CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 OVERALL ARCHITECTURE

1.1 SIMULATION COMPONENTS

A transit simulation is a useful tool for designing and optimizing public transportation systems. It allows users to replicate a typical day in a city and experiment with different configurations of the transit system to identify and address challenges and optimize the system for the needs of the population.

Due to digital disruption, transit networks have seen a significant transformation in recent years. Consequently, traditional public transport systems no longer meet the diverse needs of passengers, and so demand-responsive transit alternatives have been implemented. However, there is a shortage of simulation tools designed to test and evaluate complicated real-time demand-responsive public transit situations. Generally, a transit simulation includes three main components: transit system, background traffic, and transportation infrastructures, illustrated in Figure 1.1. The transit system includes buses, bus stops, and commuters. Meanwhile, the background traffic consists of other mobility modes such as private vehicles, taxicabs, freight vehicles, and pedestrians. In addition, the transit system and the background traffic shared use the transportation infrastructure, so they affect each other.

1.2 SIMULATION ARCHITECTURE

There are three main components of a transit simulation: the transit system, the background traffic, and the transportation infrastructure. The transit system includes the buses, bus stops, and commuters that make up the public transportation system. The background traffic consists of other modes of transportation such as private vehicles, taxis, freight vehicles, and pedestrians. Both the transit system and the background traffic share the use of the transportation infrastructure, so they can affect each other.

The transit system is made up of individual components that can be controlled and configured to meet the needs of the population. This includes the number of buses, the routes they run on, their schedule, and other factors. By simulating the transit system, users can visualize how these different components interact and how they affect the overall efficiency and effectiveness of the system.

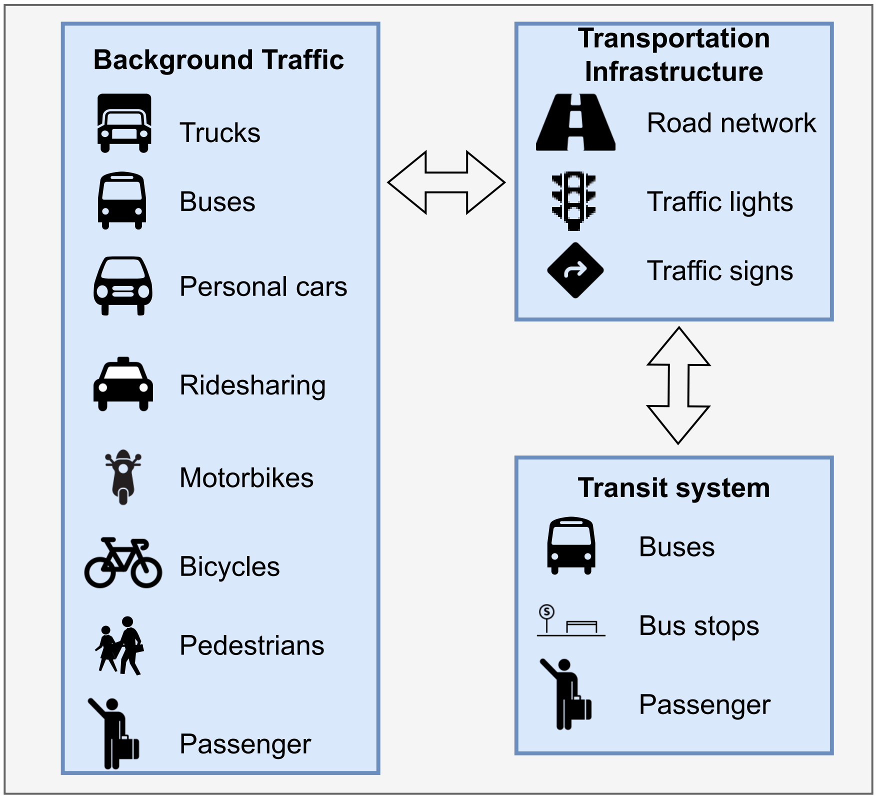


Figure 1.1. Components of a transit simulation

The background traffic consists of all other modes of transportation that are not part of the public transit system. This includes private vehicles, taxis, freight vehicles, and pedestrians. These modes of transportation can affect the transit system in various ways, such as by causing delays or congestion on the roads. By including the background traffic in the simulation, users can get a more accurate picture of how the transit system functions in the context of the overall transportation network.

The transportation infrastructure is the physical infrastructure that supports all modes of transportation, including the roads, bridges, and other infrastructure that the transit system and the background traffic use. The condition and capacity of the transportation infrastructure can have a significant impact on the efficiency and effectiveness of the transit system. By simulating the transportation infrastructure, users can identify potential bottlenecks or other issues that could affect the transit system and work to address them.

Overall, a transit simulation is a powerful tool for designing and optimizing public transportation systems. By allowing users to visualize and experiment with different configurations of the transit system, it can help to identify and address challenges and optimize the system for the needs of the population. By combining the transit system, the background traffic, and the transportation infrastructure in a single simulation, users can get a comprehensive understanding of how all of these components interact and affect each other, and use this information to design and implement more efficient and effective public transportation systems.

2 ROAD INFRASTRUCTURE

2.1 NETWORK

The urban model boundary includes Hamilton County in Tennessee, Catoosa County in Georgia, and two partial counties (Dade and Walker) in Georgia. The Chattanooga highway network included all interstates, other freeways, arterials, collectors, and a significant portion of the local roads. Here, we use Open Street Maps (OSM) to obtain the map data of the selected area. Using the Java OpenStreet Map editor to handle the map data collected from OSM might further improve the simulation's quality (JOSM).

2.1.1 TAZ Structure

Traffic analysis zone (TAZ) is a geographic area that is used to divide the planning region into small, relatively homogeneous areas in terms of land use and activity. TAZs are used to represent travel within a model study area because it is not practical or feasible to model individual households and employment. Housing and employment data are aggregated to the TAZ data, and the TAZs are used through the model process to calculate the origin and destination of trips in the model.

2.1.2 TRANSIT NETWORK DEVELOPMENT

After constructing the road network and adding cars and passengers to the map, the routes for individual vehicles are assigned. Each route begins at the origin of the vehicles and concludes at the destination shared by all vehicles. Along the trip, buses will stop to pick up people. Chattanooga Area Regional Transportation Authority (CARTA) is the main public transit service provider of the region. CARTA operates 13 fixed-route bus routes, two dial-a-ride neighborhood routes, a free shuttle route around the University of Tennessee Chattanooga campus, and two free downtown electric shuttles routes. All fixed-route bus routes (including three shuttle service routes) were modeled in the transit network. We generate an XML file including each vehicle's route information.

