Data-driven Multimodal Transportation Energy Consumption Prediction and Analysis Framework for Sustainable Transit and Transportation Planning

Final Report

by

Yuche Chen, University of South Carolina
Nathan Huynh, University of South Carolina
Gurcan Comert, Benedict College
Yunteng Zhang, University of South Carolina

Contact information
Yuche Chen, Ph.D.
300 Main St., Columbia, SC 29208
University of South Carolina
Phone: (803) 777-9105; *E-mail:* chenyuc@cec.sc.edu

August 2020

Center for Connected Multimodal Mobility (C^2M^2)

200 Lowry Hall, Clemson University
Clemson, SC 29634

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by the Center for Connected Multimodal Mobility (C^2M^2) (Tier 1 University Transportation Center) Grant, which is headquartered at Clemson University, Clemson, South Carolina, USA, from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle

Data-driven Multimodal Transportation Energy Consumption Prediction and Analysis Framework for Sustainable Transit and Transportation Planning

5. Report Date

August 2020

6. Performing Organization Code

7. Author(s)

Yuche Chen, Ph.D.; ORCID: 0000-0003-2577-2448
Gurcan Comert, Ph.D.; ORCID: 0000-0002-2373-5013
Nathan Huynh, Ph.D.; ORCID: 0000-0002-4605-5651
Yunteng Zhang; ORCID: 0000-0003-0939-0159

8. Performing Organization Report No.

9. Performing Organization Name and Address

Department of Civil and Environmental Engineering
University of South Carolina
300 Main St., Columbia, SC 29208

10. Work Unit No.

11. Contract or Grant No.

69A3551747117

12. Sponsoring Agency Name and Address

Center for Connected Multimodal Mobility (C^2M^2)
USDOT Tier 1 University Transportation Center
Clemson University
200 Lowry Hall, Clemson
Clemson, SC 29634

13. Type of Report and Period Covered

Final Report (02/01/2019 – 06/15/2020)

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The transportation sector is a major energy consumer and contributor to air pollution. Public transportation authorities can often find it challenging to choose the best projects among a spectrum of candidate transportation projects to reduce energy consumption and mitigate emissions. The objective of this project was to develop a sketch environmental planning framework to integrate traffic simulation and regulatory transportation air quality modeling (i.e., MOVES) to provide robust statistical correlations of the energy and air quality impacts with changes in traffic activities for transit operations and planning. Firstly, a traffic simulation was conducted on the road network of the city of Columbia, South Carolina. The simulations were validated with real-world traffic data acquired from iPeMS in South Carolina to demonstrate the representativeness of the traffic simulation results. iPeMS is the real-time data analysis, visualization, and reporting platform for South Carolina statewide traffic information. Secondly, outputs of traffic simulation, i.e. link-level vehicle trajectory data are processed to prepare inputs for energy/emissions simulation using EPA's Motor Vehicle Emissions Simulator. The processed inputs include vehicle mile traveled, operating mode distribution, etc. The results of emissions simulation are link-level energy/NO_x emission rates on each link for transit bus driving of the whole transportation road network. Then, we utilized machine learning models to develop a statistical correlation between the aggregated traffic activity patterns and MOVES-calculated emissions. The test set predictions showed that our models can accurately estimate MOVES-based energy and emissions rates (within a 10% mean absolute percentage error, MAPE) given aggregated traffic activity data. We also improved prediction accuracy using our neural network model, then the regular regression model. We tested the spatial transferability of our developed models to predict MOVES energy/emissions results using the road network activities of a medium-sized city. Our neural network models can achieve a 15% mean absolute percentage error relative to MOVES results. This demonstrates the transferability of the model results. With traffic data of a whole network, we are able to generate a heat map to show the energy consumption rate of operating transit buses on any link of the road network.

17. Keywords	18. Distribution Statement
Energy Consumption; Sustainable Transit Planning; Machine Learning Model	No restrictions.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	37	NA

Table of Contents

DISCLAIMER ii

LIST OF TABLES v

LIST OF FIGURES v

CHAPTER 1. Introduction and Background 3

CHAPTER 2. Literature Review 5

CHAPTER 3. Simulation and Data Preparation 8

- 3.1 Traffic Simulation 8
- 3.2 Energy Consumption Data Preparation using MOVES 14

CHAPTER 4. Feature Selection, Model Development and Results 16

- 4.1 Feature Selection 16
- 4.2 Model Development 18
- 4.3 Model Results 19

CHAPTER 5. Spatial Transferability Test and Discussion 23

- 5.1 Spatial Transferability Test 23
- 5.2 Discussion 26

CHAPTER 6. Conclusion 27

REFERENCES 29

LIST OF TABLES

Table 1. NETCONVERT Command Options 9
Table 2. Notation and Description of Model Variables 13
Table 3. Mallow's Cp for Fuel Consumption Rate on Restricted Roads 16
Table 4. Model Performance Matrix 22

LIST OF FIGURES

Figure 1. Project Framework Overview 4
Figure 2. SUMO Road Network (Left); OpenStreetMap of City of Columbia, SC (Right) 8
Figure 3. Snapshot of NETEDIT Interface 10
Figure 4. Extract of SUMO road network (Left) and real-world map counterpart (Right) 11
Figure 5. SUMO output of vehicle trajectories 12
Figure 6. Comparison between Simulation Results and Real-World Observations for Speed Distributions 12
Figure 7. Comparison between simulation results and real-world observations of detailed traffic patterns 13
Figure 8. Sample MOVES Output 15
Figure 9. Model Structure 18
Figure 10. Fuel Consumption Rate Map for City of Columbia 19
Figure 11. Model Prediction vs. MOVES Output for Fuel Consumption Rate on Restricted Access Roads 20
Figure 12. Model Prediction vs. MOVES Output for NO_x Emission Rate on Restricted Access Roads 20
Figure 13. Model Prediction vs. MOVES Output for Fuel Consumption Rate on Unrestricted Access Roads 21
Figure 14. Model Prediction vs. MOVES Output for NO_x Emission Rate on Unrestricted Access Roads 22
Figure 15. Flowchart for Project 23
Figure 16. Eichstätt SUMO network (Left) and Openstreetmap Counterpart (Right) 24
Figure 17. Model Prediction vs. MOVES Output for Fuel Consumption Rate in City of Eichstätt 25
Figure 18. Model Prediction vs. MOVES Output for NOx Rate in City of Eichstätt 25

EXECUTIVE SUMMARY

The transportation sector is a major energy consumer and contributor to air pollution. Governments around the world are taking steps to address the energy and air pollution problems caused by transportation. A portfolio of strategies should be employed to mitigate air pollution and fossil fuel dependence related to transportation. Transportation agencies are implementing a variety of projects, e.g., road network expansion, intersection signal optimization, alternative fuel fleet upgrades, and travel demand management. Public transportation offers the advantage of conveying a larger passenger volume in less space than private automobiles and thus has the potential to meet sustainability goals in the transportation sector. United States regulations mandate that all transportation projects go through a conformity analysis process for energy and air quality evaluation. The evaluation process consists of two steps: an assessment of changes in traffic activities and the quantification of impacts on energy and air quality from traffic activities. The first assessment step is usually straightforward because traffic data are archived, and transportation practitioners are experienced in traffic analysis. However, the second step is demanding. It requires a significant effort to prepare data and run specialized software tools, such as the Motor Vehicle Emissions Simulator (MOVES). Specifically, public transportation authorities can often find it challenging to choose the best projects among a spectrum of candidate projects. It is time-consuming and labor-intensive to evaluate the energy and air quality of all the candidate projects to assess the respective cost benefits in achieving sustainability goals. Therefore, transportation practitioners require a sketch planning tool that can rapidly evaluate the energy and air quality of candidate projects, whereby a limited number of preferred projects can be identified to conduct the regulatory time-consuming evaluation procedure.

The objective of this project was to develop a sketch environmental planning framework to integrate traffic simulation and regulatory transportation energy and air quality modeling (i.e., MOVES) to provide robust statistical correlations of the energy and air quality impacts with changes in traffic activities for transit operations and planning. We used advanced machine learning algorithms to estimate statistical correlations. Once established, these statistical correlations can be applied to other locations to evaluate the transportation energy and air quality impacts as long as traffic-related data are provided. Therefore, public transportation practitioners only need to focus on the first step of the two-step analysis, wherein they have the most expertise.

In this study, a traffic simulation was conducted on the road network of the city of Columbia, South Carolina. The simulations were calibrated using real-world traffic data acquired from iPeMS in South Carolina and were thus representative of real traffic. iPeMs is the real-time data analysis, visualization, and reporting platform for South Carolina statewide traffic information. The traffic simulation outputs included second-by-second travel trajectories for each vehicle and bus in the network, as well as hourly aggregated traffic patterns at the road link level at every hour, e.g., the average speed and the percentage of the idling time. For each road link, which is the smallest component of a road network, second-by-second travel trajectories were used as inputs to the MOVES model to calculate the energy consumption rate and the emission rate of NO_x (a regulated criteria pollutant) of buses. Ideally, the use of second-by-second travel trajectories for buses as inputs results in accurate transportation energy and air quality analysis. However, aggregated traffic patterns data are mostly available only to transportation authorities. Therefore, we utilized machine learning models to develop a statistical correlation between the aggregated traffic activity

patterns and the MOVES energy/emissions results, which are required for regulatory transportation air quality analysis. We divided the simulation data into a training set and a test dataset. The training set was used to develop a total of four artificial neural network models to calculate the link-level fuel consumption rate and the NO_x emission rate on restricted and unrestricted urban roads. Model features were selected iteratively to determine the optimal neural network model setup and parameters. The test set predictions showed that our models can accurately estimate MOVES-based energy consumption rates (within a 10% mean absolute percentage error) given aggregated traffic activity data. We tested the spatial transferability of our developed models using the road network activities of a medium-sized city. Our neural network models can achieve a 15% mean absolute percentage error relative to MOVES link-level energy consumption rates. This demonstrated the transferability of our model. With traffic data of a whole network, we are able to generate a heat map to show the energy consumption rate of operating transit buses on any link of the road network.

CHAPTER 1
INTRODUCTION AND BACKGROUND

The transportation sector, dominated by petroleum fuels, is a major energy consumer and a major cause of air pollution and related health effects worldwide (Frey, 2018; Krzyżanowski et al., 2005). Concerns about air quality and global climate change, have caused governments around the world to develop various strategies to reduce energy and emissions from the transportation sector. Public transportation is one such strategy that offers the advantage of conveying a larger passenger volume in less space than private automobiles (White, 2016). Additionally, emerging technologies, including automation, the Internet of Things, and sharing economy, enable innovations in transit operations to increase the potential of achieving sustainability goals in the transportation sector (Shen et al., 2018). According to the United States Bureau of Transportation Statistics, fuel cost accounts for 20% of the total operating cost in transit services (BTS, 2018). Although transit fleets are increasingly using renewable sources, the internal combustion engine (ICE) bus remains the major fuel consumer and accounts for 60% of the transit fleet across the United States (DOE, 2019). Energy-saving-oriented operation of the internal combustion engine transit bus plays a key role in achieving sustainable targets in the public transportation sector (Beaudoin et al., 2015). These targets can be achieved by providing or predicting energy consumption information for drivers and transit system managers to better plan operations of transit services (Xu et al., 2017). It is challenging to obtain this information for transit buses because of diverse driving conditions and the spatial and temporal characteristics of transit operations and routes (Silva et al., 2015).

