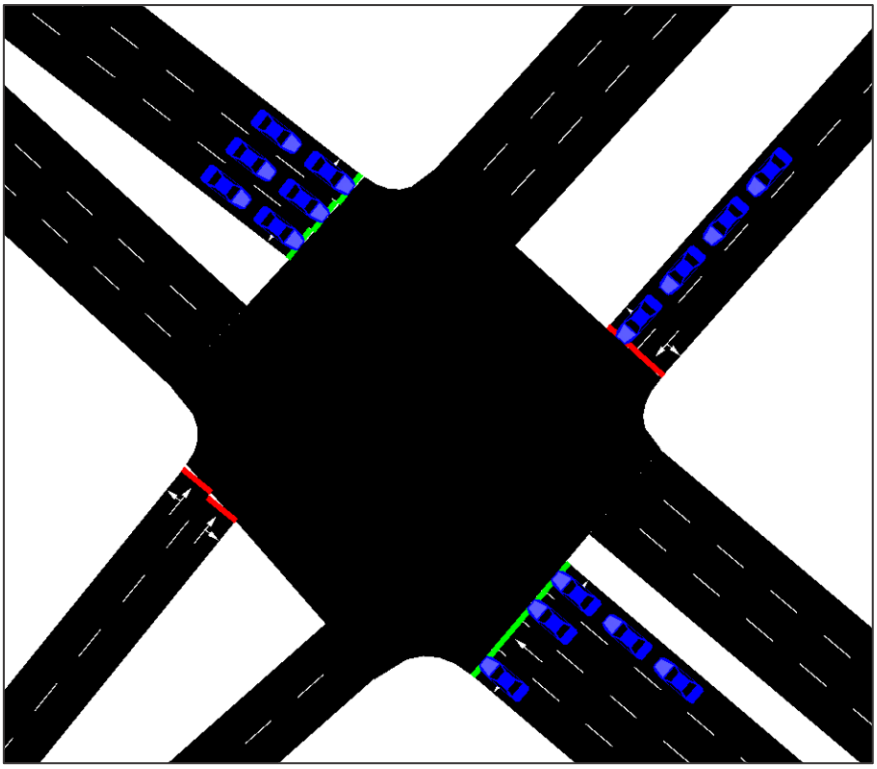

Figure 12 Vehicles at a traffic light-controlled junction within the simulation

5.3.23 Replace Vehicle Type Definitions with Emissions Distributions

The ACTIVITYGEN output by default includes basic vehicle definitions. Since we already have a pre-defined vehicle definition file that we have created with their associated emission classes we have to remove these conflicting definitions from the output of ACTIVITYGEN. Additionally, at this step, taxis are assigned their specific vehicle class (since ACTIVITYGEN assigned them to the bus type). Some additional cleaning of the file occurs – there may be a bug in ACTIVITYGEN which seems to create a subset of trips with no proper reference IDs that impact the simulation – these trips are explicitly removed.

5.3.24 Iterative Dynamic User Assignment

At this point, the trip file holds the most basic definitions required to describe a trip. Typically, this includes start and end points with some mandatory road edges along the route. When loading these trips naively into the simulation all vehicles will take the fastest path available under the assumption that they are the only vehicle in the network. This results in bottlenecks and traffic jams because of this unrealistic behaviour.

The problem of determining suitable routes that consider travel times in a traffic-loaded network is called user assignment. In SUMO there are effectively two methods to achieve this.

One technique enables us to attain a dynamic user equilibrium through iteratively calling DUAROUTER and SUMO. Using this technique weights from previous iterations are used to generate a set of route

choices per trip which are then probabilistically selected in the next simulation step. The Gawron route-choice algorithm is selected. (Gawron, 1998). The input is a weight (**w**) on the edges of the network and a set of routes (**R**) where each route (**r**) has an old cost c_r and an old probability p_r (from the last iteration) and needs a new cost c'_r and a new probability p'_r. The Gawron algorithm computes probabilities by selecting from a set of alternative routes for each driver observing: the travel time along the used route in the previous simulation step; the sum of edge travel times for a set of alternative routes and the previous probability of selecting a route.

Since we are trying to model real traffic patterns, a very important requirement that had to be added at this step is that vehicles that miss their intended departure time within 1 minute are excluded from the simulation. This is in order to prevent excessive jamming at unusual times of day as a result of overloading the scenario with more trips that the network can support.

It also became apparent that too many trips are generated for the simulation out of ACTIVITYGEN – when trying to push these through the simulation it adds unnecessary computation time and results in unrealistic behaviours. In order to overcome this the trips are only allowed to enter the simulation if they can depart on time and an additional hard cap of 10 000 vehicles running at any one point in time was set in order for the simulation to not create excessively long running high traffic scenarios during peak time.

By default, we use Dijkstra's shortest path routing algorithm.

The progress of each iteration is tracked and compared against previous iterations using the tests detailed in section 6. Once there is no measurable improvement to the equilibrium in the system the iterative process is stopped.

5.3.25 Final Iteration with Re-Routing Enabled to Generate Outputs

After we are satisfied with the simulation equilibrium, the final iteration is modified to include several other characteristics.

Re-routing devices are added to 20% of the simulated vehicles probabilistically. This means that when we run this iteration again 1 in every 5 cars will adapt within the simulation to changing traffic scenarios in simulation time. This is meant to simulated people either using GPS devices for routing or simply choosing different routes based on observed conditions.

The simulation is also set to output the emissions per vehicle in every simulation step, i.e. all emissions for all vehicles are recorded for every second of the simulation.

5.3.26 Evaluation of Outputs

Before we dive into the emission calculations, it is necessary that we evaluate the performance of the simulation across as many measurable metrics as possible in order to ascertain whether we have created a valid traffic scenario.

After considerable experimentation through trial and error, the following techniques were decided upon in order to evaluate the modelled outputs.

To validate some of these measures we used the historic traffic patterns dataset from Here Technologies. This represents travel speeds for segments of roads in 15-minute time slices aggregated over two years. Specifically, in this analysis we are evaluating the output of the simulation against an 'average Monday' for the convergence of the dynamic user assignment.

These metrics include two types, observational and comparative. The observational metrics simply need to stabilise into an expected pattern while the comparative metrics provide a way for us to evaluate the performance of the model as it is iterating through each step against historical aggregate data.

5.3.26.1 Observational

a. Volume of loaded vehicles

Since we are trying to simulate as close to a normal day as possible, some of the parameters that build the simulation such as the maximum depart delay have been configured in such a way so ensure that additional vehicles are not loaded into the simulation that result in unrealistic traffic scenarios at unusual times of the day due to excessive jamming.

b. Typical traffic pattern

Linked to the above metric, traffic should effectively spike during the morning and afternoon rush hours with a low, steady state the rest of the day. In this metric we also observe the number of vehicles 'halting' or in a totally jammed state.

c. Mean Speed

The average speed at each time step will give an indication of the load on the network as well as how jammed the scenario is at that point in time. This should be closely linked to the number of vehicles in the simulation at that point in time.

d. Mean Travel Time

Mean travel time is an interesting measure to evaluate as it behaves somewhat unintuitively in certain cases when assessed across the entire study area. For example, during the peak traffic times more vehicles couple

be driving on roads with higher speed limits which in turn means that travel times are decreased in certain cases.

5.3.26.2 Comparative

a. Linear Regression

A simple linear regression is calculated to assess how closely the simulation speeds mirror real world speeds. A perfect model of an average Monday would result in a perfectly linear relationship between the simulation outputs and the historical aggregates.

b. Relative Entropy (Kullback-Leibler divergence)

Relative Entropy is a measure in information theory that represents a way of measuring the average rate at which information is produced by a random source of data by specifically measuring how different one probability distribution is from another. We measure the relative entropy specifically of the travel speeds at each 15-minute time step against the historic traffic patterns. This gives us a very clear indication of relative information loss or gain through simulation time as the simulation behaves like the expected output.

c. Two-Sample Kolmogorov–Smirnov test

The KS test is a nonparametric test of equality comparing two continuous distributions of one dimension. The only assumption is that the two distributions are continuous – this test does not require the data to follow a normal distribution as the cumulative distribution is observed. The null hypothesis is that the two distributions are identical.

d. K-sample Anderson-Darling test

The Anderson-Darling test can be used to measure the agreement between two distributions. The null hypothesis is that the samples provided from the two distributions provided are in fact drawn from the same population. After these metrics have been measured we then produce the fundamental diagram of traffic as illustrated in figure 24.

5.3.26.3 The Fundamental Diagram of Traffic

This is a well understood way of evaluating the relation between traffic flux (vehicles per hour) and traffic density (vehicles per kilometre). (Immers and Logghe, 2002).

The fundamental diagram is an important tool for understanding the mechanics behind what is happening in the simulation for a number of reasons: the more vehicles are on a road, the slower their velocity will be; to prevent congestion and to keep traffic flow stable, the number of vehicles entering an area must be smaller or equal to the number of vehicles leaving the zone in the same time; at a critical traffic density

and critical velocity, the state of flow will change from stable to unstable & if one of the vehicles brakes in an unstable flow, the flow will collapse.

There are several methods for representing the fundamental diagram, we will be using the flow-density form. Flow-density diagrams are used to identify the traffic condition of a roadway. It clearly illustrates the intersection of free flow and congested traffic and can be used to illustrate the capacity of the road or network.

5.3.26.4 Emissions Calculations

The following emissions are tracked throughout the simulation per vehicle in milligrams per second:

i. CO_2 – Carbon Dioxide
ii. CO – Carbon Monoxide
iii. HC - Hydrocarbon
iv. NOx – Nitrogen Oxides
v. PMx – Particulate Matter

Fuel consumption is tracked throughout the simulation per vehicle in millilitres per second.

6 Results

6.1 Introduction

The following results represent the simulation after the 20th iteration including the re-router step. This 21st simulation step showed no measurable change over and above the simulation results from the dynamic user assignment as at the 20th iteration. This proves that the dynamic user assignment process was able to find an optimal equilibrium within the system which could not be further improved by allowing some of the vehicles to re-route themselves 'live' in the simulation.

6.2 Observational

6.2.1 Volume of Loaded Vehicles

Figure 13 shows that of the 346 089 valid trips generated – all of them were prepared for the simulation. Due to the max depart deviation cap as well as the maximum cap of 10 000 vehicles running at any point in time, only 222 601 (65%) were loaded into the simulation. The morning influx of vehicles into the city is clearly identifiable as the first slope increase with the afternoon exodus starting around 4PM. The small gap between the Ended vehicles (vehicles that have exited the simulation or not been allowed to enter the simulation) and the Loaded vehicles represents the increase in delayed travel times for certain commuters during peak traffic.