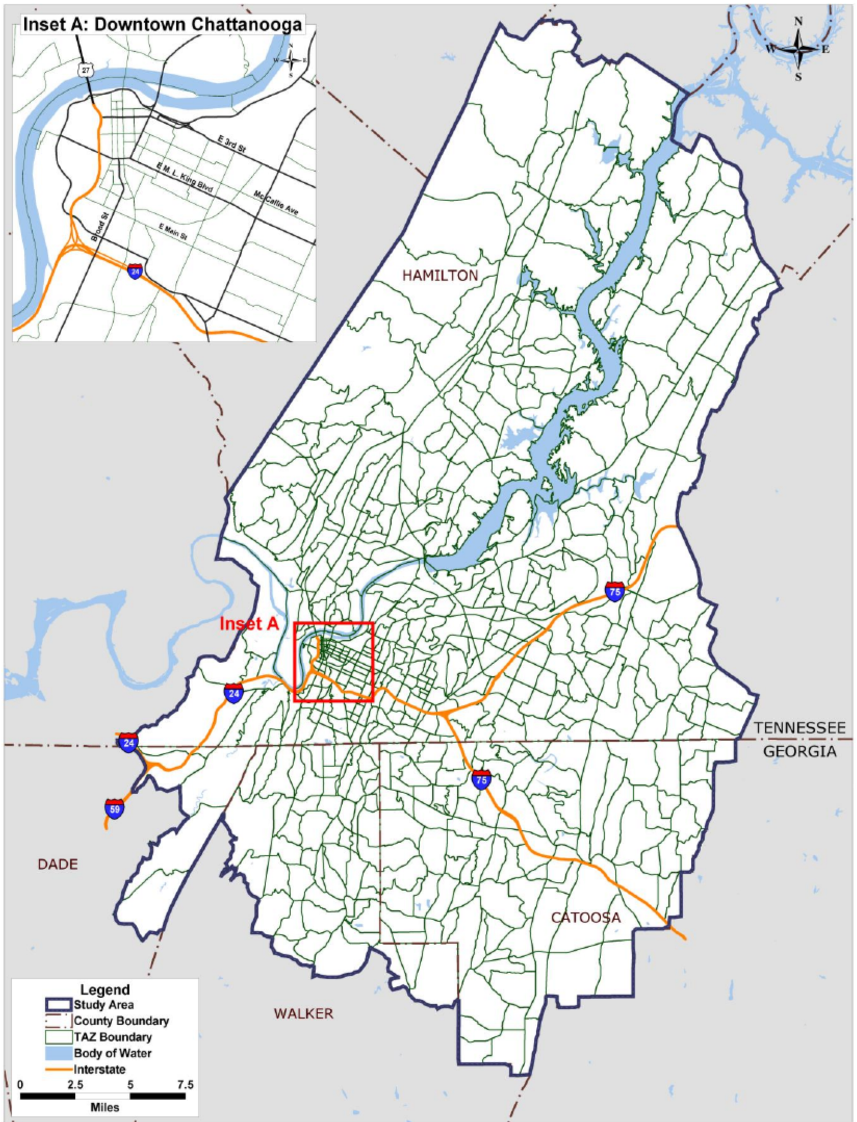


Figure 2.1. Traffic Analysis Zones

2.2 PROBLEM

2.2.1 Time of Day Model

Four time-of-day periods are incorporated into the model stream from the destination choice to assignment steps. The four time-of-day periods are AM peak, Midday, PM peak, and off-peak. Development of the time-of-day model included identifying peak travel time periods, developing peak period factors, and developing percentage of trips by purpose during each time period by direction. These factors will be used to reflect peak period traffic behavior.

2.2.2 Network Corrections

In order to simulate travel within the Chattanooga study area, a computer network must be developed that represents the street system to be modeled. Using JOSM, we have increased the connectedness of the road network. Using the JOSM editor, a user may also define the speed limits of the roads, the number of lanes, one-way roads, and traffic signal systems. Using NETCONVERT, we transform the map into an SUMOLopez et al. (2018) road network file. During this step, the map is turned into a directed graph so that SUMO can simulate traffic on it. In addition, NETCONVERT includes choices to discard particular types of highways, eliminate unconnected roads, etc. Using SUMO, we produce a map appropriate for traffic simulation. As a part of the network development process, corrections and quality checks were made to the SUMO network. Corrections made to the Chattanooga network include the following:

1. *Transit routes*
2. *Traffic lights*
3. *Modified Disconnected Intersection Nodes*
4. *Repaired Fragmented Roadway Links*
5. *Speed limit*

A poor layout of traffic signals can lead to congestion more quickly than a good configuration, making it a crucial issue in the modeling of vehicular flows. In this study, we validate manually using the ”traffic-light mode” of netedit. A SUMO traffic signal consists of eight states (light colors) and several phases. A phase is the durational combination of states. Furthermore, a set of adjustment factors will be developed to estimate the highway and street free flow speed and congested speed based on the posted speed on network links. The adjustment factors will be developed using observed local speed data.

3 BACKGROUND TRAFFIC

Urban area travel demand models are important tools for analysis of transportation plans, projects, and policies. Travel demand modeling practices vary significantly across Metropolitan Planning Organization (MPOs) in the US. In general, traditional 4-step models require less time, data, and resources to develop and validate. More advanced travel demand estimating techniques, such as in activity-based approach, usually require more resources to estimate, validate, and use, but have the benefits of improved sensitivity to policy changes. This section briefly discusses the national trends of the modeling practices.

3.1 PROBLEM

Due to the cost effectiveness, risk free, and high speed benefits , microscopic traffic simulation has been widely employed in transportation planning, design, and analysis. In recent years, the microscopic method has also been given increased prominence in traffic operations and safety investigations. Many microscopic simulation models (such as VISSIM, CORSIM, and SUMO) have been widely employed. In these simulation models, there are separate factors that are utilized to define traffic flow features (e.g. driver behavior and traffic control activities) (e.g. driver behavior and traffic control operations). Even while these microscopic simulation models include default values for certain parameters, simulation under default values sometimes provides incorrect results. Users typically have to fine-tune the settings so that traffic circumstances of genuine case studies may be appropriately portrayed. Therefore, the parameters of microscopic simulation models need to be calibrated and validated. Model calibration serves a significant role in minimizing the disparities between the simulation findings and relevant field observations, such as traffic volumes, speed, and journey time. To establish a close match between the observed and simulated traffic data, one has to execute a good calibration of microscopic traffic simulation model parameters. Because there are a significant number of unknown characteristics involved, the calibration procedure can be a time-consuming and difficult effort. As a consequence, such calibration procedure has been constructed as an optimization model in which a big search space exists due to a wide range of each key model parameters.