Researchers have developed physical models that mimic energy conversion flows within vehicles to estimate the energy consumption of transit buses under specific driving conditions. Examples of physical models are the Future Automotive Systems Technology Simulator (FASTSim) (Brooker et al., 2015), developed by the National Renewable Energy Laboratory, and the Comprehensive Modal Emissions Model (CMEM) (Scora and Barth, 2006), developed by the University of California, Riverside. These physical models normally require inputs of the vehicle components (the weight, aerodynamic coefficients, frontal area, etc.) and the vehicle operational status (the speed, acceleration, etc.) at every second to predict the energy consumption at the same frequency. These models can be used in real-time transit bus eco-driving applications but are not useful at the energy-oriented transit bus operation and route planning stage. During the planning stage, traffic condition data are normally aggregated into a specific spatial (e.g., road segment) or temporal (e.g., 15-minute average) granularity. For transportation planning, practitioners use vehicle energy and emission inventory models to evaluate energy and emissions (where CO_2 emission is a proxy for energy consumption) from the transportation sector. The Motor Vehicle Emission Simulator (MOVES), developed by the United States Environmental Protection Agency (US EPA), is one such model in widespread use by regional planning organizations and transit authorities. MOVES can predict the energy consumption and emissions for a road segment by two methods. One method is based on a predefined average speed and energy consumption or a CO_2 emission lookup table. This method is easy to use but sacrifices prediction accuracy because the same average speed on a road segment can result from a variety of driving patterns corresponding to different emission rates. In the second method, predictions are based on a vehicle operation mode matrix constructed from the second-by-second driving trajectories of all the vehicles on a road segment. This method produces very accurate results but requires highly detailed input data that is cumbersome to collect. Obtaining data from real-world GPS vehicle tracking is costly and may infringe on privacy issues. The most frequently used alternative is to generate data using traffic simulations. Researchers and practitioners have expended considerable effort into integrating MOVES with traffic simulation models to quantify transportation energy consumption and emissions. However, no microsimulation model is likely to precisely predict actual traffic changes. The calibration of simulation models to represent real-world observed traffic patterns requires an enormous expenditure of effort, making microsimulations computationally intensive.

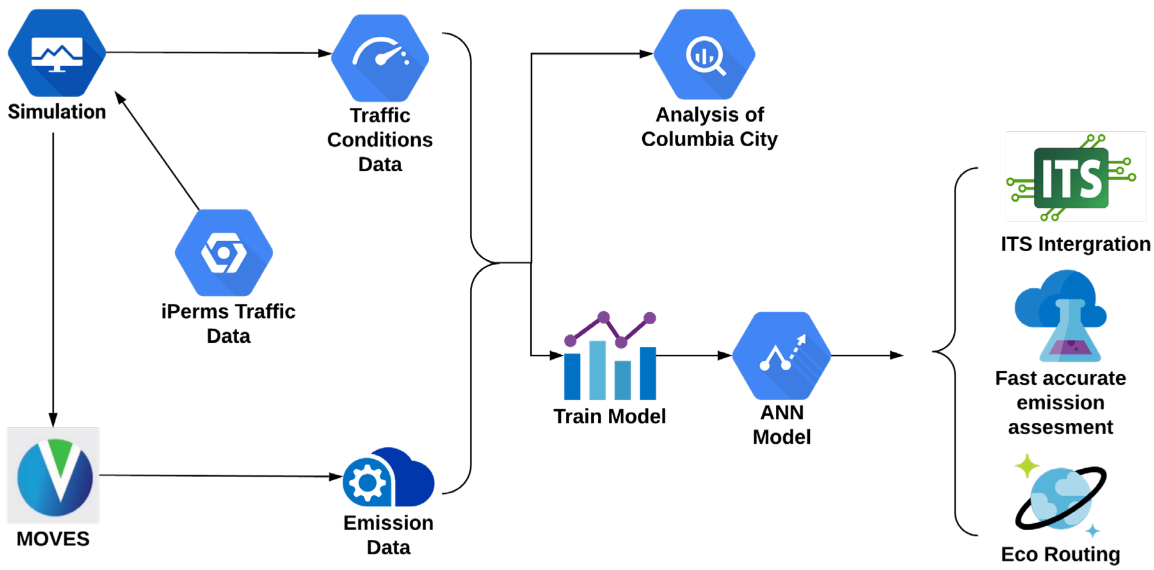


Figure 1. Project Framework Overview

The objectives of this study are as follows.

1. A traffic simulation model is integrated with a microscopic energy and emission inventory model to achieve high accuracy in transit bus energy prediction without intensive data preparation.
2. A machine learning-based surrogate model is developed to estimate vehicle-related emissions under different traffic patterns. The model can serve as a sketch planning tool for researchers and transportation air quality practitioners to quickly assess bounds on emissions benefits for traffic operational strategies.

The remainder of this report is organized as follows: Chapter 2 is a literature review of related research; the data processing and preparation are described in Chapter 3; the model development is discussed in Chapter 4; the spatial transferability of the model is presented in Chapter 5. The study is summarized and conclusions are drawn in Chapter 6.

CHAPTER 2
LITERATURE REVIEW

A detailed literature review on transportation energy and emission estimations is presented in this chapter. Transit operation and planning applications are the primary focus. In the literature, methodologies to estimate the energy consumption and emissions of transit vehicles can be classified mainly into summary or estimation models. Based on the modeling resolution, estimation models can be further categorized into microscopic and macroscopic models.

Summary models comparatively analyze trips or the daily average energy consumption of transit buses by using real-world measured data. These models do not explore the statistical relationship between fuel consumption and influencing variables but require a large number of measurements to ensure statistically robust results. Zhang et al. (2014) and Wu et al. (2015) used portable emission measurement systems to test several transit buses in Beijing and Macao, respectively, and found that natural gas buses have comparable CO_2 emission factors but higher fuel consumption relative to diesel buses. Ma et al. (2015) reported that the variance in fuel consumption from different driving styles exceeds 10% to 20% under different road conditions, even for experienced bus drivers. Giraldo and Huertas (2019) reported simultaneous measurements of fuel consumption, driving patterns, and CO_2, CO, and NO_x emission factors for diesel passenger buses under real operating conditions in high altitude cities (> 2000 masl) and mountainous regions with an average road grade of 4%. Yu et al. (2019) analyzed the annual data of 50 China IV public transit buses fueled with diesel in five cities and reported the speed and acceleration based on real road NO_x emissions and fuel consumption characteristics. Carrese et al. (2013) utilized data from buses in the city of Rome, Italy, and found variations in the average fuel consumption for buses driven by different drivers on different routes. Yu et al. (2016) analyzed the passenger load adjustment factor for estimating the diesel bus trip-level fuel consumption rate based on real-world data from Nanjing, China. Frey et al. (2007) analyzed the average fuel consumption rates of a diesel bus for different VSP bins at 2-kW/ton intervals using real-world driving data. A method for estimating the trip-level average fuel consumption rate of diesel buses was proposed using inputs from the fuel consumption rate database. These summary models provide useful information on the fuel consumption of transit buses, but the results are ad hoc and specific to vehicles at the measurement time and locations. These models can be used to access trending improvements in the fuel consumption of transit fleets but are not transferrable to other locations and applications.

Estimation models are different types of statistical models used to predict the energy consumption of transit buses. Based on the modeling granularity, these models can be further categorized into microscopic and macroscopic models.

Macroscopic models typically estimate the vehicle fuel consumption rate based on factors such as the average travel speed, the vehicle type, and the model year. The estimation relationship is typically available as a lookup or mapping table, such as those of major energy and transportation emission inventory models (Annual Energy Outlook (EIA, 2016), MOBILE6 (EPA, 2003), etc.).

The prediction parameters are often in discrete format. For example, MOVES can estimate the fuel consumption rate of vehicles given the vehicle type, fuel type, and average speed at 5-mph intervals. Macroscopic fuel consumption models have been used for eco-routing, trip assignment, and transit or passenger car planning (Penic and Upchurch, 1992; Sugawara and Niemeier, 2002). A major disadvantage of using macro models in transit planning and operation problems is that heterogeneity in driving is not considered such that two different driving trajectories with the same average speed will produce the same fuel consumption or emissions output.

The disadvantages of macroscopic models have led researchers to explore microscopic models for fuel consumption and emissions estimation. Fuel consumption or emissions are estimated at the most granular level, typically at every second of vehicular travel. These models are based on either a statistical or physical approach. The CMEM (Barth, Norbeck and Ross, 1996) was one of the first physical models and uses simplified physical descriptions to estimate the engine-out energy and emissions per second based on the total engine output power. Rakha and colleagues developed the VT-Micro model, which statistically estimates fuel consumption based on the second-by-second vehicle travel speed and acceleration (Rakha, Ahn and Trani, 2004). Microscopic fuel consumption models have been applied to eco-routing problems. Rakha, Ahn and Moran (2012) integrated the VT-Micro model into the INTEGRATION microscopic traffic assignment and simulation framework to solve an eco-routing problem. Traffic simulations were used for each link to produce second-by-second drive profiles, which were applied to VT-Micro to estimate fuel consumption. Nie and Li (2013) proposed an environmentally constrained shortest-path problem and developed a CMEM-based estimation model for carbon dioxide emissions (which are directly related to fuel consumption). The CMEM model is used to simulate the carbon dioxide emissions per second, from which optimal routing strategies are dynamically determined to minimize environmental impacts. In summary, microscopic fuel consumption estimation models are appropriate when second-by-second vehicle trajectory data is available. However, obtaining microlevel data through GPS tracking may violate privacy and is time-consuming when obtained by traffic simulations, particularly for large network applications (Nagel and Schleicher, 1994).

All the aforementioned studies lead to the conclusion that macro models are convenient but low inaccuracy, whereas data acquisition at a fine granularity for micro-models requires an excessive expenditure of effort. Although studies have attempted to streamline micro-models to reduce the effort required, these models remain cumbersome to use. The use of results from existing micromodels as inputs to a hybrid model for prediction or evaluation has not been explored. Preliminary estimates lead us to conclude that a deep-learning hybrid model can be developed to capture the complex relationship between traffic condition-related information and emission rates, such that the energy consumptions and emissions of U.S. cities could be evaluated using by existing micro-models for selected cities.

Most existing studies have used linear or nonlinear regression-based prediction methodologies for fuel consumption and emissions estimation models. Delgado et al. (2011) developed a method in which the average fuel consumption of diesel buses is first determined for representative driving cycles, and the fuel consumption rate of any driving cycle is then estimated using a linear combination of representative cycles. This method is less complex than regression or other statistical methods but cannot be used to help drivers improve fuel consumption in practice. Xie et al. (2012) developed a platform that integrates traffic simulation PARAMICS with US Environmental Protection Agency's vehicle emission model, MOVES to provide reliable fuel savings and emissions mitigation information by alternative fuel vehicles. Specifically, they utilized origin-destination data to simulate traffic on the regional road network of Greenville, SC and aggregate the traffic information into hourly data to serve as input to MOVES to estimate energy and emissions under various alternative fuel vehicle penetration scenarios. Tang et al. (2016) used a regression model with categorical variables to evaluate the spatial and temporal impacts of diesel bus energy efficiency. The considered impact factors were the time of day, the day of the week, and the road type. López-Martínez et al (2017) applied a similar approach to analyze bus data from Madrid, Spain. Hung et al. (2012) adopted a regression model with an exponential format to estimate the instantaneous fuel consumption of transit buses and passenger vans based on instantaneous speeds for accelerating, cruising and decelerating driving modes. The fuel consumption for the idling mode was also estimated based on the cumulative idling time. Pan et al. (2019) analyzed and adopted a VSP-based gradient boosted regression tree approach to estimate emission rates (CO, CO_2, HC, and NO_2) for a liquefied-natural-gas bus under real-world driving conditions but did not estimate the energy consumption. The VSP metric is the vehicle-specific power used to represent engine loads resulting from aerodynamic drag, acceleration, rolling resistance and hill climbing scaled by the vehicle mass. He et al. (2018) proposed a VSP-based binning method to assess the on-road energy consumption of battery electric buses under complex real-world usage patterns using second-by-second data from transit buses in Macao, China. Wu et al. (2013) adopted a similar VSP-based binning method to evaluate the fuel consumption of diesel buses in Beijing, China. Silva et al. (2015) used multivariate regression models to identify the following influence factors for the trip-level average energy efficiency of a bus fleet: the mixture of vehicle types, the commercial speed, and the percentage of road with grades over 5%. Wang and Rakha (2016) used the framework of the Virginia Tech comprehensive power-based fuel consumption model to improve bus fuel consumption modeling by circumventing the bang-bang control problem and determined optimum fuel economy cruising speeds between 40 and 50 km/h; a fuel consumption model was developed using the same framework to quantify the benefits of hybridization technologies for buses relative to conventional diesel bus operations (Wang and Rakha, 2016). Wang and Rakha (2017) also developed a convex second-order polynomial fuel consumption model for conventional diesel and hybrid-electric buses and found that the optimum fuel economy cruise speed ranges from 39 to 47 km/h for all buses tested on 0% to 8% grades: this optimum speed decreased as the grade and vehicle load increased.