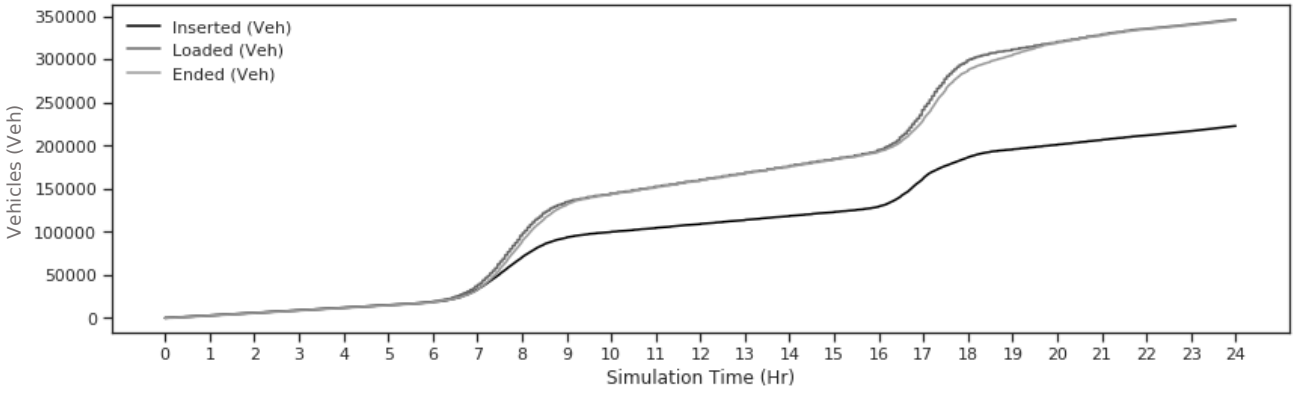

Figure 13 Vehicles inserted, loaded & ended in the simulation after Dynamic User Assignment and Re-routers were configured for the simulation

6.2.2 Typical Traffic Pattern

Figure 14 illustrates the total number of vehicles both running and halting in the simulation. A halting vehicle is a vehicle that is not moving at all due to congestion at that point in time.

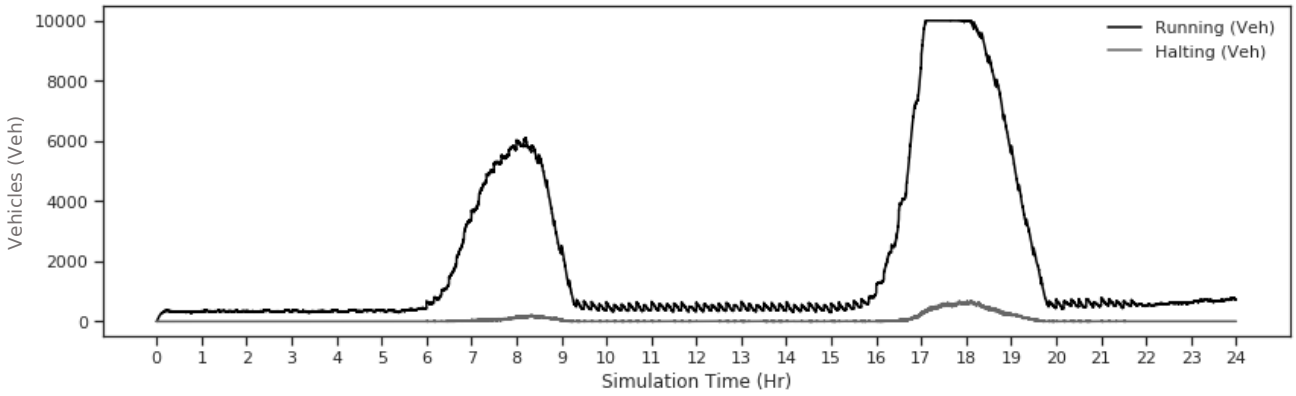

Figure 14 Traffic volume of running and halting simulated vehicles

The clearest distinction here is that the simulation can support fewer vehicles at equilibrium in the morning coming into the city when compared to the number of vehicles in the simulation during the afternoon peak period leaving the city. The hard 10 000 cap can be seen getting hit in the afternoon, implying that the simulation would be able to hold additional vehicles if possible. Interestingly, when the simulation was allowed to run unbounded by the hard cap, the same pattern of approximately 60% of the afternoon load being supported by the morning load was observed, i.e. the afternoon peak included 20 000 vehicles at its max with the morning peak including 12 000 vehicles at its max. The hard cap was a necessary limitation to maintain reasonable processing time as well as reasonable end times for peak traffic.

The fact that fewer vehicles can make their way into the simulation during the morning peak indicates that morning traffic in the study area would be considerably worse. The time period between 9AM and 4PM as well as between 8PM and 22PM includes the small number of randomly generated trips as well as personal trips as generated by ACTIVITYGEN.

6.2.3 Mean Speed

For each simulation step the mean speed of vehicles running in the simulation is recorded as illustrated in figure 15. As expected, the first few vehicles to increase the load on streets and highways with higher speed limits results in a spike at the beginning of the peak periods in terms of the average speed being travelled. The average speed quickly drops as the scenario becomes more congested. The drop in the morning average speed follows a linear pattern – this is since during this time cars are starting off their trips on uncongested roads. Due to a more concentrated volume of cars trying to leave the CBD in the afternoon, the average speed is reduced quicker during this time. The oscillating pattern observed during low traffic volumes is simply due to the way the simulation is adding and removing vehicles from the simulation.

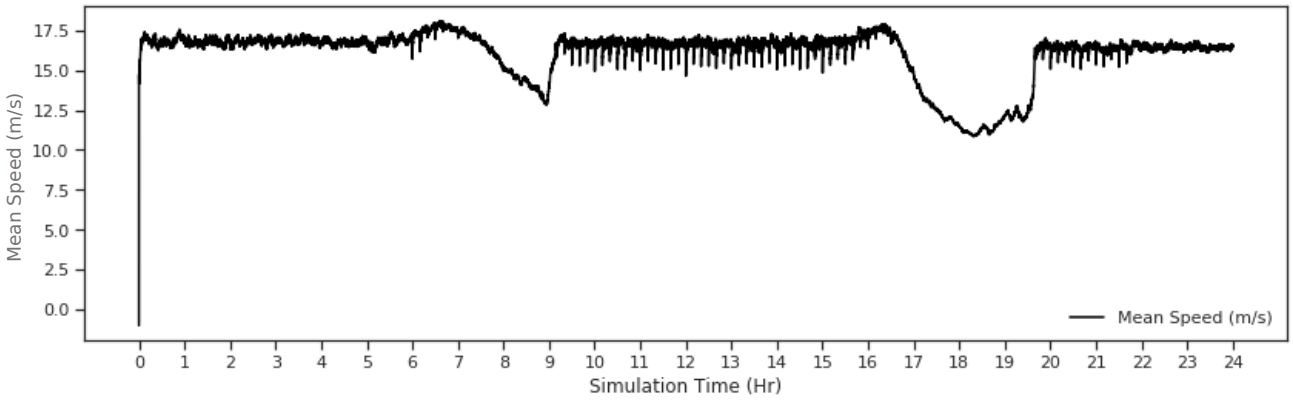

Figure 15 Average speed for all vehicles running in the simulation at each simulation step

6.2.4 Mean Travel Time

At the start of the simulation there are no vehicles running (midnight), therefore the travel time calculations in the first few simulation seconds are unrealistically underreported – this quickly stabilises to around 6 minutes and 40 seconds per trip until the morning rush where the average travel time decreases as shown in figure 16. This is due to several factors – more vehicles driving on higher speed roads as well as more vehicles conducting short trips (to school, on the bus, etc). After the afternoon rush the average travel time is the highest.

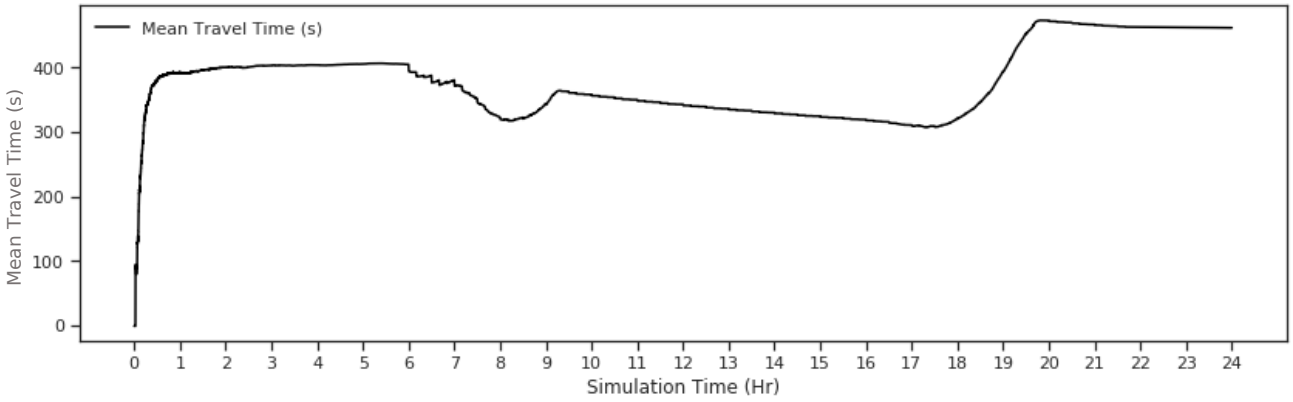

Figure 16 Average travel time for all vehicles running in the simulation at each simulation step

6.3 Comparative

6.3.1 Linear Regression

Since we are effectively comparing historical travel speed against the simulation travel speed for different steps in time, in a perfect simulation of the historical aggregate we should find this relationship to

approximate 1:1. It's clear with such low correlation coefficient (R^2) values in figure 18 that the model does not exhibit this pattern of behaviour and we can reject the null hypothesis that there is a linear relationship between historical travel speeds and the simulated travel speeds. Figure 17 shows the R^2 value and the intercept. The intercept in this type of modelling effectively holds the extra components of information not being told by simply looking at the travel speed. Ideally, the intercept should be very close to 0 with a high R^2 (accompanied by a low p-value) and should be stable throughout simulation time. It's clear that during peak traffic the relationship breaks down further. However, even during periods of low traffic volumes the relationship is not linear.