3.2 METHOD

3.2.1 Demand generation

Demand for traffic is the description of the vehicles that will circulate on the simulated map. In SUMO, the simulation of a vehicle comprises of three components: vehicle object, type of vehicle, the route. In SUMO, there are three ways to design vehicle routes: 1- Manually. The route file comprises all simulation-capable vehicles. The vehicle and route components specify each vehicle and its respective route, respectively. The definition of a vehicle must include an identification (id) and departure time (depart). Using the edges feature, a route describes the path taken by a vehicle as a list of the IDs of all the streets through which the vehicle will go. A route can be defined as part of a vehicle's definition or externally for usage by several vehicles Urquiza-Aguiar et al. (2020). 2- Randomly. Randomization permits the generation of routes for a road network. The randomTrips.py script produces a series of random routes on a map, and its settings allow for the number, class, and generation interval time of vehicles to be adjusted. 3- Traffic demand generators. Five tools can generate vehicle routes using input data such as stops and Origin-Destination matrices (O/D) including DUArouter, JTRrouter, OD2trips, MArouter, DFrouter. In this study the OD matrices received from the city were used to generation the background traffic demand file. OD2trips was used to convert OD matrices to trips. OD2trips transforms an O/D matrix into separate vehicle trips. In a given time interval, an O/D matrix specifies the number of vehicles that circulate from one district or TAZ (traffic assignment zone) to another. The input files consist of the specification of the districts and the O/D matrix or matrices, whilst the output file contains the produced journeys of vehicles from one TAZ to another. Notably, OD2trips does not require inputs from road networks. The result of this tool are defined origin/destination points for automobile journeys. In other words, OD2trips does not generate each vehicle's particular itinerary. The element TAZ can define any number of districts in the districts' file. A district definition must have at least one street with the taz Source tag and one street with the TAZ Sink tag. With the weight property, the probability of utilizing each origin and destination street are assigned. The O/D matrix can be written in three distinct formats: V, O, and Amitran. The first two are just text files. In V format, the matrix is represented in dense form. In contrast, only places with non-zero values must be supplied in the O format. If the O/D matrix is sparse, O format is thus chosen. The Amitran format follows the same XML schema as the majority of SUMO configuration files. O/D pairs with the odPair tag specify the vehicular demand in this form of O/D matrix. In addition, different time intervals can be specified using timeSlice and vehicle types can be specified with actorConfig.

3.2.2 Dynamic Traffic Assignment and Traffic Simulation Integration

With the advance of computational capacity, dynamic traffic assignment (DTA) methods can be used to assign traffic instead. DTA loads vehicles to the network at fine departure time intervals, with a time-dependent shortest path algorithm to determine "real-time" shortest path, allows vehicles to make route choice decisions based on the real-time information. DTA is usually implemented together within a traffic simulation framework. Depending on how vehicular movements are simulated, the simulation models can be categorized as macroscopic, mesoscopic, or microscopic. Macroscopic simulation uses volume-delay function to move the entire traffic flow, with no vehicle interaction simulated. Mesoscopic models do simulate vehicle dynamic states through simplified car-following or traffic flow theories without describing detailed inter-vehicle interaction to save computational time. Microscopic simulation models attempt to fully simulate individual car following behavior, with lane changes or gap acceptance etc. Traffic micro-simulation models have been used in small area networks for some time, especially for operational analysis such as traffic signal timing and traffic impact studies. But until recently the computational and data requirements have made it infeasible for larger regional models.

3.2.3 Calibration

Calibration of a microscopic traffic simulation model is the act of establishing or fine-tuning the model's parameters such that the difference between observed and simulated traffic measurements is as small as possible. In this regard, the general framework for optimization is expressed as follows. They are the minimum distance allowed between two vehicles (minGap), the maximum acceleration/deceleration as the maximum acceleration/deceleration which can be achieved by the driver, road speed limit in each leg, as the driver imperfection (sigma) and as the driver's reaction time (tau). When adjusting minGap, maximum acceleration, maximum deceleration, and the road speed limit in calibration process, those parameters give small influence to the changes in the simulation, i.e. travel time and queue. But, when adjusting parameters such as driver imperfection (sigma) and driver's reaction time (tau) the simulation give significant changes in terms of queue and travel time.

The parameters of a transportation simulation model must undergo a meticulous calibration procedure to ensure that the model's output is as accurate as feasible. Figure 3.1 describes the workflow of the calibration algorithm. A microscopic model describes the movements of specific combinations of vehicle and driver. These behaviors are the outcome of the features of drivers and their vehicles, the interactions between drivers, driver-road interaction and road characteristics, external factors, and traffic rules and control. Using demand generation tools of SUMO DUArouter, OD2trips, the O/D-matrices were imported and split into single vehicle trips. As a first phase, a specified network was created in SUMO, and detectors were installed to gather output such as speed. Next, a sensitivity analysis was conducted to determine the relevant factors that can have a significant impact on particular results. In the third stage, a genetic algorithm (GA) model was created to determine the optimal values for every relevant parameter. The GA is inspired by evolutionary theory. Its population evolves by selection, hybridization, and mutation. Selection is utilized to increase the likelihood that superior solutions will be utilized when populating new populations (solutions). Utilizing crossover and mutation to produce new solutions. The GA begins with a population generated at random and analyzes possible solutions at each generation. It has been demonstrated that the GA can achieve near-global optimums when calibrating parameters in microscopic traffic simulation models. The calibrated model's performance was evaluated using speed as the effectiveness metric. For this purpose, speeds derived from the calibrated SUMO model and the INRIX dataset were compared. The calibrated model was used as background traffic.

Traffic demand is the description of the vehicles that will circulate on the simulated map. The O/D matrices for this study are provided by the Chattanooga Hamilton County regional planning agency. An O/D matrix provides traffic flows (often vehicle flows) from each origin to each destination. With O/D matrices, traffic may take several paths to complete the trip from the origin to the destination. As there are 909 TAZ in our study area, we have 909x909 O/D matrices. Using demand generation tools of the SUMO DUArouter, OD2trips, the O/D matrices were imported and split into single vehicle trips. The simulation needs to figure out how to get from the origin edge to the destination edge for a collection of vehicles with a set of origin-destination relations (trips). In a network with high traffic, the difficulty of selecting optimal routes that account for journey times is known as user assignment. To address this issue, SUMO offers a variety of options. Using duaIterate to compute a user equilibrium, that is, it tries to identify a route for each vehicle such that no vehicle may lower its trip cost (typically the travel time) by taking a different route. It accomplishes this iteratively by: 1- using duarouter to route automobiles in a network with the lowest known edge costs (starting with empty-network travel times). 2- Invoking sumo to mimic ”actual” travel times based on the computed routes. The resulting edge costs are utilized in the net routing stage. These repetitive procedures are known as Dynamic User Assignment (DUA) and are often used in combination with a traffic simulation framework. Between successive calls of duarouter, the .rou.alt.xml format is used to record not only the current best route but also previously computed alternative routes. These routes are collected within a route distribution and used when deciding the actual route to drive in the next simulation step. This isn't always the one with the currently lowest cost but is rather sampled from the distribution of alternative routes by a configurable algorithm described below. Furthermore, a specified network was created in SUMO, and detectors were installed to collect output such as speed. The detector computes the values by calculating the entry and exit timings of the vehicle.

The settings that can be altered during SUMO calibration are mentioned below. Car-following model parameters (leader). These elements include the speed and acceleration of the leading vehicle, the response time of the following car's driver, and the characteristics of the driver. Lane-changing parameters: On multi lane roadways, the lane-changing parameters define the speed modifications associated with lane-changing and lane selection. Associated factors include front and rear gaps on the target lane, as well as speed gain throughout the maneuver (speed gain likelihood, maintain right probability, and average waiting time).

3.3 RESULT/VALIDATION

The result shown in 3.2 and 3.3 indicate the effect of average speed after calibration for both microscopic and mesoscopic model.