CHAPTER 3
SIMULATION AND DATA PREPARATION

3.1 Traffic Simulation

We used an advanced traffic simulation package, the Simulation of Urban Mobility (SUMO), to construct microscopic traffic simulation models to correlate traffic condition-related information, the energy consumption rate and the pollutant emission rate, as well as to obtain input data for training the machine learning-based model proposed in this study.

Simulation of Urban Mobility (SUMO) is an open source, highly portable, microscopic and continuous road traffic simulation package designed for large road networks. SUMO is a free and open traffic simulation suite that has been available since 2001 and can be used to model intermodal traffic systems, including road vehicles, public transport and pedestrians. Included with SUMO is a wealth of supporting tools for tasks such as route finding, visualization, network import and emission calculation. SUMO can be enhanced with custom models and provides various APIs to remotely control a simulation.

In this study, traffic simulations were performed for the road network of the city of Columbia, SC. We sourced geographic data from OpenStreetMap (OSM), a collaborative project that provides fine calibrated map data, and used SUMO to generate the road network for the simulations. The complete road network is shown in Figure 2.

Figure 2. SUMO Road Network (Left); OpenStreetMap of City of Columbia, SC (Right)

Given the large quantity of map data, the OSM data was acquired using the Overpass API (known as OSM Server Side Scripting), which is a read-only API that serves up custom selections of the OSM map data. The API acts as a database over the web: a client sends a query to the API and receives a corresponding data set.

As the OpenStreetMap is a map database created, updated and used by people around the world freely under an open license, effort is not expended to calibrate map details, especially for rarely accessed areas. Consequently, there is a considerable amount of incorrect road information in the

Columbia OSM map. Therefore, before downloading the OSM data, we expended considerable effort in manually checking the map, including supplementing missing information, such as traffic lights, speed limits and road categories, and correcting errors in the road topology, such as the number of lanes and road connections, in addition to modifying many other issues.

We used the NETCONVERT package that comes with SUMO to import the OSM map into the SUMO loadable road networks. The NETCONVERT package is a command line application, and we used the following command to generate the road network:

```
netconvert --osm-files Columbia.osm --output-file
Columbia.net.xml --geometry.remove --roundabouts.guess --
ramps.guess --junctions.join --tls.guess-signals --tls.discard-
simple --tls.join --no-internal-links --keep-edges.by-
vclass passenger --remove-edges.by-
type highway.track,highway.services,highway.unsurfaced --
proj.plain-geo true
```

where the large size of the network required the inclusion of many options, which are explained in Table 1 below.

Table 1. NETCONVERT Command Options

Option	Explanation
--osm-files	Reads OSM-network file
--output-file	Specifies name of generated network
--geometry.remove	Replaces nodes that only define edge geometry by geometry points
--roundabouts.guess	Ensures correct right-of-way at roundabouts
--ramps.guess	Identifies roads likely to have additional acceleration/deceleration lanes and adds these lanes, which are often not included in OSM data
--junctions.join	Applies a heuristic to automatically join close junctions that should be considered as single junction clusters; some junction clusters are too complex for the heuristic and must be checked manually thereafter
--tls.guess-signals	Interprets TLS nodes surrounding an intersection as signal positions for a larger TLS
--tls.discard-simple	Does not retain traffic lights at geometry-like nodes loaded from other formats
--tls.join	Clusters close tls-controlled nodes to better represent actual map
--no-internal-links	Omits internal links in junctions
--keep-edges.by-vclass	Retains edges for which VClass can be specified
--remove-edges.by-type	Remove edges of specified types

--proj.plain-geo	Writes geo-coordinates in network to facilitate future analysis

As the automation of the process can produce a considerable number of incorrect road topological errors in the resulting SUMO road network, we utilized the GUI-based tool NETEDIT to check the network and added, corrected and removed erroneously generated road lanes or junctions. The interface of NETEDIT is shown in Figure 3.

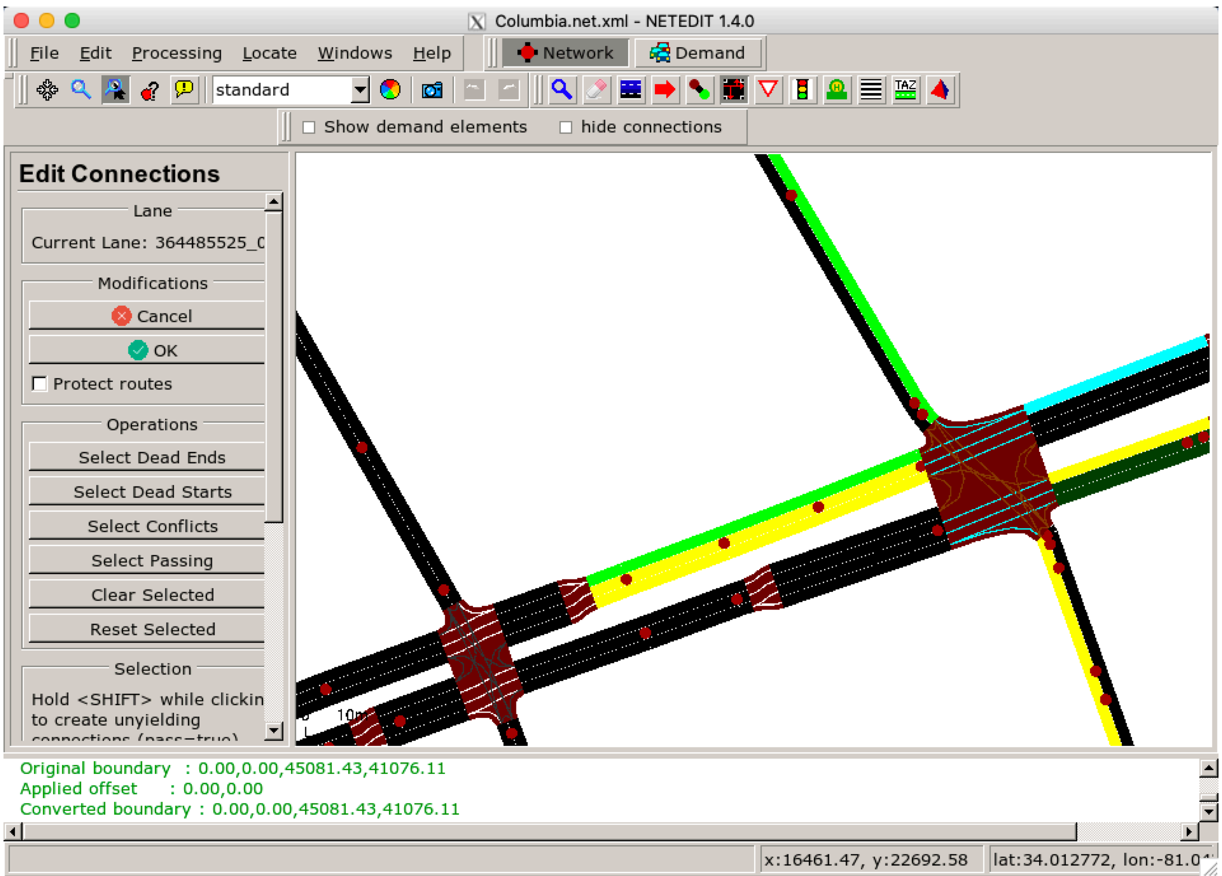


Figure 3. Snapshot of NETEDIT Interface

The SUMO road network includes very detailed road information and thus can reproduce real-world infrastructure at a high resolution. A zoomed-in view of an extract of the road network is shown in Figure 4.

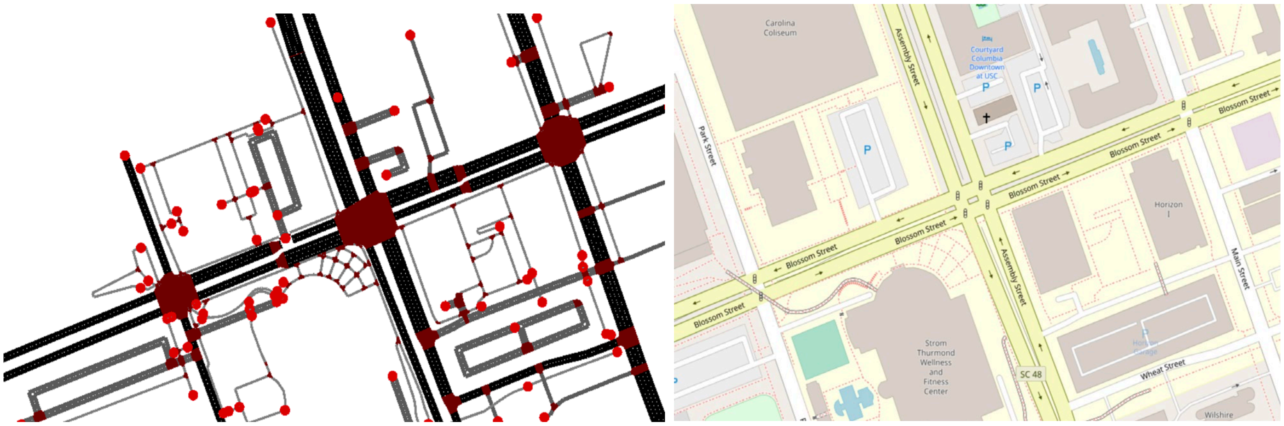


Figure 4. Extract of SUMO road network (Left) and real-world map counterpart (Right)

We utilized the Python script tool that comes with SUMO, randomTrips.py, to generate trips with randomly distributed origins and destinations. The numerous tool parameters, such as the probabilities of selecting specific nodes, the proportions of heavy-duty vehicles and the total flow going through the network, were calibrated by running the tool until the simulated traffic conditions matched the observed data.

We carefully calibrated the simulation parameters by comparing general traffic conditions and detailed traffic patterns of selected roads using traffic data obtained from iPeMS, thereby ensuring that the simulations' results reflected real-world traffic conditions for the network selected as input data to the model. iPeMS is the real-time data analysis, visualization and reporting platform for South Carolina statewide traffic information (Iteris, 2020). The platform leverages real-time and historical traffic data from HERE Technologies to provide features of traffic information in South Carolina statewide road network, such as dynamic maps to support detailed traffic analysis at different time aggregation, real-time congestion identification, historical trend reports on congestion etc.
The total simulated time span is over 8600 seconds: the network is preloaded in the first 1800 seconds of simulation, usable data are obtained from 1801 seconds to 5400 seconds, and the remaining time is used for network clearance.