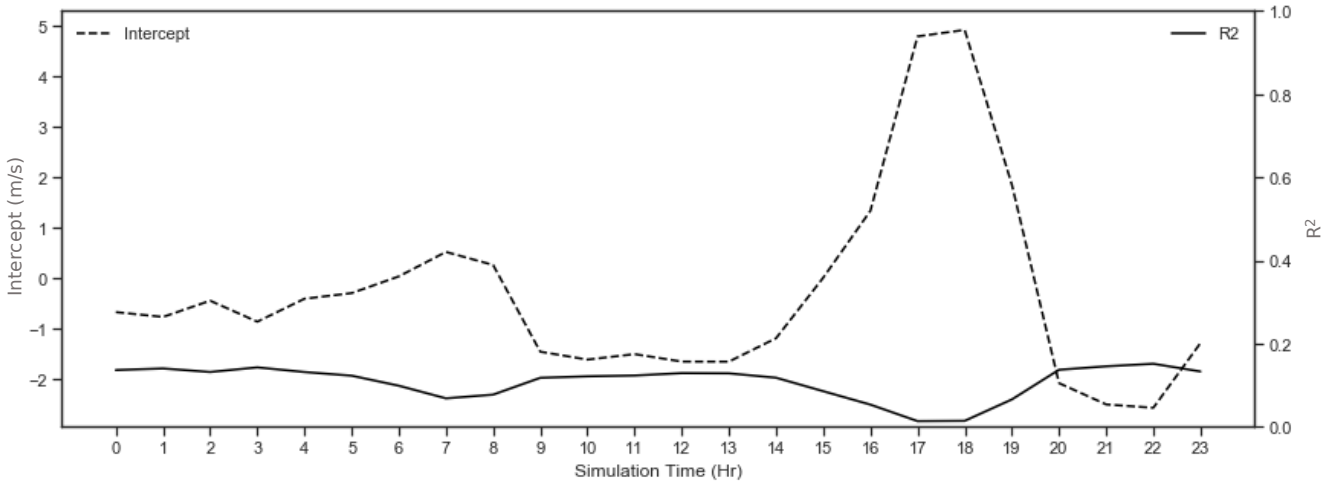

Figure 17 Coefficient of determination and intercept for linear regression model between historical & simulated travel speeds

Looking at the delta between the historical travel speeds and the simulation travel speeds (on the left of figure 18) across the full duration of the simulation we observe this relationship at a road segment level. The distribution is centred around -8m/s indicating that simulation is consistently running vehicles at approximately 8m/s faster than what is expected, even so with the current setup, we should expect the scatter plot on the right of figure 18 to show a linear pattern which is not evident.

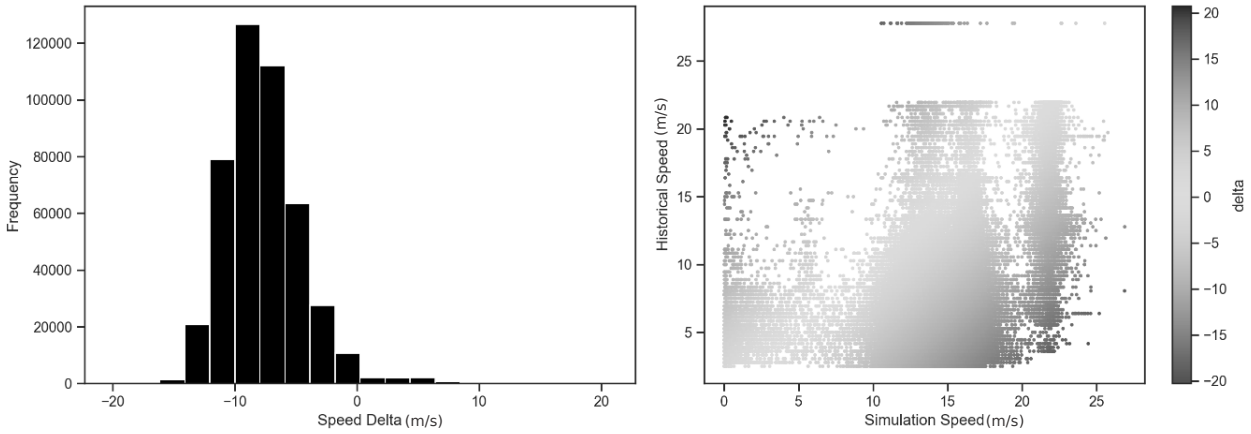

Figure 18 Speed deltas and scatter plot showing non-linear relationship between historical travel speeds and simulated travel speeds

We believe this has something to do with the way vehicle trips have been generated. In a future iteration of this study it would be worth exploring whether we could use the reduction of this delta as an objective function in order to treat this component of the simulation as an optimisation problem to better align the simulation to a specific historic condition. SUMO can leverage a weight file to assist in the determination of vehicle routes that are created. Using the historic data, we loaded a weight file into the simulation – the only noticeable difference it made was that the simulation was able to reach an equilibrium state in fewer steps, however it did not improve upon the expected linear relationship we were expecting from this test.

6.3.2 Relative Entropy (Kullback-Leibler divergence)

The important thing to note with figure 19 is the relative differences between the different times of the day. Relative entropy used in this way can tell us how much information is lost between the historic speeds and the simulation speeds. From the figure below, the period of maximum relative entropy occurs during the afternoon rush. Periods of the lowest relative entropy occur when there are very few cars on the road. Interestingly, the morning rush exhibits relatively low entropy indicating that the simulation could be a better reality fit for the morning traffic spike when compared to the afternoon period.

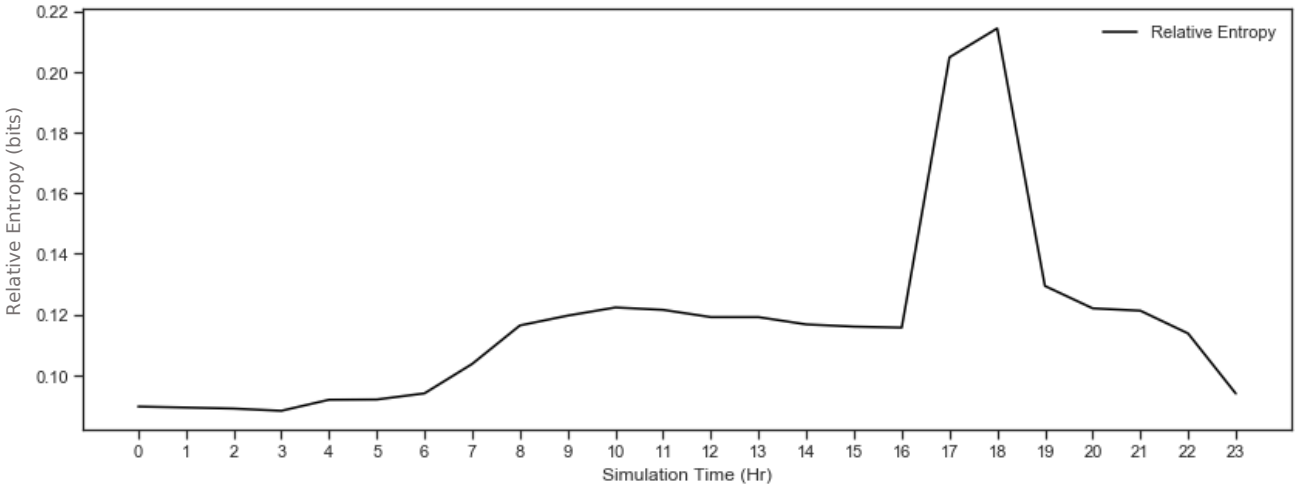

Figure 19 Relative entropy calculated through simulation time

6.3.3 Two-Sample Kolmogorov–Smirnov test

The KS test is commonly used to evaluate the null hypothesis that two distributions come from the same population. As illustrated in figure 20, with the p-value consistently at 0 we can reject the null hypothesis and conclude that these two distributions are different throughout the simulation time.

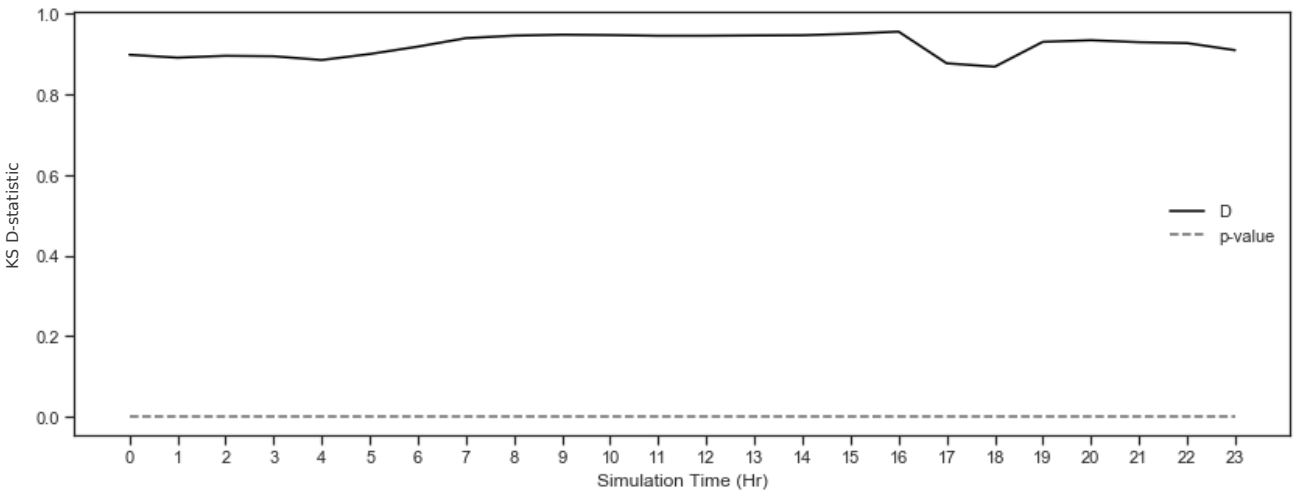

Figure 20 D-statistic and p-value for Kolmogorov-Smirnov through simulation time

This is made evident when we look at the cumulative distribution functions of all the historical speeds for the study area against the simulation speeds on the left of figure 21. For us to accept the null hypothesis of the KS test these two CDFs should be very much aligned. Even if we take the simulation speeds and adjust them by the average 8m/s we see that the distributions are still not the same as illustrated on the right of figure 21.