3.4 NOVELTY

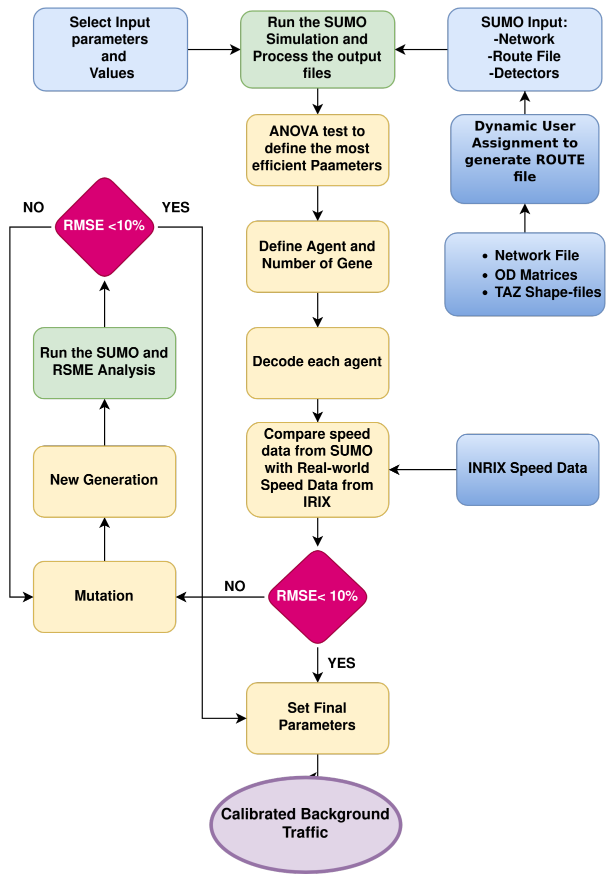


Figure 3.1. Calibration procedure

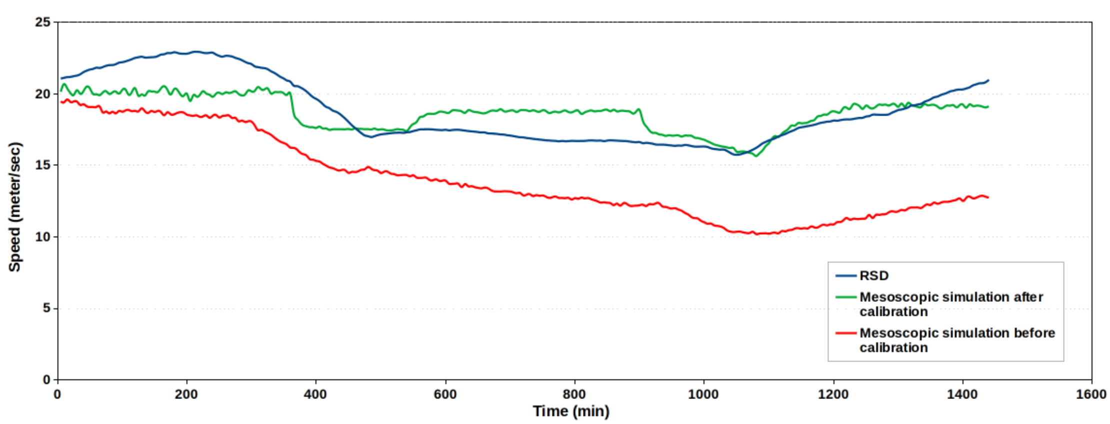


Figure 3.2. Mesoscopic calibration

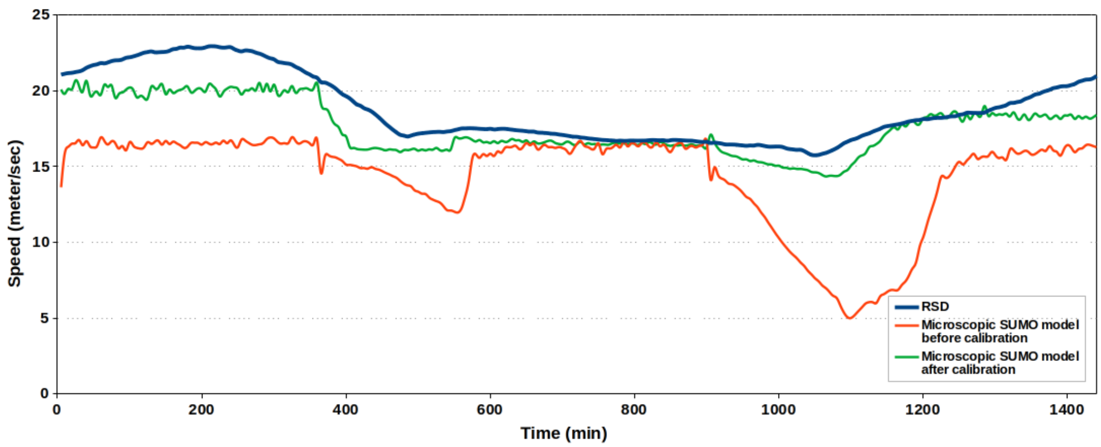


Figure 3.3. Microscopic calibration

4 TRANSIT SYSTEM

4.1 DEMAND GENERATION

4.1.1 Time of day model

Four time-of-day periods are incorporated into the model stream from the destination choice to assignment steps. The four time-of-day periods are AM peak, Midday, PM peak, and off-peak. Development of the time-of-day model included identifying peak travel time periods, developing peak period factors, and developing percentage of trips by purpose during each time period by direction. These factors were used to reflect peak period traffic behavior. The traffic assignment step for the Chattanooga model conducted by each time-of-day period.

4.1.2 Transit network development

Chattanooga Area Regional Transportation Authority (CARTA) is the main public transit service provider of the region. CARTA operates 13 fixed-route bus routes, two dial-a-ride neighborhood routes, a free shuttle route around the University of Tennessee Chattanooga campus, and two free downtown electric shuttles routes. All fixed-route bus routes (including three shuttle service routes) were modeled in the transit network.

4.1.3 Transportation demand

Urban area travel demand models are important tools for analysis of transportation plans, projects, and policies. Travel demand modeling practices vary significantly across Metropolitan Planning Organization (MPOs) in the US. In general, traditional 4-step models require less time, data, and resources to develop and validate. More advanced travel demand estimating techniques, such as in activity-based approach, usually require more resources to estimate, validate, and use, but have the benefits of improved sensitivity to policy changes. The geographic area under consideration is Chattanooga, Tennessee. The city is further divided into census tracts. We find the movement matrix of people travelling for jobs (LODES) from census tract to census tract. The generated origin-destination(OD) pairs are used as requests for our solver. Each row in either dataset represents a single trip by one person. The trip represents movement to the job location and then back to home, at certain times of the day, which are sampled from a given set of regular job start and job end times.

The sources of data can be variegated, and include: (a)Geographic data of the area (from OpenStreetMap), (b)People movement, (c)LODES dataset, (c)Residential and Work locations and (e)Microsoft Buildings Footprint dataset

4.1.4 Schedules

Our motive here is to represent all of the city/county's area and model the movement of people to and from each specified subregion. This is achieved by discretizing space and time to create the dataset. Each entry in the dataset refers to one person moving from their inital(home) location to their final(usually their workplace, may be different for different types of data sources) location. The data sources can be divided into two primary categories: (a) Data source for mass movement of people, usually on a census tract or census block group scale, (b) Data source for housing, work, and miscellaneous building locations throughout the region under consideration.