The SUMO simulation outputs a set of vehicle trajectories. Each data record contains a timestamp and the following associated vehicle information: the id, the type, the instantaneous speed and acceleration, the link ID and the relative position. As an illustration, the trajectories of all the vehicles starting from an arbitrarily selected node in the network are plotted in Figure 5.

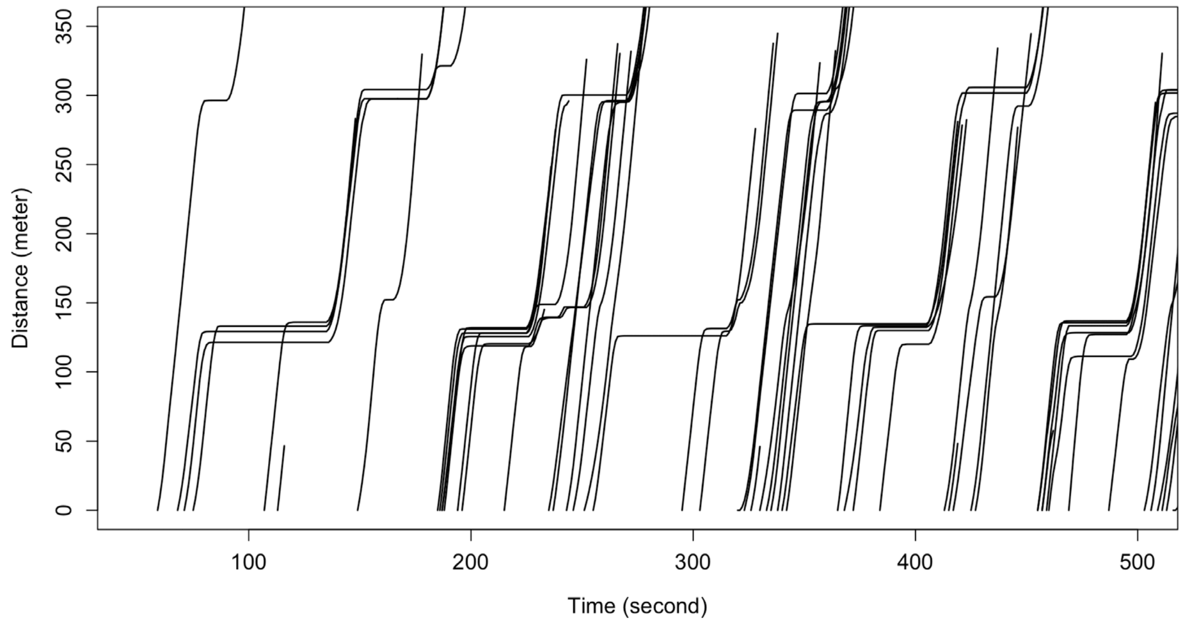


Figure 5. SUMO output of vehicle trajectories

In Figure 5, each solid line represents a vehicle trajectory, and crossing lines indicate no vehicle collisions. Crossing lines correspond to multiple lanes on the road link selected to plot the vehicle trajectories.

The microsimulation is calibrated by comparing the overall traffic speed distributions with those from real-world observations. This comparison is shown in Figure 6 below.

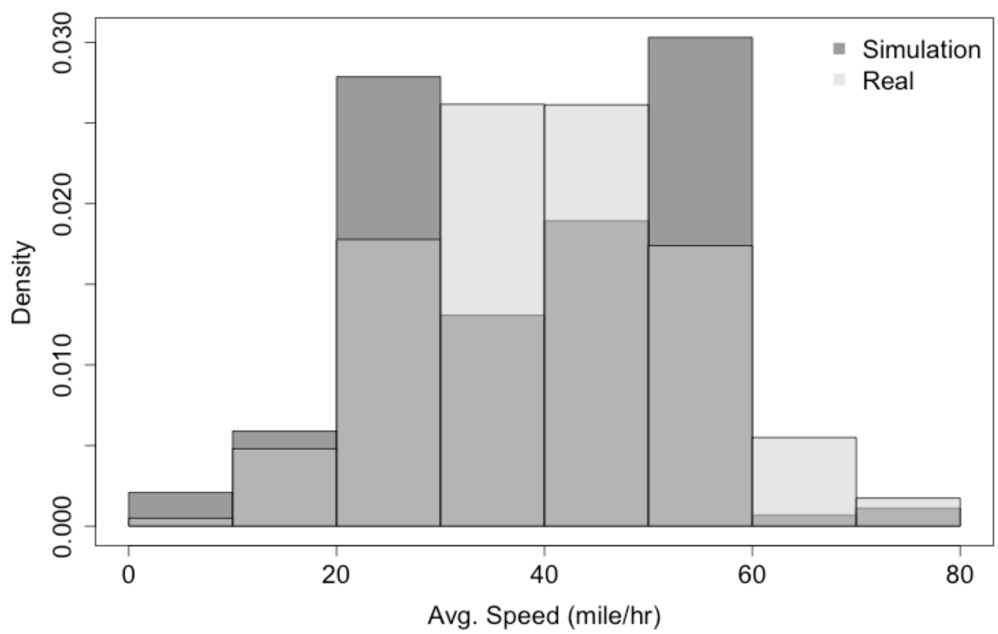


Figure 6. Comparison between Simulation Results and Real-World Observations for Speed Distributions

The simulation results for the distribution of average speeds generally match the real-world observations. Note that vehicles strictly follow the speed limit in SUMO-based simulations, which results in high frequencies for the 20 mph to 30 mph and 50 mph to 60 mph speed ranges. However, in the real world, drivers tend to drive above the speed limit, as reflected in the high frequencies of the 30 mph to 40 mph and 60 mph to 70 mph speed ranges. This discrepancy has been widely observed in microsimulations, such as those performed by Tu, Wang, and Hatzopoulou (2019).

As the iPeMS data does not contain flow information, we selected a few typical links to compare the speed-flow patterns against real-world data to validate the simulations. The purpose of this validation was to confirm that if we zoomed into individual road links, the simulation could accurately reproduce real-world traffic patterns. Figure 7 is a comparison between real-world and simulated traffic conditions for a set of selected road links. The simulations accurately reproduce the real-world traffic pattern. Note that as this study focused on estimations for peak-hour traffic, simulations for low traffic volumes were not included, however, Figure 7 shows that the projected flow-speed relation (black curve) accurately reflects real-world observations.

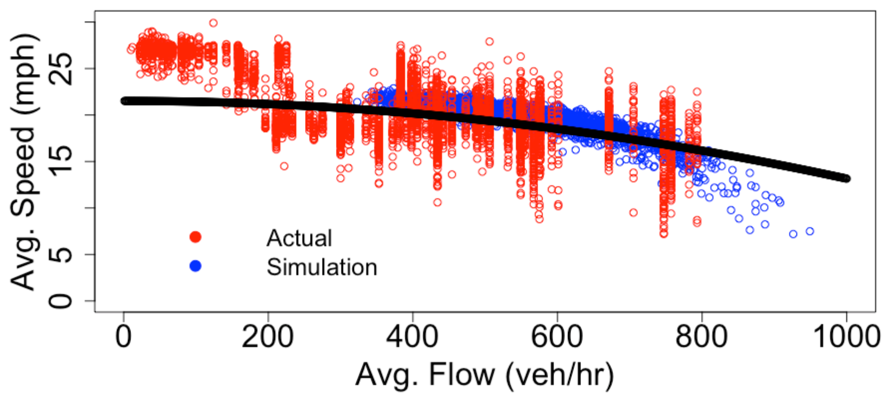


Figure 7. Comparison between the simulation results and real-world observations of detailed traffic patterns

As the raw output of SUMO simulations is a set of vehicles trajectories, including the id, speed, and position of each vehicle at any moment, we processed the data using the statistical analysis tool, R, to obtain link-based traffic condition-related information, a few of which are the average speed, the speed variance, the idling time and volume. The dataset was divided into two sub-datasets based on the road type: restricted and unrestricted urban access roads. Of these, restricted roads, such as freeways and interstates, use ramps to limit vehicle access; whereas all other urban roads are unrestricted, which typically refers to local urban streets. The traffic condition-related variables prepared at this step are listed in Table 2.

Table 2. Notation and Description of Model Variables

Variable name	Notation	Description
Average speed	v	Arithmetic mean of speeds of all vehicles going through a link during time period of interest
Avg. speed squared	v^2	Square of average speed

Variable name	Notation	Description
Avg. speed cubed	v^3	Cube of average speed
Coefficient of variation of avg. speed	c_v	Ratio of standard deviation in vehicle speeds to mean speed, which shows variability around mean speed
Percentage idling time	p_I	Vehicle idling time as percentage of total vehicle time during time period of interest
Traffic volume	f	Total number of traffic going through the link during time period of interest

These variables are chosen due to their relationship to vehicle engine demand thus energy consumption and emissions. In addition, when selecting the traffic-related variables, we focus on variables that are already archived in existing state or local traffic monitoring systems. For example, although fleet average acceleration on a link can be calculated, acceleration data are normally not reported on archived traffic system, thus it is not selected in our model. Rather, we select the coefficient of variation of speed (i.e. ratio of average and variance of speed) which can reflect changes in speed and the transient driving condition of vehicles on the road. The selected traffic-related variables will serve as the independent variables to our machine learning model to predict energy and emissions of transit buses on road network.

3.2 Energy Consumption Data Preparation using MOVES

Energy consumption data were prepared using MOVES. The EPA MOVES model is a state-of-the-science energy and emission modeling system that estimates emissions for mobile sources at the national, county, and project levels for criteria air pollutants, greenhouse gases, and air toxics. The integration of micro-simulations and MOVES has been widely utilized to estimate GHG emissions and/or air toxic emissions from on-road vehicles (Song, Yu and Zhang, 2012; Zhou et al., 2015; Abou-Senna et al., 2013; Liu et al., 2013; Tu el al., 2019).

In this study, energy consumption was measured as diesel fuel consumed (gallons). Environmental pollutant emissions data were also computed with MOVES. We selected NO_x as the indicative criteria pollutant indicator, and emission data were measured as NO_x emissions (grams).