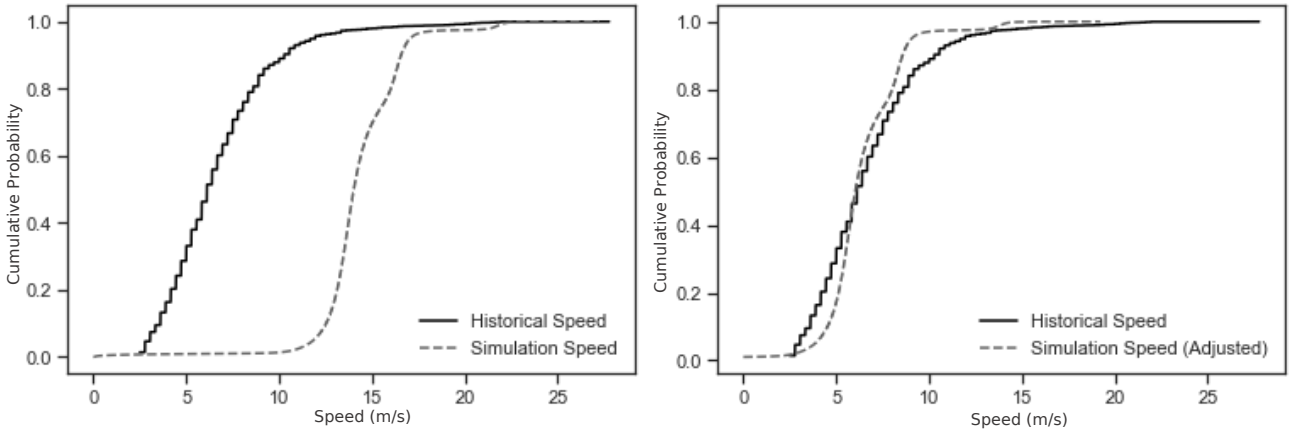

Figure 21 Cumulative distribution functions showing a misalignment between historical travel and simulation travel speeds

6.3.4 K-sample Anderson-Darling test

Like the KS test, the AD test (as illustrated in figure 22) provides a framework for us to either accept or reject the null hypothesis that the two samples come from the same population. The test statistic is many orders of magnitude higher than the 25% critical value throughout simulation time, therefore we reject the null hypothesis – the two distributions do not come from the same population.

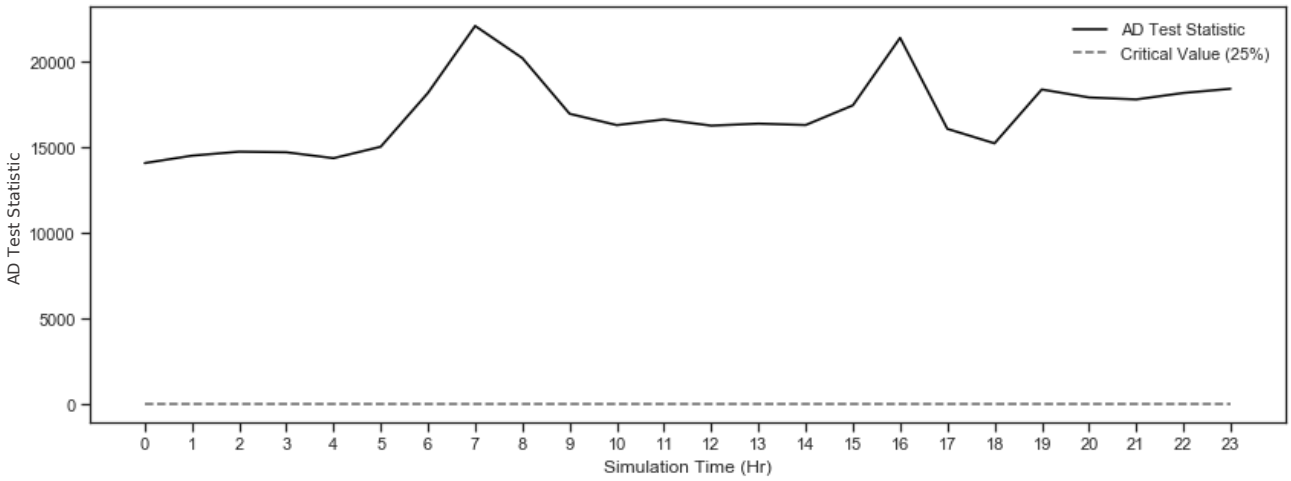

Figure 22 Test statistic and 25% critical value for Anderson-Darling test

What is interesting to note with this statistic is the shape around the afternoon peak period from 5PM where the test statistic seems to improve somewhat.

6.4 Fundamental Diagram of Traffic

Figure 23 represents the fundamental diagram of traffic for the simulation scenario in 3-hour intervals in flow-density form. This type of diagram is useful to determine at which point the optimum flow occurs.

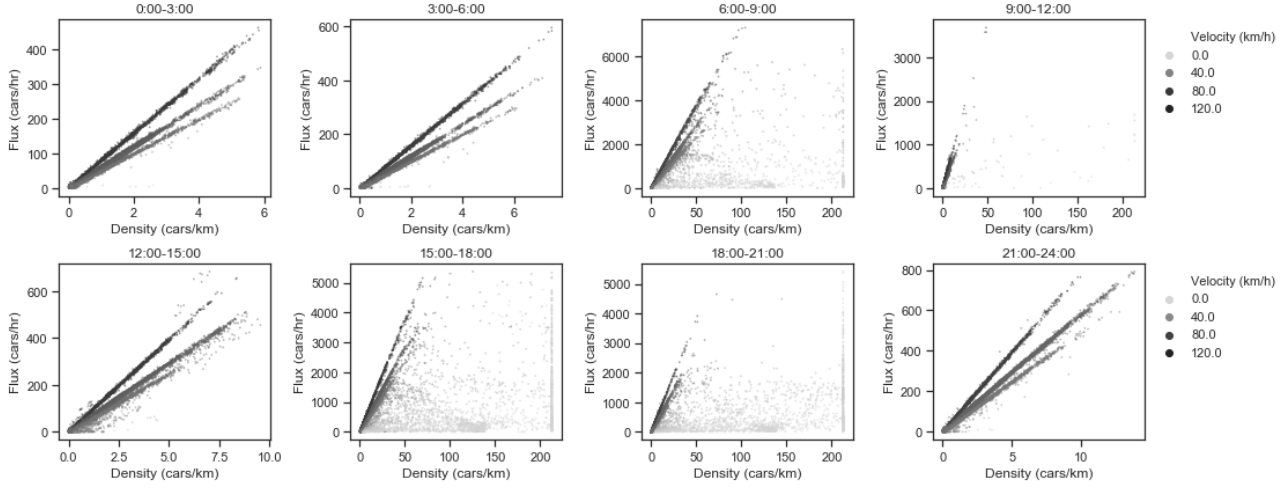

Figure 23 Fundamental diagrams of traffic flow in 3-hour increments across simulation time

Our fundamental diagrams represent the full simulation for the full network, in practice, this means that we have several fundamental diagrams for different roads all in one figure which we are evaluating in the aggregate.

This is made evident by the fact that there are a distinct set of lines in each plot in figure 24. Using the 00:00 – 03:00 example, there are three distinct collections of routes (portions of the network) supporting different volumes of cars for the same density. The diagrams with low density represent free-flowing, unbounded traffic and include the 03:00 to 6:00, 12:00 to 15:00 and 21:00 to 24:00 examples. The straight lines in each figure represent the free flow branch of the fundamental diagram.

The 6:00-9:00, 15:00-18:00 and 18:00-21:00 plots in figure 24 show many points of low velocity where vehicles are in a congested state. In an ideal flow diagram these would be more concentrated forming an arc as in the figure below. The fact that the vehicles do not form this arc means that vehicles are in a jammed state with severely hampered congested flows. An example of a true fundamental diagram without heavy congestion is illustrated in figure 24.

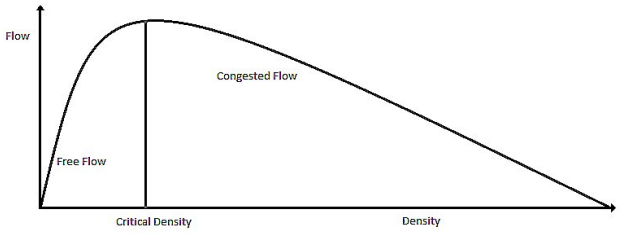

Figure 24 Expected fundamental diagram of traffic flow

The 6:00-9:00, 15:00-18:00 and 18:00-21:00 plots all exhibit a series of points forming a straight line up the right-hand side of the plot – this represents the effect of the default 'teleportation' option SUMO uses. When a vehicle has not been moving for 5 minutes it is teleported to an available location along its route to continue with its journey – this setting makes the simulation less sensitive to errors in the road network or traffic signal timings in jammed states but does leave artefacts behind in the simulation data which clearly impact the fundamental diagram.

6.5 Emissions

Generating the emissions data is a computationally expensive task which has been left for the very final stage. For our study area, the output file is 43GB in size and represents the emissions for every vehicle at each second of the simulation which covers a full day. An increased number of vehicles or increased simulation time will result in a significantly larger output.

In order to process this data a utility was developed that allows us to read portions of the file aggregated into one-minute slices into memory in order to generate the following plots in a piecemeal manner.

Table 3 shows the total emissions were recorded for the simulated day.

Table 3 Aggregated emissions for the full simulation runtime by vehicle type

Vehicle Type	CO_2 (kg)	CO (g)	HC (g)	NOx (g)	PMx (g)	fuel (L)
busstandard	3 469	6 003	1 544	27 385	654	1 480
passengeralternative	1 173	11 537	200	650	26	606
passengerdiesel	105 634	60 408	13 539	435 966	18 065	39 740
passengerpetrol	152 714	1 262 783	18 722	90 019	2 902	65 660
randompassengeralternative	433	2 757	64	228	8	224
randompassengerdiesel	33 073	18 265	4 223	123 442	5 468	12 444
randompassengerpetrol	46 903	167 051	4 494	25 917	633	20 165
taxi	16 688	1 835	714	57 369	66	6 279
Total	**360 087**	**1 530 639**	**43 502**	**760 976**	**27 822**	**146 596**

6.5.1 Emissions by Type

There is a very direct correlation with the plots in this section and figure 14 showing the number of running vehicles through simulation time. The plots in this section show the aggregated total emissions each measure is pumping into the atmosphere (or burnt in the case of fuel) through simulation time split by the generic vehicle types that were initially defined for the simulation.

An interesting observation here is that the bus and taxi schedules can be clearly seen through their corresponding emissions. Additionally, when the simulation was in its congested peak afternoon state, buses, taxis and the random trips were not able to enter the simulation on schedule and thus were skipped. Once they were able to start entering the simulation again they joined the congested network and subsequently increased their emission outputs relative to less busy times as expected.

The emission type that produces the most output by pure volume is carbon dioxide with the main contributors being the passenger petrol and diesel vehicles emitting over 500kg of CO_2 per second across the simulated network in the afternoon peak as illustrated in figure 25.