4.1.5 People's Movement data

Usually such datasets are obtained from census bureaus, which collect countywide information on people's travelling patterns, and their preferred destinations. We use the LEHD Origin-Destination Employment Statistics (LODES) United States Census Bureau (2017) data, which is publicly available from the United States Census Bureau. These may change over the years. This data is aggregated for each census block group, having the number of people travelling between census block groups. It also contains the number of workers, classified into age groups, wage groups and industry sector.

Another source could be individual organizations which use tracking technologies to record the activity of its user-base. These give us a temporal trend of the population's travels, but is not usually spatially widespread and mostly confined to cities. One such dataset is Safegraph SafeGraph (2020), which is also on a census block group scale of aggregation. This counts the number of people moving between their residence census block groups and different types of destinations (like, offices, grocery stores, entertainment places) which is also in a specific census block group. This includes the frequency and count of visits, and hence the a robust temporal distribution.

4.1.6 Building locations

As we have seen in the previous subsection, the data is usually aggregated on the census block group scale and we have no specific information about the exact home or work locations that are needed to formulate specific Origin-Destination pairs. It is imperative we find the locations of individual houses and workplaces in the concerned census block groups. Two such hierarchical methods are described here. OpenStreetMap OpenStreetMap contributors (2017) provides a labelled collection of buildings. They are individual geometric shapes on a two dimentional (2D) plane. These buildings are then represented as a point, which are the centroids of the building's shape. Since these buildings are tagged, we can classify them into homes or workplaces, and form a primary notion of the exact locations of where people are moving. One drawback of this method of data collection is that for smaller or less populous areas, the buildings are not properly tagged, making the dataset very small, and restricted to major population centers.

Census block groups that lie on the outskirts or far from cities are usually missing any sort of tagged buildings, hence a distinct lack of home or residential buildings can be seen. We are unsure of the exact locations of where people are starting and ending their trips. To get get some spatial clarity and increase the granularity of the people's movement locations, we use a secondary layer of building location information. This data layer usually covers a large number of buildings both in urban and rural areas, but is untagged, i.e., the buildings cannot be specifically classified as homes, workplaces, or if they are used for other purposes. Microsoft produced a U.S.-wide vector building dataset in 2018 Team (2018) that was generated from aerial images available to Bing Maps using deep learning methods for object classification Microsoft (2018). We use the US Buildings Footprint to find the set of untagged buildings. Combining the tagged and untagged buildings we can essentially cover the landscape under consideration.

4.1.7 Alternate Method

Use a SUMO tool od2trips to generate the person trips by incorporating transportation demand (in O format) and taz.xml. The geographic area under consideration is Chattanooga, Tennessee. The city is divided into census tracts. We find the movement matrix of people travelling for jobs (LODES) from census tract to census tract. The generated origin-destination(OD) pairs are used as requests for our solver. Each row in either dataset represents a single trip by one person. The trip represents movement to the job location and then back to home, at certain times of the day, which are sampled from a given set of regular job start and job end times. Another source could be individual organizations that use tracking technologies to record the activity of their user base. These give us a temporal trend of the population's travels but are not usually spatially widespread and mostly confined to cities. One such dataset is Safegraph.

4.2 CONVERT GTFS TO SUMO FORMAT

The General Transit Feed Specification (GTFS) is a data specification that allows public transit agencies to publish their transit data in a format that can be consumed by a wide variety of software applications. Today, the GTFS data format is used by thousands of public transport providers. GTFS is split into a schedule component that contains schedule, fare, and geographic transit information and a real-time component that contains arrival predictions, vehicle positions and service advisories. A GTFS feed is composed of a series of text files collected in a ZIP file. Each file models a particular aspect of transit information: stops, routes, trips, and other schedule data. The details of each file are defined in the GTFS reference. Figure 4.1 shows the manual Procedure of transit simulation using SUMO. We need to find bus stops' positions on network so first using TraCI to interact with SUMO (convertGeo.py) then get the position info of stops (including edge ID,

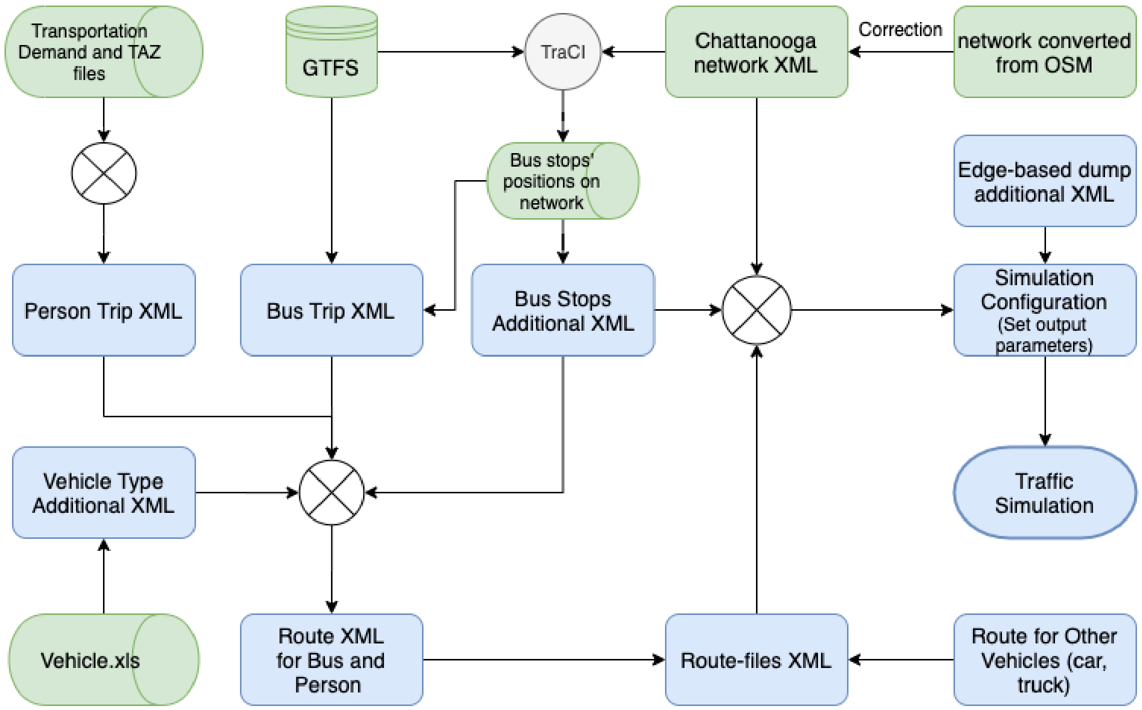


Figure 4.1. Procedure of transit simulation using SUMO

lane position and lane index) based on geo coordinates in GTFS. Finally, Create bus stop additional xml.