The MOVES model estimates emissions from the distribution of the vehicle operating mode (opMode) for every link over a specified time period, which is calculated using the instantaneous speed, acceleration, and VSP of each vehicle. In MOVES, an emission rate (in grams per second) is associated with each opMode. MOVES calculates the total emissions by summing the product of the time of buses spent in each opMode and the associated emission rate of buses. At this step, we first obtain link-level opMode distribution for buses from the vehicle trajectory outputs of the traffic simulation step and use them as the inputs into MOVES model. The remaining setup of the MOVES model is given as:

- Level: Project level
- Time horizon: May 2019, weekday, morning, 8 am-9 am
- On-road vehicles: Diesel transit buses
- Road type: Urban restricted access roads for highways; urban unrestricted access roads for urban streets

- Geographical Setup:
    - Region: South Carolina, Richland County
    - Meteorological information: MOVES default database
    - Fleet age distributions: MOVES default database

The MOVES outputs (see example in Figure 8) were the total fuel consumed and the NO_x emissions for each road link. A snapshot of the MOVES output for highway fuel consumption is shown in Figure 8. The fuel consumption and NO_x emissions will be served as dependent variable in the machine learning model, which is to produce energy and emissions results as close as possible to the results generated by MOVES. In this step, MOVES energy and emissions results are treated as ground truth values, due to the facts that 1) the MOVES model has been well calibrated to represent real-world situations, and 2) the purpose of this project is to create a sketch planning tool that can perform MOVES analysis in a rapid format.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	MOVESRunI(	IterationId	YearID	MonthID	DayID	HourID	StateID	CountyID	ZoneID	LinkID	PollutantID	ProcessID	SourceTypeI(	regClassId	FuelTypeID	fuelSubtypei	ModelYearID	RoadTypeID	SCC	engTechId	sectorId	hpId	EmissionQuant	EmissionQua	EmissionQuantS
2	1	1	2019	5	5	9	45	45079	450790	1119	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	10964.09961	\N	\N
3	1	1	2019	5	5	9	45	45079	450790	1118	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	16859.09961	\N	\N
4	1	1	2019	5	5	9	45	45079	450790	1117	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	3067.159912	\N	\N
5	1	1	2019	5	5	9	45	45079	450790	1116	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	2620.439941	\N	\N
6	1	1	2019	5	5	9	45	45079	450790	1115	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	179.1459961	\N	\N
7	1	1	2019	5	5	9	45	45079	450790	1114	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	2284.800049	\N	\N
8	1	1	2019	5	5	9	45	45079	450790	1113	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	8667.209961	\N	\N
9	1	1	2019	5	5	9	45	45079	450790	1112	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	2056.550049	\N	\N
10	1	1	2019	5	5	9	45	45079	450790	1111	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	9275.290039	\N	\N
11	1	1	2019	5	5	9	45	45079	450790	1110	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	121313	\N	\N
12	1	1	2019	5	5	9	45	45079	450790	1109	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	78688.20313	\N	\N
13	1	1	2019	5	5	9	45	45079	450790	1108	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	26766.09961	\N	\N
14	1	1	2019	5	5	9	45	45079	450790	1107	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	7913.870117	\N	\N
15	1	1	2019	5	5	9	45	45079	450790	1106	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	7384.620117	\N	\N
16	1	1	2019	5	5	9	45	45079	450790	1105	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	7516.240234	\N	\N
17	1	1	2019	5	5	9	45	45079	450790	1104	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	8410.879883	\N	\N
18	1	1	2019	5	5	9	45	45079	450790	1103	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	7739.990234	\N	\N
19	1	1	2019	5	5	9	45	45079	450790	1102	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	9919.570313	\N	\N
20	1	1	2019	5	5	9	45	45079	450790	1101	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	7479.259766	\N	\N
21	1	1	2019	5	5	9	45	45079	450790	1100	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	6760.459961	\N	\N
22	1	1	2019	5	5	9	45	45079	450790	1099	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	3548.51001	\N	\N
23	1	1	2019	5	5	9	45	45079	450790	1098	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	2245.780029	\N	\N
24	1	1	2019	5	5	9	45	45079	450790	1097	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	11683.90039	\N	\N
25	1	1	2019	5	5	9	45	45079	450790	1096	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	5101.379883	\N	\N
26	1	1	2019	5	5	9	45	45079	450790	1095	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	6311.439941	\N	\N
27	1	1	2019	5	5	9	45	45079	450790	1094	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	61376.60156	\N	\N
28	1	1	2019	5	5	9	45	45079	450790	1093	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	8546.799805	\N	\N
29	1	1	2019	5	5	9	45	45079	450790	1092	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	15478.90039	\N	\N
30	1	1	2019	5	5	9	45	45079	450790	1091	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	10101.7002	\N	\N
31	1	1	2019	5	5	9	45	45079	450790	1090	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	8373.290039	\N	\N
32	1	1	2019	5	5	9	45	45079	450790	1089	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	14013.09961	\N	\N
33	1	1	2019	5	5	9	45	45079	450790	1088	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	44303.69922	\N	\N
34	1	1	2019	5	5	9	45	45079	450790	1087	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	6867.470215	\N	\N
35	1	1	2019	5	5	9	45	45079	450790	1086	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	12139	\N	\N
36	1	1	2019	5	5	9	45	45079	450790	1085	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	30955.19922	\N	\N
37	1	1	2019	5	5	9	45	45079	450790	1084	91	1	42	\N	\N	\N	\N	4	\N	\N	\N	\N	724.0650024	\N	\N

Figure 8. Sample MOVES Output

CHAPTER 4
FEATURE SELECTION, MODEL DEVELOPMENT AND RESULTS

4.1 Feature Selection

We first selected a set of variables commonly included in previous emission prediction models: these variables formed a pool for features to be selected and used in our model training. The variables are related to traffic conditions and are as follows: the average speed and its higher orders, the variance in the speed, the percentage of vehicle idling time, and the traffic volume. Note that these variables can be captured by existing traffic monitoring systems, such that the proposed model can be applied to evaluate city emissions.

Before feature selection, we tested the importance of each feature using Mallow's Cp, which is commonly used to assess the fit of a regression model.

In using Mallow's Cp to evaluate the fit of a multiple regression model that involves several predictor variables, all smaller models for a subset of all the predictors are assessed by comparison with the full model, and the unexplained error because of the exclusion of some variables is calculated. Mallow's Cp is usually calculated as follows:

$$C_p = \frac{SSE_p}{S^2} - N + 2P$$

where $SSE_p = \sum_{i=1}^{N}(Y_i - Y_{pi})^2$ is the residual sum of squares from a model with a set of P – 1 predictor variables plus an intercept (a constant); Y_{pi} is the predicted value of the ith observation of Y from the P-1 predictor variables; S^2 is the estimate of the variance σ^2, and N is the sample size.

We evaluated all the models using a subset of explanatory variables, and the results are shown in Table 3. For concision, we only include the results for the fuel consumption rate estimation model on restricted access roads. The remaining results can be found in Attachment A.

Table 3. Mallow's Cp for Fuel Consumption Rate on Restricted Roads

P-1	Variables	Cp	R-square	Adj-R-square
1	c_v	139.0633	0.1392	0.1385
1	p_I	217.5549	0.0852	0.0844
1	v	283.9363	0.0395	0.0386
1	v^2	315.6438	0.0177	0.0168
1	v^3	328.1083	0.0091	0.0082
1	f	336.4856	0.0033	0.0024
2	c_v, p_I	87.2934	0.1763	0.1748
2	v, c_v	102.3802	0.1659	0.1644
2	v^2, c_v	112.8198	0.1587	0.1572
2	v^3, c_v	124.4067	0.1507	0.1492
2	c_v, f	140.6244	0.1395	0.138
2	v^2, p_I	216.909	0.087	0.0854
2	v^3, p_I	217.7258	0.0865	0.0848
2	v, p_I	218.7467	0.0857	0.0841

2	f, p_I	219.4355	0.0853	0.0836
2	v, v^2	222.8403	0.0829	0.0813
3	v, c_v, p_I	40.4958	0.2099	0.2077
3	v^2, c_v, p_I	48.5358	0.2043	0.2022
3	v^3, c_v, p_I	61.9842	0.1951	0.1929
3	v^2, v^3, c_v	78.7539	0.1835	0.1813
3	c_v, f, p_I	87.8503	0.1773	0.175
3	v, v^3, c_v	90.2608	0.1756	0.1734
3	v, v^2, c_v	96.1493	0.1715	0.1693
3	v, c_v, f	97.0178	0.1709	0.1687
3	v^2, c_v, f	106.9173	0.1641	0.1619
3	v^3, c_v, f	121.1534	0.1543	0.152
4	v^2, v^3, c_v, p_I	20.8357	0.2248	0.222
4	v, v^3, c_v, p_I	35.3384	0.2148	0.212
4	v, c_v, f, p_I	36.2979	0.2141	0.2113
4	v, v^2, c_v, p_I	40.0564	0.2115	0.2087
4	v^2, c_v, f, p_I	42.3904	0.2099	0.2071
4	v^3, c_v, f, p_I	57.8962	0.1993	0.1964
4	v, v^2, v^3, c_v	73.1328	0.1888	0.1858
4	v^2, v^3, c_v, f	80.2586	0.1839	0.1809
4	v, v^3, c_v, f	90.975	0.1765	0.1735
4	v, v^2, c_v, f	95.3513	0.1735	0.1705
5	v, v^2, v^3, c_v, p_I	5.2107	0.2369	0.2335
5	v^2, v^3, c_v, f, p_I	21.9209	0.2254	0.2219
5	v, v^3, c_v, f, p_I	35.2893	0.2162	0.2127
5	v, v^2, c_v, f, p_I	37.965	0.2144	0.2108
5	$v, v2, v3, cv, f$	74.9984	0.1889	0.1852
5	v, v^2, v^3, f, p_I	126.9214	0.1531	0.1493
6	v, v^2, v^3, c_v, f, p_I	7	0.2371	0.2329

The general rule for using Mallow's Cp is to select a model for which $C_p \leq P$ and is close to P. $C_p \geq P$ indicates that the partial model is under fitted compared with the full model, that is, the model does not have adequate explanatory capabilities compared to the full model. However, $C_p \ll P$ indicates overfitting, which is also reflected in a low Adj-R squared value. An overfitted model cannot be used to make robust predictions with new data.

The table shows that the best-fit model has 5 (out of 6) variables, including $v, v^2, v^3, c_v,$ and p_I, $C_p = 5.2107$ ($P = 6$) and an adjusted R squared of 0.2335. However, the adjusted R squared is not significantly lower than that of the full model, which suggests that all 6 predictor variables should be included to provide as much information as possible.

Then, with the information obtained from Mallow's Cp approach, we adopted the random forest algorithm to determine the optimal set of features for a neural network model. This algorithm tests different combinations of features and neural network specifications to determine an optimal set of features for constructing a model. The results showed that all the features were important for the model, in agreement with the conclusions obtained from the Mallow's Cp process.

4.2 Model Development

We used the selected features, that is, the average speed, the square and cube of the average speed, the variational coefficient of the average speed, the percentage idling time and the traffic volume, to train the model with the MOVES output to serve as the ground truth for a supervised neural network training. Figure 9 shows the traditional fully connected structure adopted for the deep-learning neural network.

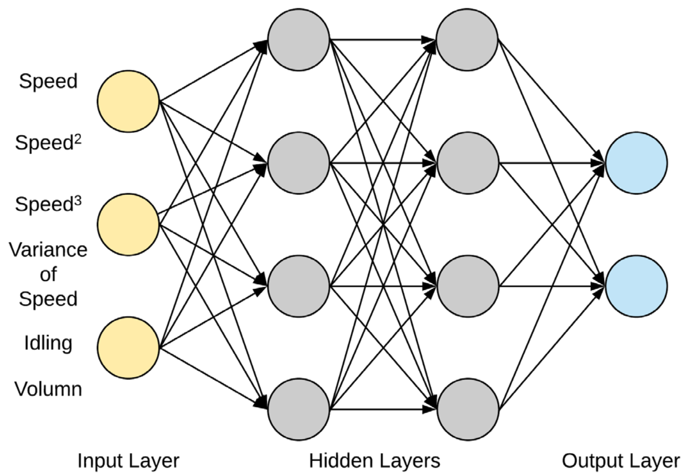


Figure 9. Model Structure

We reserved 20% of the model data for model testing and used the remaining 80% for model training.
We tested many different network specifications to identify the network with the minimum squared error.

- Restricted access road
    - Energy consumption rate estimation
      The model has a total of 5 layers: the input layer has 6 neurons; the second, third, and fourth layers have 8, 6, and 4 neurons, respectively; and the last layer, which is the output layer, has 1 neuron.
    - Pollution estimation (NO_x)
      The model has a total of 5 layers, with the same distribution of neurons given above.

- Unrestricted access road
    - Energy consumption rate estimation
      The model has a total of 5 layers: the input layer has 6 neurons; the second, third, and fourth layers have 8, 5, and 2 neurons, respectively; and the last layer, which is the output layer, has 1 neuron.
      Pollution estimation (NO_x)
      The model has a total of 5 layers, with the same distribution of neurons given above.