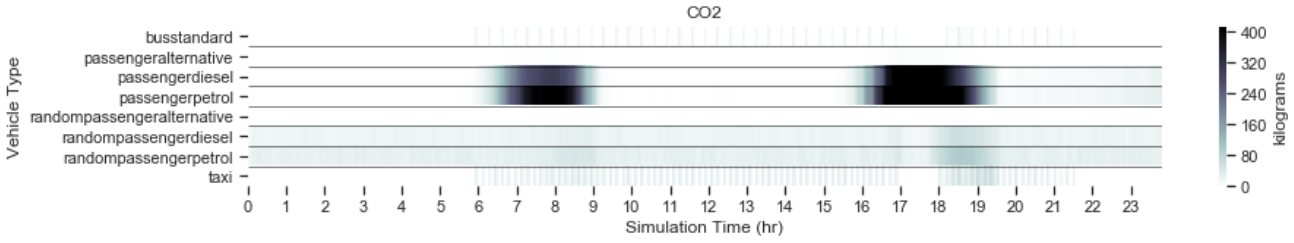

Figure 25 CO_2 Emission Heatmap

Petrol vehicles clearly seem to emit the most carbon monoxide through simulation time as shown in figure 26. Carbon monoxide is a toxic substance that reacts with certain aldehydes and nitrogen oxides producing photochemical smog.

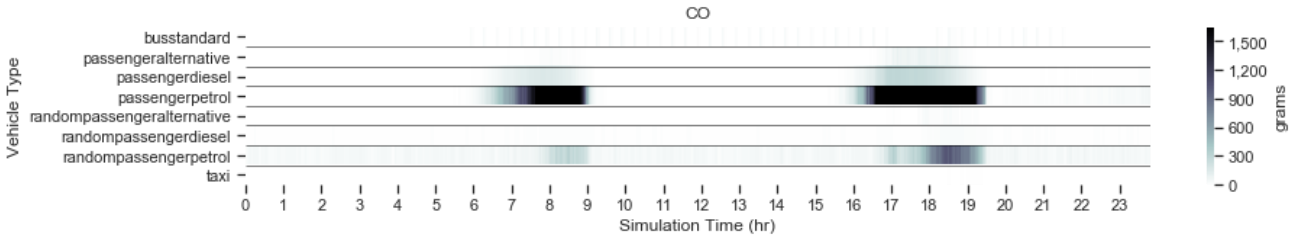

Figure 26 CO Emission Heatmap

Hydrocarbons are one of the lowest emissions by total grams dumped into the atmosphere as shown in figure 27. Hydrocarbons emitted in this way represent unburnt petroleum or diesel being emitted due to inefficient combustion.

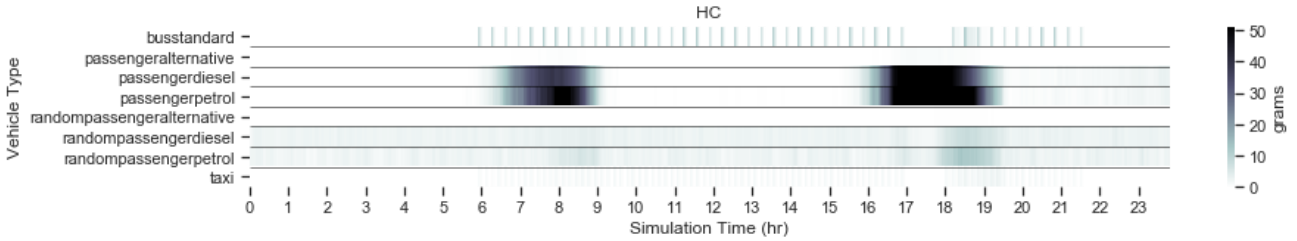

Figure 27 HC Emission Heatmap

Diesel vehicles are responsible for the largest emissions of nitrogen oxides as illustrated in figure 28. Nitrogen oxides react with ozone to produce photochemical smog.

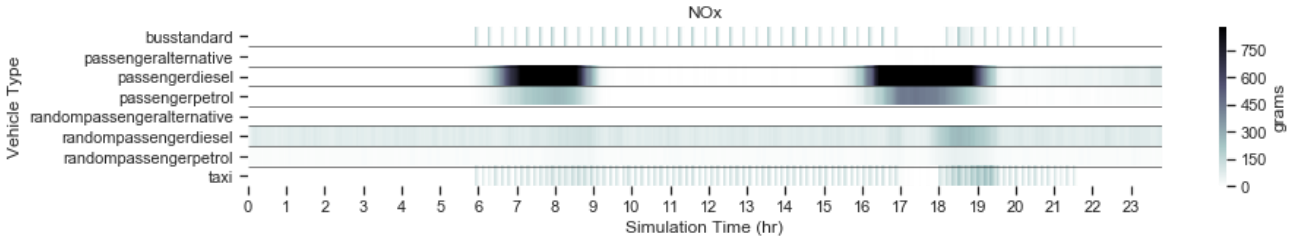

Figure 28 NOx Emission Heatmap

Diesel vehicles are clearly responsible for the majority of particulate matter emissions as illustrated in figure 29. They are known to cause wheezing, coughing and shortness of breath and typically include heavy metals that are toxic.

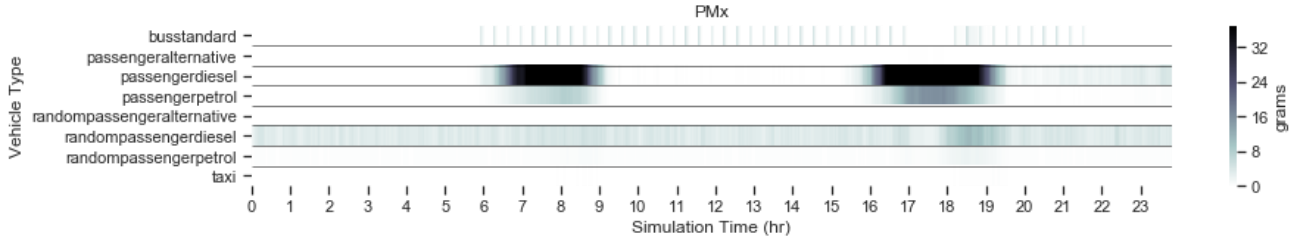

Figure 29 PMx Emission Heatmap

As expected, more fuel is burnt when vehicles are in a congested state as illustrated in figure 30. Reducing congestion will reduce fuel consumption which in turn will reduce all the other emissions tracked in our simulation.

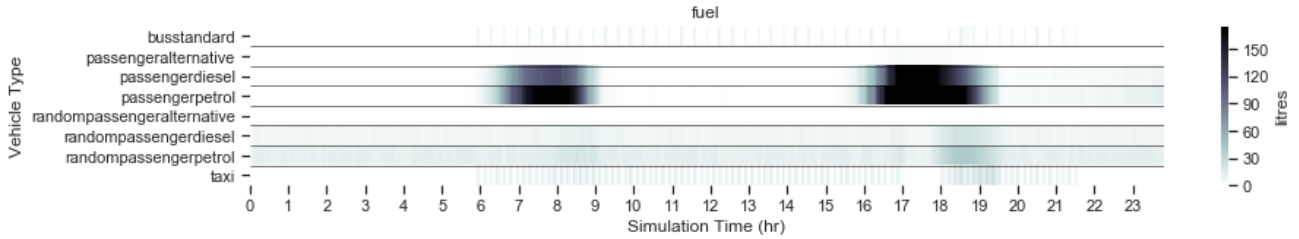

Figure 30 Fuel Consumption Heatmap

6.5.2 Emissions Heatmap

Figure 31 was created by running a kernel density estimator over all of the per vehicle per second emission points weighted by carbon dioxide emissions in order to visually represent which parts of the road network through simulation time contributed the most to the above-mentioned emission outputs.

This allows us to visually interpret the effect the underlying network configuration may have on resultant emission outputs.

The heatmap is rendered using a 1% low/99% high clip to slightly emphasize the regions of interest.

Major routes, major interchanges as well as primary and secondary roads linking the central business district to the surrounding areas resulted in the majority of emission outputs.

Patterns of heightened emissions seem to form at regions typical of stop streets, lane mergers and bends in the road.

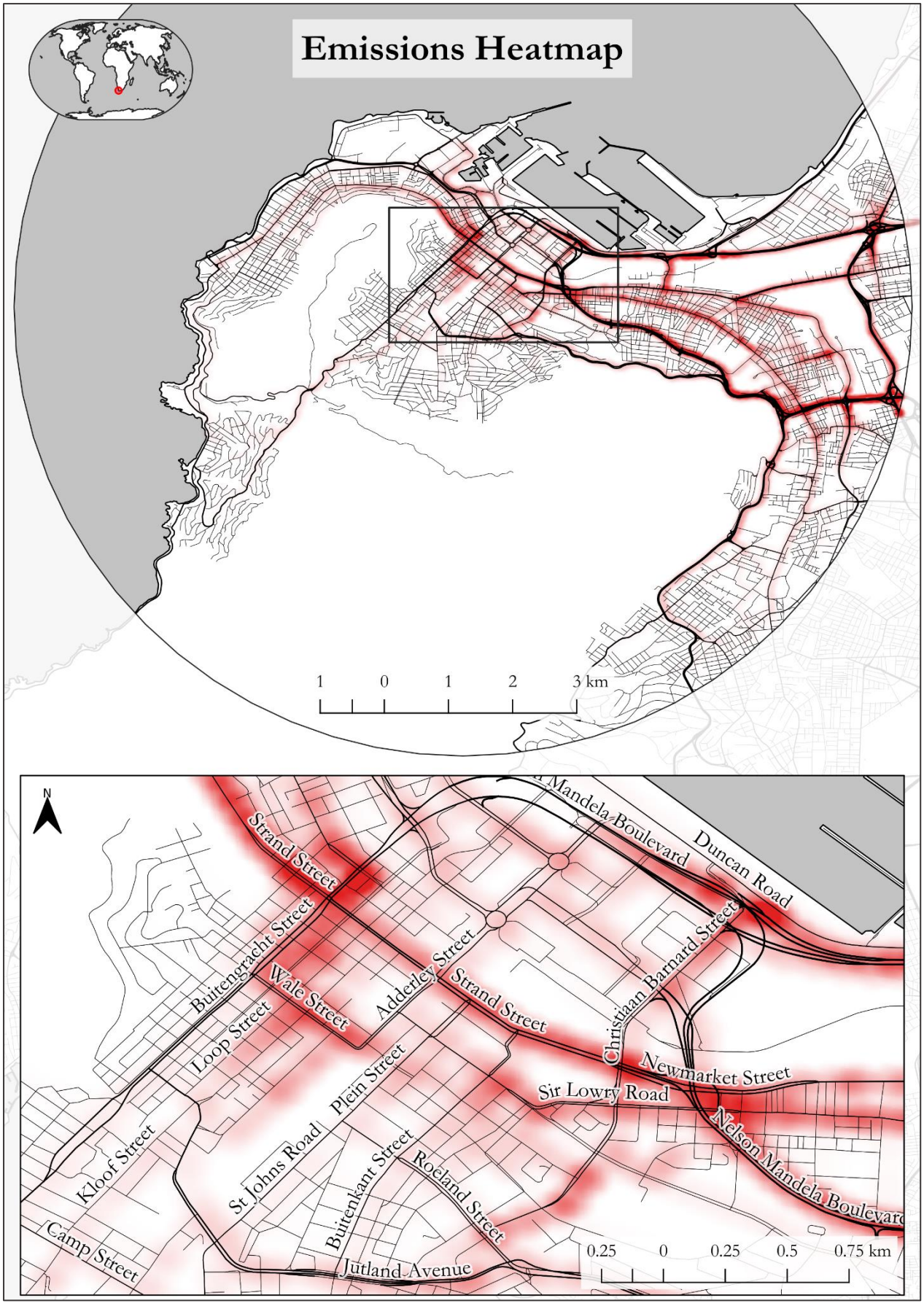

Figure 31 Geographic representation of the emissions heatmap for the full simulation time

7 Discussion

Borshchev and Filippov (2004) illustrate that agent-based modelling is a valid methodology to use when we need to move past the limitations of system dynamics and discrete event approaches. Specifically, in the case of active objects such as people and vehicles where timing, event ordering and individual behaviour is key.