4.2.1 Literature review

4.2.2 Our approach

5 BACKGROUND TRAFFIC ELIMINATION

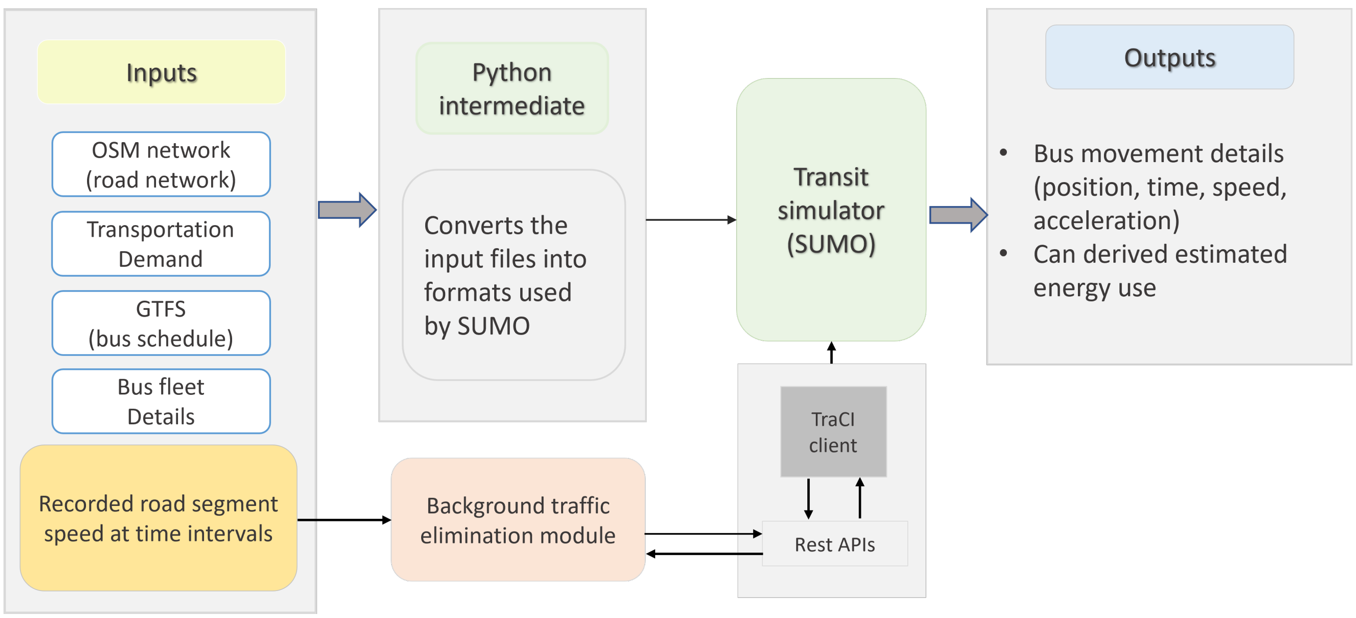


Figure 5.1. Transit simulation without background traffic

The public transit simulation is conventionally done by feeding all the data at once. This makes it slower to complete and requires more computational processing. To make this task easier and much faster, we leverage the use of calibrated background speeds in our proposed system, simulation with background traffic elimination (BTE-Sim). Both methods are shown here.

5.1 INPUTS TO THE SYSTEM

The simulation considers multiple inputs to set up the desired scenario:

a) The road network is selected from OpenStreetMap (OSM) – represents the city/regions roads

b) The transportation demand is obtained by generating OD pairs from publicly available datasets, like LODES (for jobs), or also from city organizations

c) The bus schedules are found from GTFS

d) We need to identify the number of individual cars on the road as well, to get a proper assessment of the traffic speeds

e) The calibrated background speeds (either from the full simulation, APC, or INRIX data)

This data is fed into the python intermediate which structures and converts them into SUMO (an open-source transit simulator) usable formats. SUMO performs the simulation for the desired periods (usually 24 hours) and outputs the movement details of each bus – like their speed, acceleration, distance covered, and so on. We can then use this data to find the average speed of each traffic lane and calibrate those lanes to the found speeds. This is the speed due to the background traffic (all other vehicles, except the buses). This is used further down the line in the updated type of simulation, which is a simulation using background traffic elimination.

5.2 WORKING

This form of simulation requires the same inputs as previously used, along with the calibrated background speeds, as shown in Fig.5.1. This background traffic speed can be received from one of three sources:

a) the background traffic speeds of the simulation discussed above

b) the automated passenger count (APC) dataset for the buses of the transit agency

c) INRIX, which is a dataset containing the speeds of vehicles on discrete road segments at very fine time intervals.

We can choose the background speed from any one of the above sources and we show in Experiment 1 [sec.6.2], amongst the choices, which background speed data is preferred for our system. Since we do not need to simulate the movement of all vehicles and focus only on bus movements, it can run for a much shorter duration. Thus, greatly reducing the execution time of a day's simulation, providing faster results (around 5 minutes on a regular PC). It functions similarly to the previous scenario, but the lane speeds are individually modified here, using a component of SUMO, called the TraCI (traffic control interface) client. It allows "online", as in while the sim is running, to control the lanes and buses. The outputs are details of the movements of the bus, which can be used to check for optimality or for other purposes, such as energy calculations

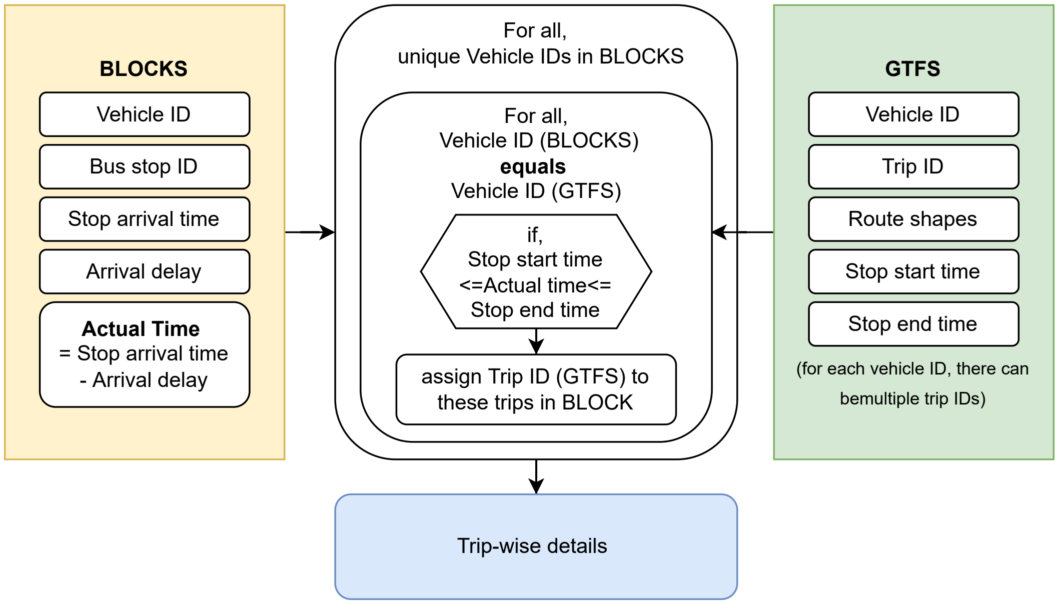


Figure 5.2. Block to trip conversion

5.3 BLOCK TO TRIP CONVERSION

The outputs we get from the simulation are in the form of blocks of trips. We need to reformat them into separate trips. This is done by using the conversion schema, as shown in Fig. 5.2. In order to reformat the outputs from a transit simulation into separate trips, we outline the steps needed to take the data in the form of blocks of trips and reformat it into individual trips and involves identifying the relevant fields for each trip such as the unique bus identifier, the time of the stops, and they delay in arrival. Then they are organized into separate records or rows for each trip and are compared to the GTFS data. The conversion schema produces the trip identifier for each movement of the bus in the original block-wise data.