4.3 Model Results

We conducted simulations to evaluate the link-based energy consumption rate and emission rate for transit buses during peak-hour traffic. Figure 10 shows a Columbia city map, where the colored links indicate the levels of the energy consumption rate. Green indicates a low energy consumption rate (below 0.15 gallons/mile for the fuel consumption rate), blue indicates a medium rate (0.15 gallons/mile to 0.225 gallons/mile), and red indicates a high rate (above 0.225 gallons/mile).

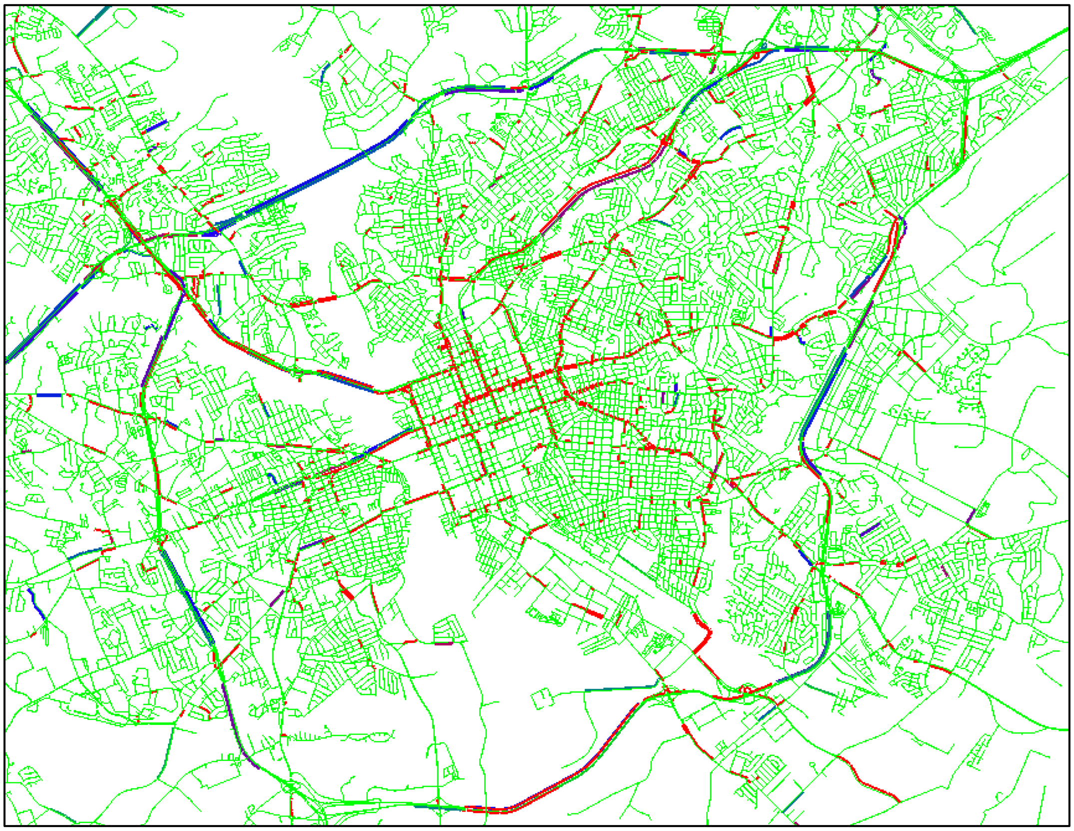
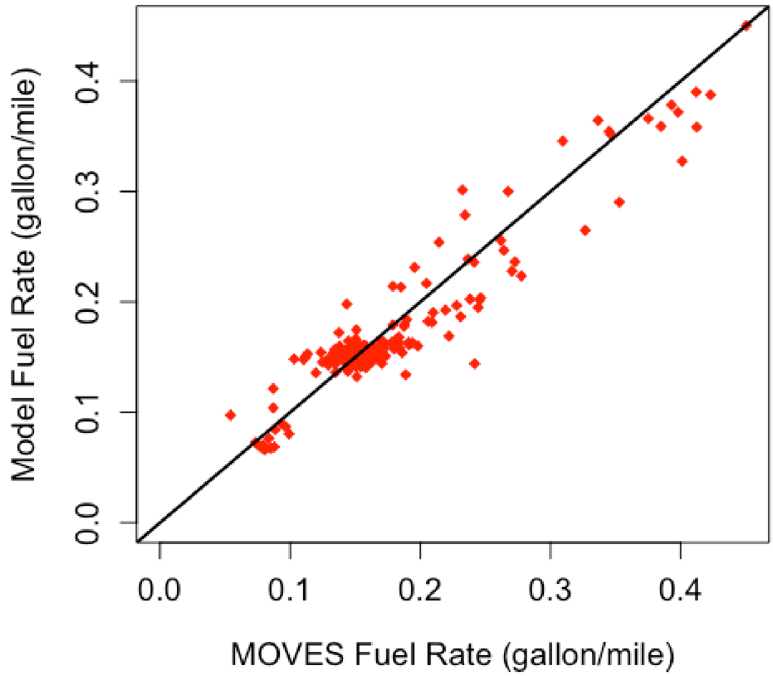


Figure 10. Fuel Consumption Rate Map for City of Columbia

After completing the neural network model training, we tested the model with a preserved dataset for testing, and the results are shown in Figures 11, 12, 13 and 14.



Error = 9.7%

Figure 11. Model Prediction vs. MOVES Output for Fuel Consumption Rate on Restricted Access Roads

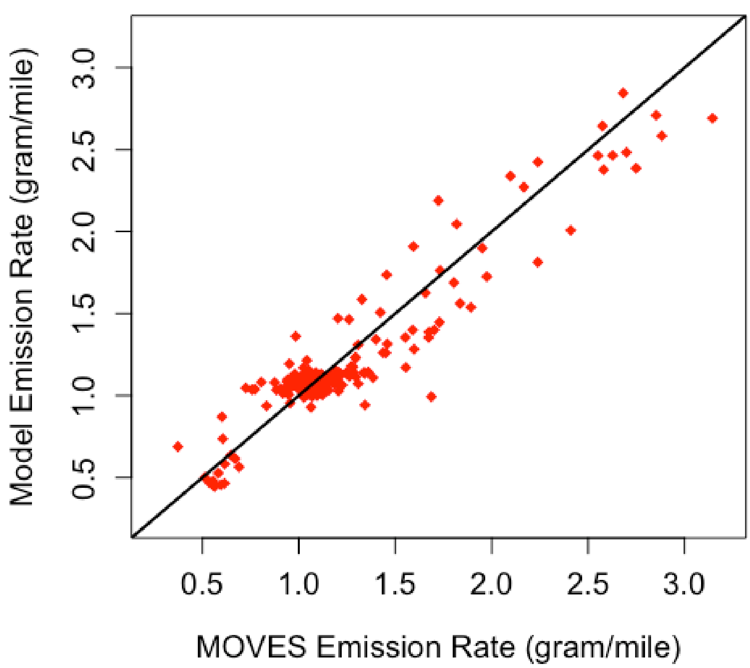


Figure 12. Model Prediction vs. MOVES Output for NO_x Emission Rate on Restricted Access Roads

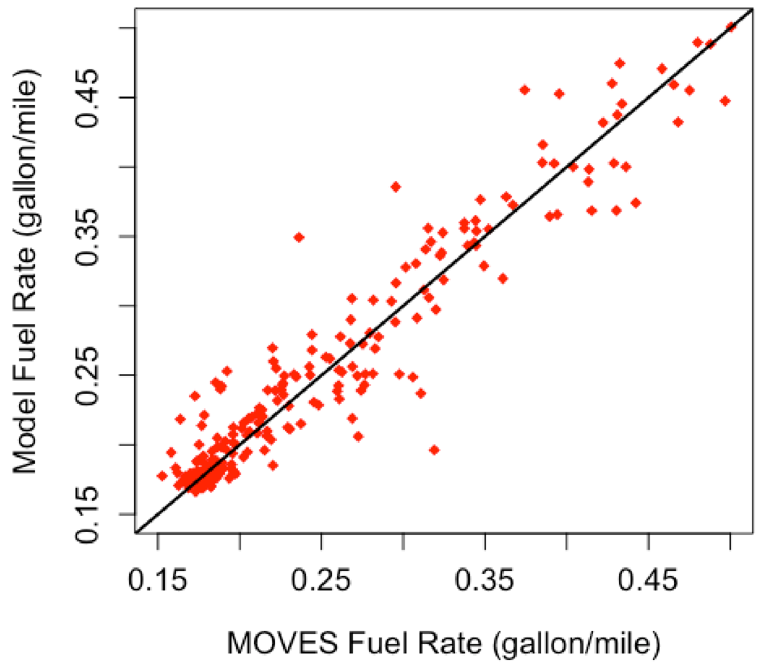


Figure 13. Model Prediction vs. MOVES Output for Fuel Consumption Rate on Unrestricted Access Roads

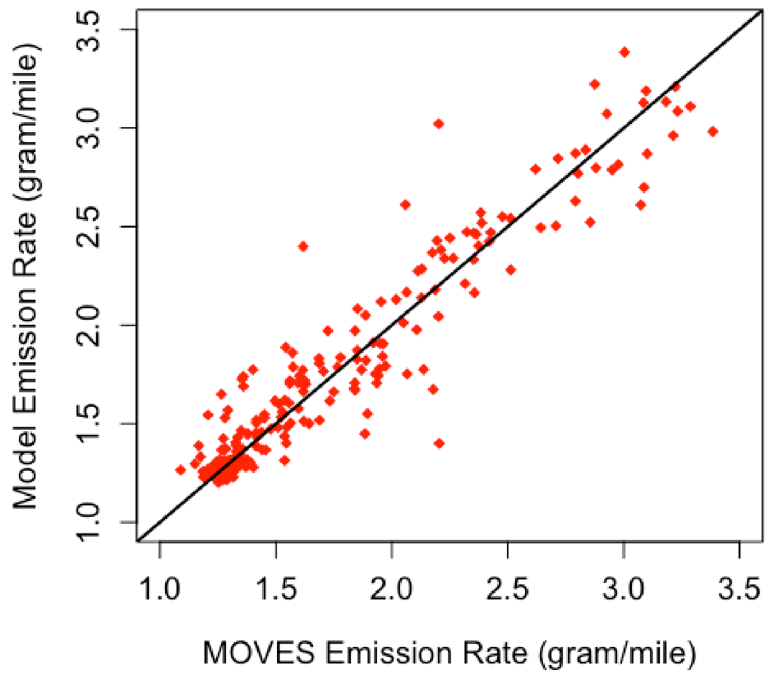


Figure 14. Model Prediction vs. MOVES Output for NO_x Emission Rate on Unrestricted Access Roads

The MAPE of the predictions is summarized in Table 4.

Table 4. Model Performance Matrix

Road type	**Energy**	**NO_x**
Urban restricted access	6.8%	6.6%
Urban unrestricted access	9.7%	9.9%

CHAPTER 5
SPATIAL TRANSFERABILITY TEST AND DISCUSSION

5.1 Spatial Transferability Test

We tested the applicability of the model to cities other than Columbia, SC by performing simulations and a MOVES run for another medium-sized city, Eichstätt in Germany. We then applied our trained model and compared the output to that of the MOVES model. The flowchart in Figure 15 summarizes the content of the preceding and current chapters to illustrate the project structure.