In this paper we have shown that it is possible to develop an agent-based simulation of urban mobility using synthetic trip data.

In terms of the overarching hypothesis, we have shown that using synthetic trip data in such a simulation is a feasible way to create granular emission outputs based on how agents act on one another within the simulation at a high level of detail through simulation time.

Considerable effort was spent going through the road network and capturing the traffic signal locations across the city using the OpenStreetMap iD editor – these edits are now a part of OSM and publicly available for other to use across many different applications that can leverage this publicly available dataset.

Bankes (2002) raises that considerable emphasis has been placed on representational design in agent-based modelling. However, in order for this type of modelling to realise its potential, Bankes argues that focus needs to be redirected to loading of use cases to be explored in the simulation, uncertainty analysis, calibration of models to data and methodologies for using models to answer specific questions to solve problems.

Using a series of statistical tests, we have shown that there are a number of checks that need to be performed to understand what is going on in the model and how it relates to historic aggregates.

Because of the microscopic nature of the simulation, considerable processing time and effort is required to ensure that things run as expected in the simulation. Small errors in the underlying network topology, scheduling of routes, scheduling of traffic signal and vehicle definitions have compounding effects that impact measured outputs.

Uppoor et al. (2014) found that simplistic assumptions about the macroscopic or microscopic dynamics of road traffic can greatly affect the network topology – this has the added risk of introducing bias in the performance evaluation of these simulated environments. The authors noted that future work should focus on finding rigorous means to validate the generated trips with realistic signal propagation information as well as including comprehensive network connectivity analysis.

Observationally, the simulation in this paper is modelling a full day with heavy congestion in the morning and afternoon peaks that conforms to expected typical traffic patterns that can be expected of such a simulation. Comparatively, the simulation travel speeds do not align to the historical aggregate travel speeds, so we cannot say that we have constructed a simulation of an average Monday over the past two years (which the historic dataset represents).

Due to the fact that the simulation conforms to the observational expectations of such a simulation as shown in section 7.2, although we cannot say that we have modelled an historical average through the comparative tests in section 7.3, we would be confident in saying that it is a valid simulation of a *possible* day within the study area within the limitations explicitly modelled in this study.

Model calibration and effective validation are indispensable steps to ensure the reliability of the developed model and the output of subsequent scenario-based experiments. The accuracy of results is dependent on the quality of both the calibration and validation process (Milam, 2000).

Due to the granular nature of the patterns of behaviour being modelling there is a lot of room for improvement to be able to generate a generic enough simulation that can approximate a given week day for an area of interest utilising synthetic trip data.

Uppoor et al. (2014) used an origin-destination matrix approach focussing on the generation and analysis of a synthetic large-scale urban vehicular mobility dataset. They concluded that more real-world data is required to allow for a more rigourous validation process. It was inconclusive whether or not the dataset they had generated for Köln would be sufficiently accurate for networking studies.

There are some unknowns and additions that would have to be a part of the simulation in future iterations to make the outputs more credible, these include:

1. Having a better way of knowing the true number of vehicles that should be running through simulation time – the hard cap we set of 10 000 maximum running vehicles at any one time was determined through iterative runs and observing the behaviour of the model across the observational and comparative metrics discussed. This is in line with what Uppoor et al. (2014) found with regards to requiring real world data to properly calibrate the model.
2. Having the real traffic signal timings and configurations used in the study area – the configuration used in this model is a best guess.
3. In the model we have not explicitly simulated dedicated bus lanes or car-pooling lanes which are common in these types of urban environments and have a direct impact on travel time.

4. Although 20% of the trips generated were completely random, the patterns of behaviour this set of trips represents should be explicitly modelled – these include things such as business deliveries, trucks delivering/collecting cargo from the port or tourist-related traffic.
5. The simulation should be able to run without the teleportation function – this would mean having a 100% topologically accurate representation of the network (and its control systems) as well as having the correct scaling of demand on each road segment through simulation time in order to prevent the excessive jamming scenarios that this functionality addresses. Similarly, the necessary setting of only allowing vehicles to enter the simulation if they are able to at the modelled step in time would need to be addressed by the same work. This would result in a better representation of the fundamental diagram of traffic as discussed in section 7.4.
6. In this simulation we are referencing the HBEFA v3.1 vehicular emissions database – similar databases should be compared and contrasted to validate the emission outputs. Over and above this, in situ measurements could be taken for a study area and compared to simulated outputs within an agent-based simulation similar to what has been created in this study. Elkafoury et al (2015) focussed on comparing Carbon Monoxide emissions measured in situ against HBEFA V3.1 for New Borg El-Arab in Egypt and found that the HBEFA model underestimated emission figures. Elkafoury et al. (2015) went on to develop their own regression model to try and predict CO emissions and found it to be more accurate than the HBEFA estimations which were not aligned to the vehicle classes and behaviours of the study area.
7. Subsequent work should focus specifically on treating the generation of the trips and the dynamic user equilibrium configuration as an optimisation problem focussed on minimising the error between a simulated measure (such as travel speeds) and historic data. In doing so, it means we will be able to use high-level population statistics to create accurate models of reality without the need for considerable in situ measurements – this would accelerate the extent and scale at which this type of study can be carried out in the absence of more granular data. Paz et al. (2015) proposed a Memetic Algorithm which combines a genetic algorithm and simulated annealing that would allow for accurate calibration of such models; however this type of calibration is directly dependant on having access to vehicle count data at different parts of the network through the time period you intend on simulating.

Once we are at a point where we can statistically prove that the model represents historic aggregates then this type of model becomes a very powerful scenario planning tool as made evident by Ragab, M., Hashim, I. and Asar, G. (2017) who explored some simulation scenarios for Egypt that showed an increase in road speed can reduce emissions and fuel consumption and reducing traffic volumes can reduce emissions and fuel consumption (with minimal impact on travel time).

8 Conclusion

This paper investigated how a computer simulation of individual vehicles interacting with one another on a microscopic scale within the City of Cape Town, South Africa using derived trip data from aggregate statistics in an automated way can be used to emulate traffic conditions within the city.

Using this simulated environment, emissions that these vehicles pump into the environment namely CO_2, CO, HC, NOx, PMx (and the requisite fuel consumption) were calculated after careful configuration and testing of the simulation environment.

Modelling this type of problem in this way can enable researchers to understand not only the dynamic interaction of agents acting in a system, but also which specific characteristics about the relationships between these agents have the largest impact on reducing total emissions. Modelling at such a fine scale means the researcher has many 'levers' they can pull once they have a properly calibrated model.

Although the simulation was not able to represent an average Monday for the study area, the simulation is still valid in many of the observational characteristics for a given passenger road network. In conducting this exercise, we have also explored the complexities involved in modelling users of a road networks at a granular level using aggregate data as a source.

This paper has also shown the importance of running tests on the simulation outputs – one may be able to very easily create a traffic scenario using these techniques, however more advanced modelling needs to be employed to ensure that the synthetic data produces a simulation that is able to approximate true conditions for the study area.

There are many aspects of urban mobility that can be added and extended upon this research, this is only scratching the surface of what is possible with such a detailed model. Global policy is shifting to accommodate 'greener' thinking. Exploring emissions on a microscopic level thanks to technological advances can aid in how these policies evolve over time and can provide insight into how we impact our environment through daily commutes.

Future studies can build on the framework laid out in this paper and it is recommended that for this type of synthetic trip generation to be valid for modelling a realistic 'real world' day considerable attention should be paid towards developing a methodology that allows a user of the simulation to calibrate the simulation to a known state as closely as possible in an automated manner.

Agent-based simulations have developed over time into a useful tool that allows us to model complex activities. The research conducted in this paper forms the basis for what is required to allow researchers to explore system-wide implications of agent-modelled actions, behaviours and decisions. Particularly in the context of transportation network optimisation and planning, these kinds of models provide us the

necessary tools to simulate both changes to the network and events on the network to inform decision making which could ultimately result in the development of smarter urban infrastructure based on empirical evidence.

9 References

Bankes, S. (2002). Agent-based modeling: A revolution?. Proceedings of the National Academy of Sciences, 99(Supplement 3), pp.7199-7200.

Bazzan, A. and Klügl, F. (2013). A review on agent-based technology for traffic and transportation. The Knowledge Engineering Review, 29(03), pp.375-403.

Borshchev, A. and Filippov, A. (2004). From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools. The 22nd International Conference of the System Dynamics Society, July 25 - 29, 2004, Oxford, England.

Dias, J., Henriques Abreu, P., Castro Silva, D., Fernades, G., Machado, P. and Leitão, A. (2013). Preparing Data for Urban Traffic Simulation using SUMO. Proceeding: SUMO 2013 User Conference.

Dias, J., Machado, P., Silva, D. and Abreu, P. (2014). An Inverted Ant Colony Optimization approach to traffic. Engineering Applications of Artificial Intelligence, 36, pp.122-133.

Dlr.de. (2018). DLR - Institute of Transportation Systems - SUMO – Simulation of Urban MObility. [online] Available at: https://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/ [Accessed 8 Oct. 2018].

Elkafoury, A., Negm, A., Aly, M., Bady, M. and Ichimura, T. (2015). Develop dynamic model for predicting traffic CO emissions in urban areas. Environmental Science and Pollution Research, 23(16), pp.15899-15910.