6 SYSTEM MODULARITY

6.1 CHANGING ROUTE TRAFFIC PATTERNS

Setting: As mentioned above, the edge-speed data of BTE-Sim can be collected from different sources. In this experiment, we use BTE-Sim from three sources to simulate the transit system on Jan 11, 2022. The first source is from a prior run of TransitGym which uses the OD matrix to generate the background traffic. The second source is The INRIX traffic and road speed service INRIX (2021) which is collected from connected cars and mobile devices, cameras and sensors on roadways, and major events expected to affect traffic and is available through INRIX IQ a SaaS-based cloud platform. The third is APC, which as described earlier, records the time that buses arrived at each bus stop. In all the cases, we know the travel time of buses between two bus stops from the data. Moreover, we also know the distance of the bus stops, so we can estimate the average speed of the buses when moving from any bus stop to another.

Result: In Fig. 6.1, the absolute ToA error gets amplified for APC data while INRIX provides comparatively better results. BTE-Sim performs the best when using the background traffic times generated using TransitGym. Thus, we choose the background traffic speeds from TransitGym to be our traffic speed data for the rest of the simulations.

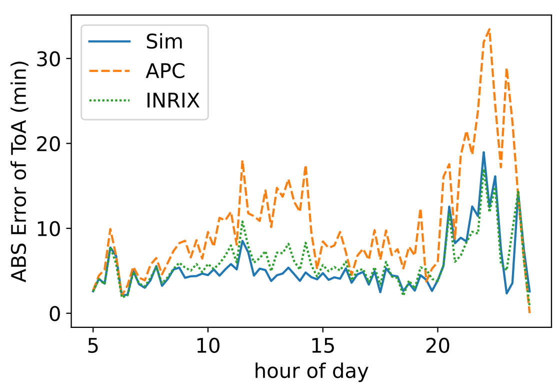


Figure 6.1. Comparing background traffic sources for BTE-Sim

6.2 CHANGING ROUTE TRANSPORTATION DEMAND

Setting: The transit demand (i.e., the number of transit commuters, and their travel routes) can change across dates. We use three different OD datasets for simulating the scenario of changing OD demand data. The datasets are disparate as each of them contains a varying number of people moving, to different locations throughout a day. For the same day of Jan 11, 2022 in Chattanooga, we test these OD variations to show the absolute ToA error.

Result: BTE-Sim shows minimal differences in the three OD situations as seen in Fig. 6.2 The changes in absolute ToA error are minuscule for a day's operation. This demonstrates that BTE-Sim is well equipped to maintain steady simulations even under varying situations.

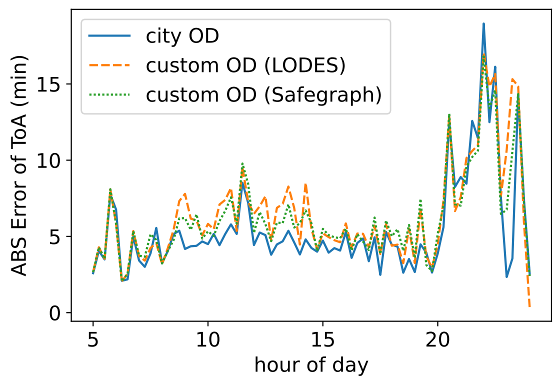


Figure 6.2. BTE-Sim for different OD data

6.3 CHANGING BUS SCHEDULE (GTFS)

Setting: BTE-Sim simulates for two weeks (08/15/2021 and 01/10/2022).

Result: As we can see in Fig 6.3, the simulated values are usually very less scattered, with the ToA values having very short inter-quartile ranges for a given day, with a comparatively higher dispersion on 01-12. Both the mean and maximum values of the absolute error of ToA are consistently under 10 minutes. With these, we can confirm that BTE-Sim has a very low error margin in simulating regular traffic and the operations of the transit systems on different days.

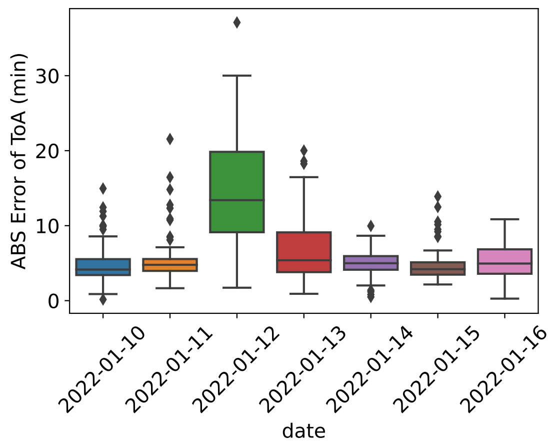


Figure 6.3. BTE-Sim for different GTFS

7 ANALYSIS

7.1 COMPARISON WITH TRANSIT GYM

Setting: We conduct simulations for the city of Chattanooga and for the day of January 11, 2022. We use real data from the day's transit operation in the simulation. To generate the background traffic for TransitGym we use the OD travel matrix provided by the city's planning agency.

Result: To evaluate the simulation, we focus on the difference of Time of Arrival (ToA) at bus stops between the simulation and real world times, on the same date. It is measured in minutes. Fig. 7.1 depicts the absolute error of ToA of Transit-gym and BTE. This demonstrates that using historical edge speeds, BTE can mimic effects of the background traffic, without which it would have no delays. But in accomodating background traffic, it has the capacity to change dynamically with the traffic conditions that may arise in the city. From this, we can also infer that BTE performs comparatively better at simulating the movement of buses across the city, having a consistently lesser ToA than TransitGym for an entire day's operation.

7.2 BTE-SIM IMPROVES THE SIMULATION TIME

Setting: For the city of Chattanooga, we investigate scenarios of varying number of vehicles on the road. There could be more vehicles, and to mimic that we increase the number of vehicles in

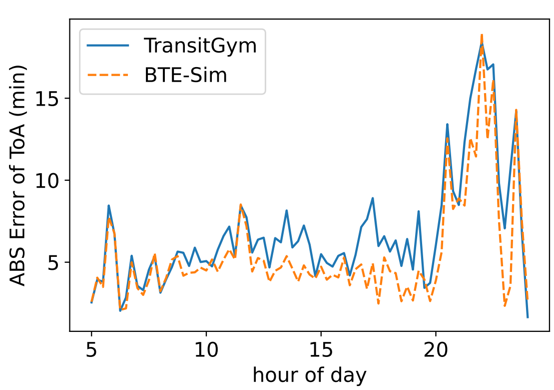


Figure 7.1. Comparing Transit-gym and BTE-Sim on absolute error of Time of Arrival

the OD matrix.

Result: Table 7.1 records the execution times of both the simulation methods. For a baseline of 100,000 vehicles, we can notice that BTE-Sim runs more than 12 times faster than traditional Transit-Gym. And with a varying number of vehicles, the computation speed of the simulations also changes drastically. As the number of vehicles increased by a factor of 4, TransitGym execution time increased by 8 times, whereas BTE-Sim computation time increased by a factor of 2. BTE-Sim runtime increases minimally with a huge increase in vehicular traffic, and is highly adaptable to traffic volume changes and can be re-run for increased traffic scenarios without much time penalty.