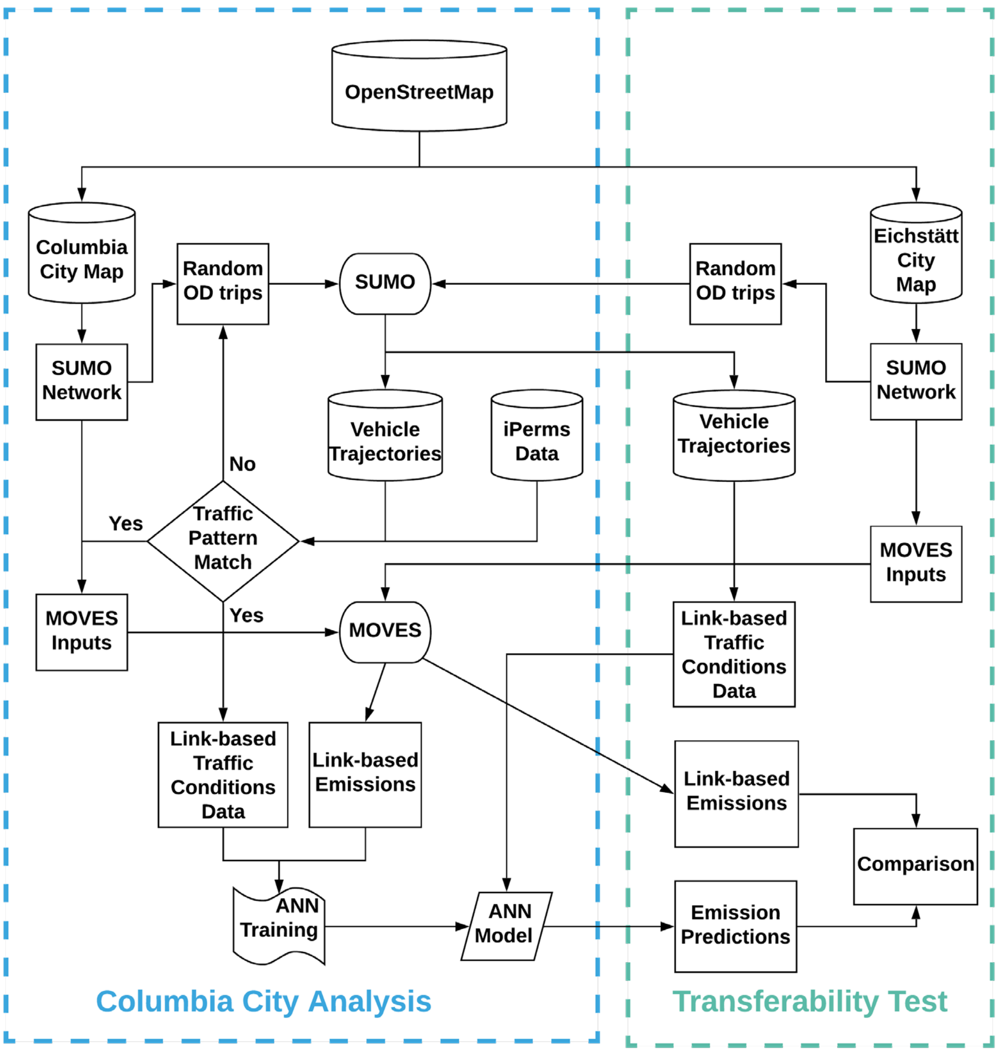


Figure 15. Flowchart for the Project

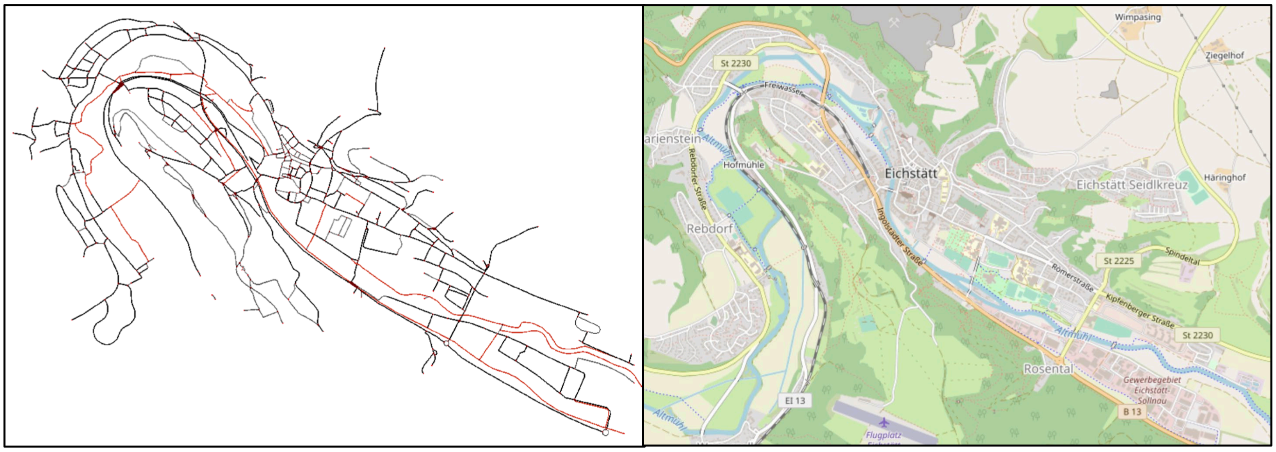


Figure 16. Eichstätt SUMO network (Left) and Openstreetmap Counterpart (Right)

The Eichstätt road network is shown in Figure 16. We conducted microsimulations and MOVES modeling as in Chapter 3. The microsimulation outputs were further processed and fed into the model trained in Chapter 4, and the model output was compared with that of MOVES. The results are shown in Figures 17 and 18 below.

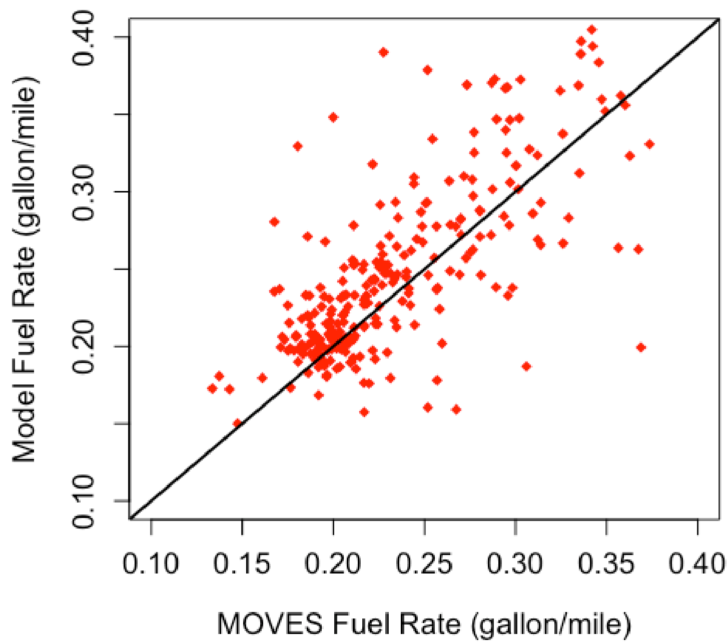


Figure 17. Model Prediction vs. MOVES Output for Fuel Consumption Rate in City of Eichstätt

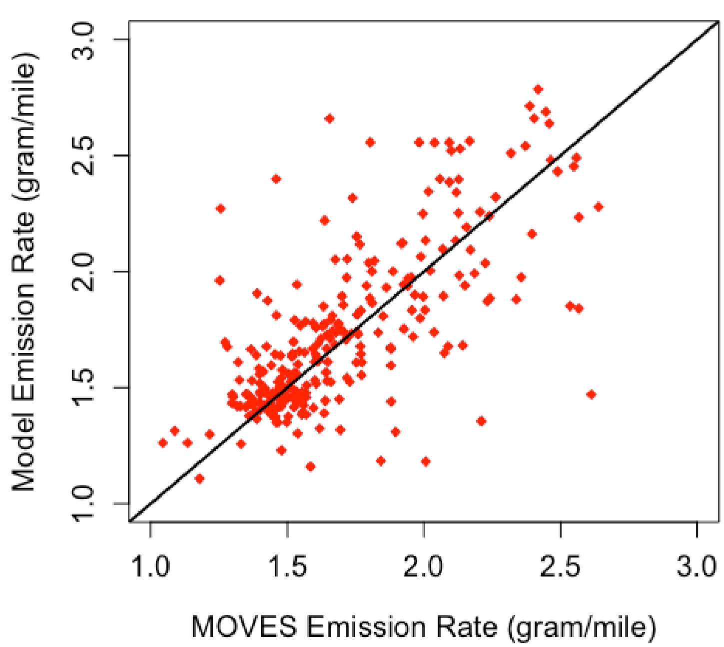


Figure 18. Model Prediction vs. MOVES Output for NO_x Rate in City of Eichstätt

The MAPE is 12.3% for the predicted overall fuel consumption and 10.2% for NO_x. The results show that the deep-learning model trained in Chapter 4 accurately predicts the emission rates using simulation data from a completely different city. This result suggests that the model has strong applicability to other urban cities and thus can be used for energy consumption and emission evaluations for many different cities across the U.S.

5.2 Discussion

Accurate predictions were obtained by applying the proposed model to a completely different urban city outside the region for which the model was originally trained, suggesting the model has a strong spatial transferability and can be applied to other U.S. cities for planning or evaluation purposes. However, although high model accuracy is obtained using MAPE as the performance measure, the Model Prediction vs. MOVES Output plot for the emission rate shows some deviations between the model predictions and the ground truth data, indicating the model is not fully capable of accurately predicting emission rates of road links for other urban cities. These deviations mainly occurred at some high emission rate data points. The model overestimated the emission rate at these points.

One highly likely reason for the overestimation by the model is the absence of critical features of traffic control-related variables, such as signal phase patterns. Another likely reason is that some traffic patterns were not captured by the simulations that were used to generate the traffic condition data for the model training. In this case, the model cannot make accurate predictions under the missing traffic conditions.

We also speculate that the deviation may result from differences in road topologies, such as the percentage of signal-controlled road links or highways, positions of highways, distributions of link speed limits, types of link connections, among many others. Differences in road topologies can lead to drastically different traffic activities, and these differences may not have been sufficiently captured by the traffic condition variables we selected.

In addition, the model comprises two sub-models: one for urban highways and one for urban local streets. We speculate that refining the road type could improve the model capability.

CHAPTER 6
CONCLUSION

Sustainable design and environmental protection are becoming the top priorities in urban transit planning. One barrier to realizing these goals is the absence of efficient tools to evaluate the energy and emissions impacts of transit activities under various traffic conditions. The state-of-the-art tool, MOVES, has low accuracy when using aggregated and macroscopic level data, whereas microscopic models are computationally intensive in terms of calibration and generating results, especially for large-scale urban networks.

In this project, a deep-learning-based surrogate model is developed that can efficiently and accurately predict vehicle-related energy consumptions and emissions for urban transit buses under different traffic patterns. This model can serve as a sketch planning tool for public transportation planners to design fuel-sustainable and pollution-mitigating transit routes, as well as an evaluation tool for researchers and transportation air quality practitioners to quickly assess bounds on emissions benefits of traffic operational strategies.

In machine learning, data quality is essential for a successful model. To prevent the "garbage in, garbage out" scenario, the proposed model was developed by meticulously calibrating the network for the city of Columbia, and extensive microsimulations were also carefully calibrated using real-world observed traffic data to generate an abundant quantity of good-quality data for machine learning training. MOVES modeling was then executed to acquire emission rate data for use as the ground truth in the training and testing process.

In a knowledge-driven approach, a set of traffic condition-related variables are initially identified as candidate prediction features, which include the average speed, the square and cube of the average speed, the coefficient of variation of the average speed, the idling time as a percentage of the total vehicle time and the traffic volume. The model features were selected using both a traditional statistical technique (Mallow's Cp method) and a machine learning algorithm (the random forest). The results show that all of the aforementioned traffic condition-related variables are important. Note that all these variables can be measured by the majority of existing traffic monitoring equipment on the market.