Environmental Protection Agency. 2005. Air Emission Sources. Available from internet: http://www.epa.gov/air/emissions/ [Accessed 8 Oct. 2018]

Fakir, S. (2018). SA's Carbon Tax Bill nearly there, but not yet enough to do the job | Daily Maverick. [online] Daily Maverick. Available at: https://www.dailymaverick.co.za/opinionista/2018-03-15-sas-carbon-tax-bill-nearly-there-but-not-yet-enough-to-do-the-job/ [Accessed 19 Oct. 2018].

Flitsch, C., Kastner, K., Bosa, K. and Neubauer, M. (2018). Calibrating Traffic Simulation Models in SUMO Based upon Diverse Historical Real–Time Traffic Data — Lessons Learned in ITS Upper Austria. EPiC Series in Engineering, 2, pp.25-42.

Flötteröd, G., Chen, Y. and Nagel, K. (2011). Behavioral Calibration and Analysis of a Large-Scale Travel Microsimulation. Networks and Spatial Economics, 12(4), pp.481-502.

Gawron, C. (1998). An Iterative Algorithm to Determine the Dynamic User Equilibrium in a Traffic Simulation Model. International Journal of Modern Physics, 09(03), pp.393-407.

GHG National Inventory Report South Africa 2000 - 2010. (2014). [ebook] Department of Environmental Affairs, p.46. Available at: https://www.environment.gov.za/sites/default/files/docs/greenhousegas_invetorysouthafrica.pdf [Accessed 13 Oct. 2018].

Hbefa.net. (2018). HBEFA - Handbook Emission Factors for Road Transport. [online] Available at: http://www.hbefa.net/e/index.html [Accessed 10 Oct. 2018].

Heppenstall, A. (2012). Agent-based models of geographical systems. Publisher: Springer.

Immers, L. and Logghe, S. (2002). [online] Traffic Flow Theory. Available at: https://www.mech.kuleuven.be/cib/verkeer/dwn/H111part3.pdf [Accessed 16 Dec. 2018].

Its.mit.edu. (2018). MITSIMLab | INTELLIGENT TRANSPORTATION SYSTEMS LAB. [online] Available at: https://its.mit.edu/software/mitsimlab [Accessed 8 Oct. 2018].

Jeihani, M.; James, P.; Saka, A.; Ardeshiri, A. 2015. Traffic Recovery Time Estimation under Different Flow Regimes in Traffic Simulation, Journal of Traffic and Transportation Engineering (English Edition) 2(5): 291-300.

Krajzewicz, D., Erdmann, J., Behrisch, M. and Bieker, L. (2012). Recent Development and Applications of SUMO – Simulation of Urban MObility. International Journal on Advances in Systems and Measurements,, 5(3 & 4).

Krauß, S. (1998). Microscopic Modeling of Traffic Flow: Investigation of Collision Free Vehicle Dynamics. Hauptabteilung Mobilitat und Systemtechnik - Koln.

Kun, C.H.E.N.; Lei, Y.U. 2007. Microscopic TrafficEmission. Simulation and Case Study for Evaluation of Traffic Control Strategies, Journal of Transportation. Systems Engineering and Information Technology 7(1): 93-99.

Matsim.atlassian.net. (2018). MATSim vs .... [online] Available at: https://matsim.atlassian.net/wiki/spaces/MATPUB/pages/84705338/MATSim+vs.+... [Accessed 12 Oct. 2018].

MATSim.org. (2018). MATSim.org. [online] Available at: https://matsim.org/ [Accessed 8 Oct. 2018].

Milam, R., 2000. Recommended Guidelines for the Calibration and Validation of Traffic Simulation Models. Available at: http:/www.fehrandpeers.com/ [Accessed 12 Oct. 2018]

Paz, A., Molano, V., Martinez, E., Gaviria, C. and Arteaga, C. (2015). Calibration of traffic flow models using a memetic algorithm. Transportation Research Part C: Emerging Technologies, 55, pp.432-443.

Pmg.org.za. (2018). Carbon Tax Draft Bill: National Treasury Briefing | PMG. [online] Available at: https://pmg.org.za/page/Carbon%20Tax%20Draft%20Bill:%20National%20Treasury%20briefing?via=homepage-feature-card [Accessed 14 Oct. 2018].

Ragab, M., Hashim, I. and Asar, G. (2017). Impact of Road Traffic on Air Emissions: Case Study Kafr El-Sheikh City, Egypt. International Journal for Traffic and Transport Engineering, 7(3).

Reporter, S. (2018). What Cape Town's R481m traffic congestion project will bring | IOL Business Report. [online] Iol.co.za. Available at: https://www.iol.co.za/business-report/economy/what-cape-towns-r481m-traffic-congestion-project-will-bring-15456391 [Accessed 16 Sep. 2018].

Saidallah, M., El Fergougui, A. and Elalaoui, A. (2016). A Comparative Study of Urban Road Traffic Simulators. MATEC Web of Conferences, 81, p.05002.

SourceForge. (2018). Transportation Analysis and Simulation. [online] Available at: https://sourceforge.net/projects/transims/ [Accessed 8 Oct. 2018].

Stevanovic, A., Stevanovic, J., Zhang, K. and Batterman, S. (2009). Optimizing Traffic Control to Reduce Fuel Consumption and Vehicular Emissions. Transportation Research Record: Journal of the Transportation Research Board, 2128, pp.105-113.

Sumo.dlr.de. (2018). Simulation of Urban Mobility - Wiki. [online] Available at: http://sumo.dlr.de/wiki/ [Accessed 6 Oct. 2018].

Swarts, S. (2017). Modelling South African traffic for large networks - An extension of the gravity model for traffic demand modelling. University of Stellenbosch, South Africa.

TomTom. (2016). TomTom Traffic Index. Available at: https://www.tomtom.com/en_gb/trafficindex/ [Accessed 7 Oct. 2018]

Transactioncapital.co.za. (2017). Mini Bus Taxi Industry in South Africa. [online] Available at: https://www.transactioncapital.co.za/downloads/taxi/TC%20website%20SA%20Taxi%20section.pdf [Accessed 7 Oct. 2018].

Uppoor, S., Trullols-Cruces, O., Fiore, M. and Barcelo-Ordinas, J. (2014). Generation and Analysis of a Large-Scale Urban Vehicular Mobility Dataset. IEEE Transactions on Mobile Computing, 13(5), pp.1061-1075.

Series from Lund University

Department of Physical Geography and Ecosystem Science

Master Thesis in Geographical Information Science

1. *Anthony Lawther:* The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland (2008).
2. *Rickard Hansen:* Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research (2008).
3. *Emil Bayramov:* Environmental monitoring of bio-restoration activities using GIS and Remote Sensing (2009).
4. *Rafael Villarreal Pacheco:* Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia (2009).
5. *Siri Oestreich Waage:* a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).
6. *Edgar Pimiento:* Shallow landslide susceptibility – Modelling and validation (2010).
7. *Martina Schäfer:* Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010).
8. *August Pieter van Waarden-Nagel:* Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010).
9. *Samira Muhammad:* Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010).
10. *Fredros Oketch Okumu*: Using remotely sensed data to explore spatial and temporal relationships between photosynthetic productivity of vegetation and malaria transmission intensities in selected parts of Africa (2011).
11. *Svajunas Plunge:* Advanced decision support methods for solving diffuse water pollution problems (2011).
12. *Jonathan Higgins:* Monitoring urban growth in greater Lagos: A case study using GIS to monitor the urban growth of Lagos 1990 - 2008 and produce future growth prospects for the city (2011).
13. *Mårten Karlberg:* Mobile Map Client API: Design and Implementation for Android (2011).
14. *Jeanette McBride:* Mapping Chicago area urban tree canopy using color infrared imagery (2011).
15. *Andrew Farina:* Exploring the relationship between land surface temperature and vegetation abundance for urban heat island mitigation in Seville, Spain (2011).

16. *David Kanyari*: Nairobi City Journey Planner: An online and a Mobile Application (2011).

17. *Laura V. Drews:* Multi-criteria GIS analysis for siting of small wind power plants - A case study from Berlin (2012).

18. *Qaisar Nadeem:* Best living neighborhood in the city - A GIS based multi criteria evaluation of ArRiyadh City (2012).

19. *Ahmed Mohamed El Saeid Mustafa:* Development of a photo voltaic building rooftop integration analysis tool for GIS for Dokki District, Cairo, Egypt (2012).

20. *Daniel Patrick Taylor*: Eastern Oyster Aquaculture: Estuarine Remediation via Site Suitability and Spatially Explicit Carrying Capacity Modeling in Virginia's Chesapeake Bay (2013).

21. *Angeleta Oveta Wilson:* A Participatory GIS approach to *unearthing* Manchester's Cultural Heritage '*gold mine'* (2013).

22. *Ola Svensson:* Visibility and Tholos Tombs in the Messenian Landscape: A Comparative Case Study of the Pylian Hinterlands and the Soulima Valley (2013).

23. *Monika Ogden:* Land use impact on water quality in two river systems in South Africa (2013).

24. *Stefan Rova:* A GIS based approach assessing phosphorus load impact on Lake Flaten in Salem, Sweden (2013).

25. *Yann Buhot:* Analysis of the history of landscape changes over a period of 200 years. How can we predict past landscape pattern scenario and the impact on habitat diversity? (2013).

26. *Christina Fotiou:* Evaluating habitat suitability and spectral heterogeneity models to predict weed species presence (2014).

27. *Inese Linuza:* Accuracy Assessment in Glacier Change Analysis (2014).

28. *Agnieszka Griffin:* Domestic energy consumption and social living standards: a GIS analysis within the Greater London Authority area (2014).

29. *Brynja Guðmundsdóttir:* Detection of potential arable land with remote sensing and GIS - A Case Study for Kjósarhreppur (2014).

30. *Oleksandr Nekrasov:* Processing of MODIS Vegetation Indices for analysis of agricultural droughts in the southern Ukraine between the years 2000-2012 (2014).

31. *Sarah Tressel:* Recommendations for a polar Earth science portal

    in the context of Arctic Spatial Data Infrastructure (2014).

32. *Caroline Gevaert:* Combining Hyperspectral UAV and Multispectral Formosat-2 Imagery for Precision Agriculture Applications (2014).

33. *Salem Jamal-Uddeen:* Using GeoTools to implement the multi-criteria evaluation analysis - weighted linear combination model (2014).

34. *Samanah Seyedi-Shandiz:* Schematic representation of geographical railway network at the Swedish Transport Administration (2014).

35. *Kazi Masel Ullah:* Urban Land-use planning using Geographical Information System and analytical hierarchy process: case study Dhaka City (2014).

36. *Alexia Chang-Wailing Spitteler:* Development of a web application based on MCDA and GIS for the decision support of river and floodplain rehabilitation projects (2014).