#Vehicles	TransitGym	BTE-Sim
100K	27.7 minutes	2.21 minutes
400K	4 hours 4 minutes	5.11 minutes
800K	16 hours 51 minutes	7.81 minutes
1400K	41 hours 18 minutes	8.27 minutes

Table 7.1. Simulation time of TransitGym and BTE-Sim for scenarios with different number of vehicles

7.3 BTE-SIM SIMULATES DIFFERENT DATES

Setting: We demonstrate the operation of BTE-Sim by simulating for a week from Jan 10, 2002 to Jan 16, 2022. Note that the transit setting of various dates is different. For example, a trip may be offered on Monday but unavailable on Tuesday.

Result: As we can see in Fig 7.2, the simulated values are usually very less scattered, with the ToA values having very short inter-quartile ranges for a given day, with a comparatively higher dispersion on 01-12. Both the mean and maximum values of absolute error of ToA are consistently under 10 minutes. With these we can confirm that BTE-Sim has a very low error margin at simulating regular traffic and the transit systems operations over different days.

Through all the experimentation we successfully show that BTE-Sim is capable of performing consistently under dynamic scenarios for city wide transportation operations. It is able to handle large scale traffic, and people' movements along with its primary focus of public transportation. The simulation time is significantly faster than existing methods and continue to be so under increased traffic load.

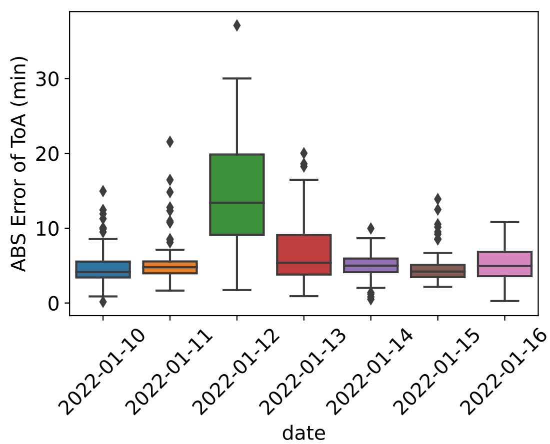


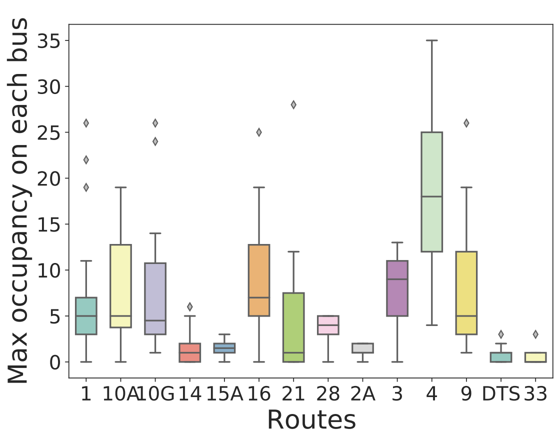
Figure 7.2. BTE-Sim over different dates

8 CASE STUDIES & RESULTS

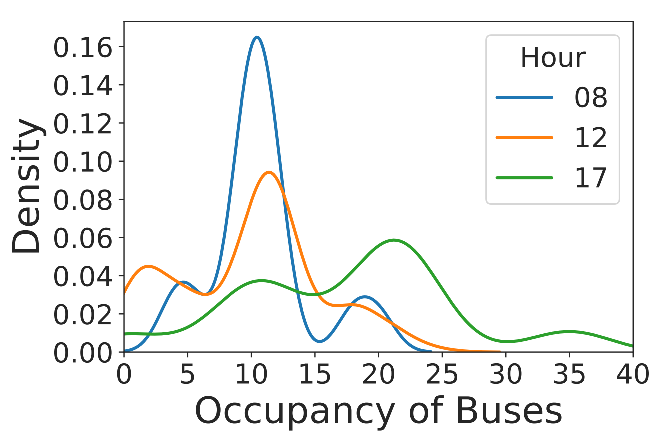
8.1 CHATTANOOGA - 01/11/2022

On January 11, 2022, the transit simulation of Chattanooga was conducted to analyze the movement of buses during the day. The simulation examined multiple parameters including occupancy, speed, distance traveled, and energy usage. The results show the occupancy variance and the variation in the density of occupancy of the buses (Fig. 8.1a). Occupancy for different hours of the day is illustrated in Fig. 8.1b The average speed of the buses was 22.5 mph, with a maximum speed of 35 mph and a minimum of 12 mph. The total distance traveled by the buses was 1,739 miles, with an average energy usage of 1.9 megawatt-hours.

Overall, the simulation provided valuable insights into the efficiency and performance of the Chattanooga transit system. It identified areas where improvements could be made, such as increasing the occupancy of the buses and reducing energy usage. The simulation also highlighted the importance of monitoring and analyzing various parameters to optimize the transit system and ensure that it is meeting the needs of the community. By analyzing these parameters regularly, the transit system can continuously improve and provide a reliable and efficient service to its users.



(a) Maximum occupancy



(b) Density of Route 4

Figure 8.1. Analysis examples for the transit system of Chattanooga on January 11, 2022, using BTE-Sim

9 SUMMARY

- We proposed the Background Traffic Elimination module to speed up transit simulations while obtaining competitive results compared to conventional simulations.
- We present a novel downstream task of transit simulations i.e., evaluating the OD matrices.
- The underlying road infrastructure is addressed. We generate the necessary data and define it according to our use case.
- The input data are listed and discussed. The sources and methods of finding them are discussed.
- ‘The system’s working is described starting with an overview and explanation of the details, like the inputs and the working.
- We conducted comprehensive experiments to demonstrate the operation of the system, on evolving transit scenarios, and multiple days to demonstrate its modularity.
- Simulate an entire day and show sample outputs.
- Could be used to make energy estimations for the desired periods of transit simulation

10 FUTURE WORK

- Filter trips in an optimized way to reduce the reliance on any particular dataset, like Clever bustime, or APC.
- Introduce door-to-door vans to simulate on-demand transit and micro-transit

REFERENCES

INRIX (2021). "Inrix delivers products for the automotive and transportation industries such as real-time parking and traffic information and solutions that facilitate the safe testing and deployment of autonomous vehicles.

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., and Wießner, E. (2018). "Microscopic traffic simulation using sumo." *2018 21st international conference on intelligent transportation systems (ITSC)*, IEEE, 2575–2582.

Microsoft (2018). "Us building footprints.

OpenStreetMap contributors (2017). "Planet dump retrieved from https://planet.osm.org .

SafeGraph (2020). "Places schema. retrieved from https://docs.safegraph.com/docs/places-schema.

Team, B. M. (2018). "Computer generated building footprints for the united states.

United States Census Bureau (2017). "Longitudinal Employer-Household Dynamics - Origin-Destination Employment Statistics.

Urquiza-Aguiar, L., Coloma-Gómez, W., Bautista, P. B., and Calderón-Hinojosa, X. (2020). "Comparison of sumo's vehicular demand generators in vehicular communications via graph-theory metrics." *Ad Hoc Networks*, 106, 102217.