The calibrated simulations were used to evaluate the link-based transit bus energy consumption rate and pollutant emission rate of the city of Columbia under peak-hour traffic conditions. The model can accurately predict the energy consumption rate and pollutant emission rate: the MAPEs for the estimated fuel consumption rate and the NO_x emission rate are 6.8% and 6.6%, respectively, for highways, and 9.7% and 9.9%, respectively, for urban roads.

The transferability of the model to other urban cities was tested by conducting microsimulations for a completely different medium-sized city in Germany, followed by MOVES modeling to obtain corresponding emission rate data. The deep-learning model was applied to the microsimulation data, and the predicted emission rates were compared to the MOVES results. The application of the model to predict emissions rates of road links on the Germany city produces the MAPE of 12.3% and 10.2% for fuel consumption rate and the NO_x rate, respectively.

This project focuses on urban road networks. Further studies are required to explore the development of similar models for rural areas across the U.S. The road grade was not included as a predictive feature in this project because of the absence of corresponding data. However, the road grade plays a critical role in automobile fuel consumption rate and is therefore correlated with the emission rate. This suggests further studies using road grade data to establish a more comprehensive evaluation model.

REFERENCES

Abou-Senna, H., Radwan, E., Westerlund, K. and Cooper, C.D., 2013. Using a traffic simulation model (VISSIM) with an emissions model (MOVES) to predict emissions from vehicles on a limited-access highway. Journal of the Air & Waste Management Association, 63(7), pp.819-831.

Beaudoin, J., Farzin, Y.H. and Lawell, C.Y.C.L., 2015. Public transit investment and sustainable transportation: A review of studies of transit's impact on traffic congestion and air quality. Research in Transportation Economics, 52, pp.15-22.

Brooker, A., Gonder, J., Wang, L., Wood, E. et al., "FASTSim: A Model to Estimate Vehicle Efficiency, Cost, and Performance," SAE Technical Paper 2015-01-0973, 2015, doi:10.4271/2015-01-0973

Bureau of Transportation Statistics, 2018. Transportation Economic Trends 2018. Washington, D.C.: United States Department of Transportation. https://doi.org/10.21949/1502599.

Carrese, S., Gemma, A. and La Spada, S., 2013. Impacts of driving behaviours, slope and vehicle load factor on bus fuel consumption and emissions: a real case study in the city of Rome. Procedia-Social and Behavioral Sciences, 87, pp.211-221.

Delgado, O.F., Clark, N.N. and Thompson, G.J., 2011. Modeling transit bus fuel consumption on the basis of cycle properties. Journal of the Air & Waste Management Association, 61(4), pp.443-452.

Department of Energy, 2019. Transit Buses by Fuel Type in United States. Accessed December 18, 2019. https://afdc.energy.gov/data/

e Silva, J.D.A., Moura, F., Garcia, B. and Vargas, R., 2015. Influential vectors in fuel consumption by an urban bus operator: Bus route, driver behavior or vehicle type?. Transportation Research Part D: Transport and Environment, 38, pp.94-104.

E.H. Pechan & Associates, Inc. ADVANCES IN PROJECT LEVEL ANALYSES FINAL REPORT, 2010 (Report No. 10.10.001/9005.000). Retrieved from the U.S. Department of Transportation, Federal Highway Administration website: https://w ww.fhwa.dot.gov/ENVIRONMENT/air_quality/conformity/research/project_level_analyses/pla.pdf

EPA, U., 2015. MOVES2014a User Guide. EPA-420-B-15-095.

Frey, H.C., 2018. Trends in on road transportation energy and emissions. Journal of the Air & Waste Management Association, 68(6), pp.514-563.

Frey, H.C., Rouphail, N.M., Zhai, H., Farias, T.L. and Gonçalves, G.A., 2007. Comparing real-world fuel consumption for diesel-and hydrogen-fueled transit buses and implication for emissions. Transportation Research Part D: Transport and Environment, 12(4), pp.281-291.

Giraldo, M. and Huertas, J.I., 2019. Real emissions, driving patterns and fuel consumption of in-use diesel buses operating at high altitude. Transportation Research Part D: Transport and Environment, 77, pp.21-36.

Guensler, R., Liu, H., Xu, X., Xu, Y. and Rodgers, M.O., 2016, January. MOVES-Matrix: setup, implementation, and application. In 95th Annual Meeting of the Transportation Research Board, Washington, DC.
Vancouver

He, X., Zhang, S., Ke, W., Zheng, Y., Zhou, B., Liang, X. and Wu, Y., 2018. Energy consumption and well-to-wheels air pollutant emissions of battery electric buses under complex operating conditions and implications on fleet electrification. Journal of cleaner production, 171, pp.714-722.

Hung, W.T., Tong, H.Y. and Cheung, C.S., 2005. A modal approach to vehicular emissions and fuel consumption model development. Journal of the Air & Waste Management Association, 55(10), pp.1431-1440.

Iteris, 2020. South Carolina Department of transportation completes upgrade to Iteris ClearGuide for improved statewide mobility. (n.d.). Iteris, Inc. https://www.iteris.com/news/south-carolina-department-transportation-completes-upgrade-iteris-clearguide-improved

Krzyżanowski, M., Kuna-Dibbert, B. and Schneider, J. eds., 2005. Health effects of transport-related air pollution. WHO Regional Office Europe.

Lajunen, A., 2014. Energy consumption and cost-benefit analysis of hybrid and electric city buses. Transportation Research Part C: Emerging Technologies, 38, pp.1-15.

Liu, H., Wang, Y., Chen, X. and Han, S., 2013. Vehicle emission and near-road air quality modeling in Shanghai, China, based on taxi GPS data and MOVES revised emission inventory. Transp. Res. Rec, 2340, pp.38-48.

López-Martínez, J.M., Jiménez, F., Páez-Ayuso, F.J., Flores-Holgado, M.N., Arenas, A.N., Arenas-Ramirez, B. and Aparicio-Izquierdo, F., 2017. Modelling the fuel consumption and pollutant emissions of the urban bus fleet of the city of Madrid. Transportation Research Part D: Transport and Environment, 52, pp.112-127.

Ma, H., Xie, H., Huang, D. and Xiong, S., 2015. Effects of driving style on the fuel consumption of city buses under different road conditions and vehicle masses. Transportation Research Part D: Transport and Environment, 41, pp.205-216.

Mata, C., de Oliveira Leite, W., Moreno, R., Agudelo, J.R. and Armas, O., 2016. Prediction of NO_x emissions and fuel consumption of a city bus under real operating conditions by means of biharmonic maps. Journal of Energy Engineering, 142(4), p.04016018.

Pan, Y., Chen, S., Qiao, F., Ukkusuri, S.V. and Tang, K., 2019. Estimation of real-driving emissions for buses fueled with liquefied natural gas based on gradient boosted regression trees. Science of the Total Environment, 660, pp.741-750.

Rakha, H., Yue, H. and Dion, F., 2011. VT-Meso model framework for estimating hot-stabilized light-duty vehicle fuel consumption and emission rates. Canadian Journal of Civil Engineering, 38(11), pp.1274-1286.

Scora, G. and Barth, M., 2006. Comprehensive modal emissions model (cmem), version 3.01. User guide. Centre for environmental research and technology. University of California, Riverside, 1070

Shen, Y., Zhang, H. and Zhao, J., 2018. Integrating shared autonomous vehicle in public transportation system: A supply-side simulation of the first-mile service in Singapore. Transportation Research Part A: Policy and Practice, 113, pp.125-136.

Song, G., Yu, L. and Zhang, Y., 2012. Applicability of traffic microsimulation models in vehicle emissions estimates: Case study of VISSIM. Transportation research record, 2270(1), pp.132-141.

Sbayti, H., El-Fadel, M. and Kaysi, I., 2002. Effect of roadway network aggregation levels on modeling of traffic-induced emission inventories in Beirut. Transportation Research Part D: Transport and Environment, 7(3), pp.163-173.

Tang, Xiru, Zhu, Y., Xu, L., 2016. The Analysis of Space-time Characteristics of Bus Operation and Energy Consumption based on ArcGIS. Energy Procedia, 104, pp.456-461.

Wang, A., Tu, R., Gai, Y., Pereira, L.G., Posen, I.D. and Hatzopoulou, M., 2019. Capturing the uncertainties in regional emission estimates related to vehicle electrification can improve the robustness of decision-making (No. 19-04153).

Wang, C., Ye, Z., Yu, Y. and Gong, W., 2018. Estimation of bus emission models for different fuel types of buses under real conditions. Science of The Total Environment, 640, pp.965-972.

Wang, J. and Rakha, H.A., 2016. Fuel consumption model for conventional diesel buses. Applied energy, 170, pp.394-402.

Wang, J. and Rakha, H.A., 2016. Modeling fuel consumption of hybrid electric buses: model development and comparison with conventional buses. Transportation Research Record, 2539(1), pp.94-102.

Wang, J. and Rakha, H.A., 2017. Convex fuel consumption model for diesel and hybrid buses. Transportation Research Record, 2647(1), pp.50-60.

Wang, Y., Szeto, W.Y., Han, K. and Friesz, T.L., 2018. Dynamic traffic assignment: A review of the methodological advances for environmentally sustainable road transportation applications. Transportation Research Part B: Methodological, 111, pp.370-394.

White, P.R., 2016. Public transport: its planning, management and operation. Routledge.

Wu, X., Zhang, S., Wu, Y., Li, Z., Zhou, Y., Fu, L. and Hao, J., 2015. Real-world emissions and fuel consumption of diesel buses and trucks in Macao: from on-road measurement to policy implications. Atmospheric Environment, 120, pp.393-403.

Wu, Y., Yu, L., Song, G. and Xu, L., 2013. Feasibility study of fuel consumption prediction model by integrating vehicle-specific power and controller area network bus technology. Transportation Research Record, 2341(1), pp.66-75.

Xie, Y., Chowdhury, M., Bhavsar, P., & Zhou, Y. (2012). An integrated modeling approach for facilitating emission estimations of alternative fueled vehicles. Transportation Research Part D: Transport and Environment, 17(1), 15-20.

Xu, Y., Li, H., Liu, H., Rodgers, M.O. and Guensler, R.L., 2017. Eco-driving for transit: An effective strategy to conserve fuel and emissions. Applied energy, 194, pp.784-797.

Yao, Z., Wei, H., Liu, H. and Li, Z., 2013. Statistical vehicle specific power profiling for urban freeways. Procedia-Social and Behavioral Sciences, 96, pp.2927-2938.

Zeng, W., Miwa, T. and Morikawa, T., 2015. Exploring trip fuel consumption by machine learning from GPS and CAN bus data. Journal of the Eastern Asia Society for Transportation Studies, 11, pp.906-921.

Yu, H., Li, M., Li, J., Liu, Y., Lv, H. and Ma, K., 2019. Real-road NO_x Emission and Fuel Consumption Characteristics of China IV Public Transit Buses. Energy Procedia, 158, pp.4623-4628.

Yu, Q., Li, T. and Li, H., 2016. Improving urban bus emission and fuel consumption modeling by incorporating passenger load factor for real world driving. Applied energy, 161, pp.101-111.

Zhang, S., Wu, Y., Liu, H., Huang, R., Yang, L., Li, Z., Fu, L. and Hao, J., 2014. Real-world fuel consumption and CO_2 emissions of urban public buses in Beijing. Applied Energy, 113, pp.1645-1655.

Zhou, X., Tanvir, S., Lei, H., Taylor, J., Liu, B., Rouphail, N.M. and Frey, H.C., 2015. Integrating a simplified emission estimation model and mesoscopic dynamic traffic simulator to efficiently evaluate emission impacts of traffic management strategies. Transportation Research Part D: Transport and Environment, 37, pp.123