37. *Alessandro De Martino:* Geographic accessibility analysis and evaluation of potential changes to the public transportation system in the City of Milan (2014).

38. *Alireza Mollasalehi:* GIS Based Modelling for Fuel Reduction Using Controlled Burn in Australia. Case Study: Logan City, QLD (2015).

39. *Negin A. Sanati:* Chronic Kidney Disease Mortality in Costa Rica; Geographical Distribution, Spatial Analysis and Non-traditional Risk Factors (2015).

40. *Karen McIntyre:* Benthic mapping of the Bluefields Bay fish sanctuary, Jamaica (2015).

41. *Kees van Duijvendijk:* Feasibility of a low-cost weather sensor network for agricultural purposes: A preliminary assessment (2015).

42. *Sebastian Andersson Hylander:* Evaluation of cultural ecosystem services using GIS (2015).

43. *Deborah Bowyer:* Measuring Urban Growth, Urban Form and Accessibility as Indicators of Urban Sprawl in Hamilton, New Zealand (2015).

44. *Stefan Arvidsson:* Relationship between tree species composition and phenology extracted from satellite data in Swedish forests (2015).

45. *Damián Giménez Cruz*: GIS-based optimal localisation of beekeeping in rural Kenya (2016).

46. *Alejandra Narváez Vallejo:* Can the introduction of the topographic indices in LPJ-GUESS improve the spatial representation of environmental variables? (2016).

47. *Anna Lundgren:* Development of a method for mapping the highest coastline in Sweden using breaklines extracted from high resolution digital elevation models (2016).

48. *Oluwatomi Esther Adejoro:* Does location also matter? A spatial analysis of social achievements of young South Australians (2016).

49. *Hristo Dobrev Tomov:* Automated temporal NDVI analysis over the Middle East for the period 1982 - 2010 (2016).

50. *Vincent Muller:* Impact of Security Context on Mobile Clinic Activities

    A GIS Multi Criteria Evaluation based on an MSF Humanitarian Mission in Cameroon (2016).

51. *Gezahagn Negash Seboka:* Spatial Assessment of NDVI as an Indicator of Desertification in Ethiopia using Remote Sensing and GIS (2016).

52. *Holly Buhler:* Evaluation of Interfacility Medical Transport Journey Times in Southeastern British Columbia. (2016).

53. *Lars Ole Grottenberg*: Assessing the ability to share spatial data between emergency management organisations in the High North (2016).

54. *Sean Grant:* The Right Tree in the Right Place: Using GIS to Maximize the Net Benefits from Urban Forests (2016).

55. *Irshad Jamal:* Multi-Criteria GIS Analysis for School Site Selection in Gorno-Badakhshan Autonomous Oblast, Tajikistan (2016).

56. *Fulgencio Sanmartín:* Wisdom-volkano: A novel tool based on open GIS and time-series visualization to analyse and share volcanic data (2016).

57. *Nezha Acil:* Remote sensing-based monitoring of snow cover dynamics and its influence on vegetation growth in the Middle Atlas Mountains (2016).

58. *Julia Hjalmarsson:* A Weighty Issue: Estimation of Fire Size with Geographically Weighted Logistic Regression (2016).

59. *Mathewos Tamiru Amato:* Using multi-criteria evaluation and GIS for chronic food and nutrition insecurity indicators analysis in Ethiopia (2016).

60. *Karim Alaa El Din Mohamed Soliman El Attar:* Bicycling Suitability in Downtown, Cairo, Egypt (2016).

61. *Gilbert Akol Echelai:* Asset Management: Integrating GIS as a Decision Support Tool in Meter Management in National Water and Sewerage Corporation (2016).

62. *Terje Slinning:* Analytic comparison of multibeam echo soundings (2016).

63. *Gréta Hlín Sveinsdóttir:* GIS-based MCDA for decision support: A framework for wind farm siting in Iceland (2017).

64. *Jonas Sjögren:* Consequences of a flood in Kristianstad, Sweden: A GIS-based analysis of impacts on important societal functions (2017).

65. *Nadine Raska:* 3D geologic subsurface modelling within the Mackenzie Plain, Northwest Territories, Canada (2017).

66. *Panagiotis Symeonidis*: Study of spatial and temporal variation of atmospheric optical parameters and their relation with PM 2.5 concentration over Europe using GIS technologies (2017).

67. *Michaela Bobeck:* A GIS-based Multi-Criteria Decision Analysis of Wind Farm Site Suitability in New South Wales, Australia, from a Sustainable Development Perspective (2017).

68. *Raghdaa Eissa*: Developing a GIS Model for the Assessment of Outdoor Recreational Facilities in New Cities Case Study: Tenth of Ramadan City, Egypt (2017).

69. *Zahra Khais Shahid*: Biofuel plantations and isoprene emissions in Svea and Götaland (2017).

70. *Mirza Amir Liaquat Baig*: Using geographical information systems in epidemiology: Mapping and analyzing occurrence of diarrhea in urban - residential area of Islamabad, Pakistan (2017).

71. *Joakim Jörwall*: Quantitative model of Present and Future well-being in the EU-28: A spatial Multi-Criteria Evaluation of socioeconomic and climatic comfort factors (2017).

72. *Elin Haettner*: Energy Poverty in the Dublin Region: Modelling Geographies of Risk (2017).

73. *Harry Eriksson*: Geochemistry of stream plants and its statistical relations to soil- and bedrock geology, slope directions and till geochemistry. A GIS-analysis of small catchments in northern Sweden (2017).

74. *Daniel Gardevärn:* PPGIS and Public meetings – An evaluation of public participation methods for urban planning (2017).

75. *Kim Friberg:* Sensitivity Analysis and Calibration of Multi Energy Balance Land Surface Model Parameters (2017).

76. *Viktor Svanerud:* Taking the bus to the park? A study of accessibility to green areas in Gothenburg through different modes of transport (2017).

77. *Lisa-Gaye Greene*: Deadly Designs: The Impact of Road Design on Road Crash Patterns along Jamaica's North Coast Highway (2017).

78. *Katarina Jemec Parker*: Spatial and temporal analysis of fecal indicator bacteria concentrations in beach water in San Diego, California (2017).

79. *Angela Kabiru*: An Exploratory Study of Middle Stone Age and Later Stone Age Site Locations in Kenya's Central Rift Valley Using Landscape Analysis: A GIS Approach (2017).

80. *Kristean Björkmann*: Subjective Well-Being and Environment: A GIS-Based Analysis (2018).

81. *Williams Erhunmonmen Ojo*: Measuring spatial accessibility to healthcare for people living with HIV-AIDS in southern Nigeria (2018).

82. *Daniel Assefa*: Developing Data Extraction and Dynamic Data Visualization (Styling) Modules for Web GIS Risk Assessment System (WGRAS). (2018).

83. *Adela Nistora*: Inundation scenarios in a changing climate: assessing potential impacts of sea-level rise on the coast of South-East England (2018).

84. *Marc Seliger*: Thirsty landscapes - Investigating growing irrigation water consumption and potential conservation measures within Utah's largest master-planned community: Daybreak (2018).

85. *Luka Jovičić*: Spatial Data Harmonisation in Regional Context in Accordance with INSPIRE Implementing Rules (2018).

86. *Christina Kourdounouli*: Analysis of Urban Ecosystem Condition Indicators for the Large Urban Zones and City Cores in EU (2018).

87. *Jeremy Azzopardi*: Effect of distance measures and feature representations on distance-based accessibility measures (2018).

88. *Patrick Kabatha*: An open source web GIS tool for analysis and visualization of elephant GPS telemetry data, alongside environmental and anthropogenic variables (2018).

89. *Richard Alphonce Giliba*: Effects of Climate Change on Potential Geographical Distribution of Prunus africana (African cherry) in the Eastern Arc Mountain Forests of Tanzania (2018).

90. *Eiður Kristinn Eiðsson*: Transformation and linking of authoritative multi-scale geodata for the Semantic Web: A case study of Swedish national building data sets (2018).

91. *Niamh Harty*: HOP!: a PGIS and citizen science approach to monitoring the condition of upland paths (2018).

92. *José Estuardo Jara Alvear*: Solar photovoltaic potential to complement hydropower in Ecuador: A GIS-based framework of analysis (2018).

93. *Brendan O'Neill*: Multicriteria Site Suitability for Algal Biofuel Production Facilities (2018).

94. *Roman Spataru*: Spatial-temporal GIS analysis in public health – a case study of polio disease (2018).

95. *Alicja Miodońska*: Assessing evolution of ice caps in Suðurland, Iceland, in years 1986 - 2014, using multispectral satellite imagery (2019).

96. *Dennis Lindell Schettini*: A Spatial Analysis of Homicide Crime's Distribution and Association with Deprivation in Stockholm Between 2010-2017 (2019).

97. *Damiano Vesentini*: The Po Delta Biosphere Reserve: Management challenges and priorities deriving from anthropogenic pressure and sea level rise (2019).

98. *Emilie Arnesten*: Impacts of future sea level rise and high water on roads, railways and environmental objects: a GIS analysis of the potential effects of increasing sea levels and highest projected high water in Scania, Sweden (2019).

99. *Syed Muhammad Amir Raza*: Comparison of geospatial support in RDF stores: Evaluation for ICOS Carbon Portal metadata (2019).

100. *Hemin Tofiq*: Investigating the accuracy of Digital Elevation Models from UAV images in areas with low contrast: A sandy beach as a case study (2019).

101. *Evangelos Vafeiadis*: Exploring the distribution of accessibility by public transport using spatial analysis. A case study for retail concentrations and public hospitals in Athens (2019).

102. *Milan Sekulic*: Multi-Criteria GIS modelling for optimal alignment of roadway by-passes in the Tlokweng Planning Area, Botswana (2019).

103. *Ingrid Piirisaar*: A multi-criteria GIS analysis for siting of utility-scale photovoltaic solar plants in county Kilkenny, Ireland (2019).

104. *Nigel Fox*: Plant phenology and climate change: possible effect on the onset of various wild plant species' first flowering day in the UK (2019).

105. *Gunnar Hesch*: Linking conflict events and cropland development in Afghanistan, 2001 to 2011, using MODIS land cover data and Uppsala Conflict Data Programme (2019).

106. *Elijah Njoku*: Analysis of spatial-temporal pattern of Land Surface Temperature (LST) due to NDVI and elevation in Ilorin, Nigeria (2019).

107. *Katalin Bunyevácz*: Development of a GIS methodology to evaluate informal urban green areas for inclusion in a community governance program (2019).

108. *Paul dos Santos*: Automating synthetic trip data generation for an agent-based simulation of urban mobility (2019).