BIG DATA ANALYTICS FOR BATTERY ELECTRIC BUS

ENERGY MODELLING AND PREDICTION

BIG DATA ANALYTICS FOR BATTERY ELECTRIC BUS ENERGY

MODELLING AND PREDICTION

By

HATEM MAHMOUD AHMED ABDELATY

B.Sc., M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the

Requirements for the Degree Doctor of Philosophy

DOCTOR OF PHILOSOPHY (2021)

McMaster University

Hamilton, Ontario

(Civil Engineering)

TITLE:	Big Data Analytics for Battery Electric Bus Energy Modelling and Prediction
AUTHOR:	Hatem Abdelaty
SUPERVISOR:	Dr. Moataz Mohamed (Assistant Professor)
NUMBER OF PAGES:	xxi, 262

ABSTRACT

Battery electric buses (BEBs) bring several advantages to public transportation systems. With fixed routes and scheduled trips, the implementation of BEBs in the transit context is considered a seamless transition towards a zero greenhouse gases transit system. However, energy consumption uncertainty is a significant deterrent for mainstream implementation of BEBs. Demonstration and trial projects are often conducted to better understand the uncertainty in energy consumption (E_C). However, the BEB's energy consumption varies due to uncertainty in operational, topological, and environmental attributes.

This thesis aims at developing simulation, data-driven, and low-resolution models using big data to quantify the E_C of BEBs, with the overarching goal of developing a comprehensive planning framework for BEB implementation in bus transit networks. This aim is achieved through four interwind objectives.

1) Quantify the operational and topological characteristics of bus transit networks using complex network theory. This objective provides a fundamental base to understanding the behaviour of bus transit networks under disruptive events.

2) Investigate the impacts of the vehicular, operational, topological, and external parameters on the E_C of BEBs.

3) Develop and evaluate the feasibility of big-data analytics and data-driven models to numerically estimate BEB's E_C.

4) Create an open-source low-resolution data-based framework to estimate the E_C of BEBs. This framework integrates the modelling efforts in objectives 1-3 and offers practical knowledge for transit providers.

Overall, the thesis provides genuine contributions to BEB research and offers a practical framework for addressing the E_C uncertainty associated with BEB operation in the transit context. Further, the results offer transit planners the means to set up the optimum transit operations profile that improves BEB energy utilization, and in turn, reduces transit-related greenhouse gases.

ACKNOWLEDGEMENTS

In the name of Allah, the most gracious, the most merciful

First, I would like to express my sincere thanks to my advisor Dr. Moataz Mohamed for his excellent direction, support, and encouragement during the entire work. I am very grateful for the opportunity to develop my work at McMaster University and in this knowledge domain under his supervision, as well as to learn from his experience. I have to admit his superior patience with students and with me in unmatched. He continuously keeps an eye on every detail related to the work and strongly helps and supports his students in solving any concerns regarding the work and personal life. Wishing Dr. Moataz and his family all the best in their life, May Allah bless you with prosperity, success and achieve every goal in your life.

I would also like to thank my committee members: Dr. Wael El-Dakhakhni and Dr. Mohamed Hussein, for their observations and time. Their monitoring of my work has motivated me to improve and to continue my training.

I am very grateful to the co-authors who have shared the work of my papers and whose contributions have made their publication possible.

Finally, heartfelt thanks to my family and friends in Canada and Egypt. I thank them for all their understanding, patience, and love during these four years. Thanks to my parents, who supported me during these times and still support me during my life, I cannot achieve anything in my life without your loyal prayers. Thanks to my lovely sisters, Ola and Doha, and thanks to my lovely brothers Mohamed and Moaz.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

a_{ij}	Existence of a connection between nodes i and j (1 if it exists or 0 if it does not)
a_{jk}	Existence of a connection between nodes j and k (1 if it exists or 0 if it does not)
a_{ik}	Existence of a connection between nodes i and k (1 if it exists or 0 if it does not)
C	Travel time cost incurred when all links are existing in the network
C_a	Travel time cost of removing a link
$C_B(i)$	Betweenness Centrality of node (i)
$C_c(i)$	Closeness Centrality of node (i)
C_i	Clustering Coefficient of node (i)
Cn_{ij}	Connectivity between city i and city j
Cn_{ij}^{Air}	Dynamic weighed Connectivity between city i and city j by air transportation
Cn_{ij}^{Train}	Dynamic weighed Connectivity between city i and city j by rail transportation
C/C_R	Ratio between weighted Clustering Coefficient and its mean randomization value
D_{ij}^V	Ratio between the route and global highest average speed
d_{ij}	Shortest Path between node/station (i) and node/station (j)
$E(G)$	Efficiency of network G
$f(APL)$	Maximum value for Average Path Length after removing nodes in the network
e^m	Total number of multiple links
H_N	Total number of hub stations in the bus transit network
H_{ij}	Number of links in the shortest path between station (i) and station (j)
k_i	Degree of a node/station (i) (total number of links connected to a node/station (i))
K_{av}	Average Degree of the bus transit networks
K_{av}^2	Second moment Degree of the bus transit networks
K_{max}	Maximum number of links connected to a single bus station
K_T	Summation of the Degree for each bus station in the network
$K(G-S)$	Number of connected components split after eliminating nodes
$L(G)$	Average Shortest Path Length
L_T	Total number of links in the network
L_m	Total number of multiple links between bus stations
L/L_R	Ratio between the inverse weight of path length and its mean randomization value
N	Number of nodes/stations in the network
N_{Total}	Total number of bus stations in the bus transit network.
N_{TE}	Total number of transfer and end stations
r_{ij}	Length of the scaled path selected by passenger intuitive logic algorithm
S	Total number of eliminated nodes
v_c^t	Total number of subway lines passing through a transfer station minus one
v^t	Total number of transfer stations
W_{ij}^{Air}	Link weight between node/city(i) and node/city(j) related with air transportation
W_{ij}^{Train}	Link weight between node/city(i) and node/city(j) related with rail transportation
L_i	Total number of links between K_i neighbours

λ_i	Eigenvalues of the adjacency matrix of a graph.
$\sigma_{ij}(i)$	Total number of shortest paths that pass-through node/station (i)
σ_{kj}	Total number of shortest paths between each two nodes/stations in the network
$\rho_{uw}(i)$	Probability that a path from nodes (u) to (w) passes through a node (i)
ε	Length difference between two routes
γ	Transfer count difference
ξ	Route transfer count divergence threshold
λ	Route length divergence threshold
l_i	Load of a node (i)
α	Tolerance parameter
ρ	Connectivity of the network
e-Buses	Electric buses
EVs	Electric vehicles
BEB	Battery electric bus
FCEB	Fuel cell electric bus
GHG	Green house gas
E_C	Energy consumption
m	Bus mass
A_F	Frontal area of the bus
HP	Motor power
C_d	Drag coefficient
C_r	Rolling resistance coefficient
C_B	Battery capacity
T_B	Battery temperature
SoC	State of charge
SoC_i	Initial state of charge
T_C	Traffic condition
T_T	Travel time
S_S	Spacing between stops
S_D	Stop Density
P_L	Passenger loading
V_a	Average driving speed
a	Acceleration rates
d	Deceleration rates
g	Road gradient
L	Route length
R_C	Road condition
T_A	Ambient temperature
p_a	Air density
V_W	Wind speed
D_{Agg}	Driver aggressiveness
HVAC	Heating, ventilation, and air conditioning
AUX	Auxiliary power
Br_{reg}	Regenerative brake

HSR	Hamilton street railway
MRA	Multiple regression analysis
IM	Interpolation method
ML	Machine learning
NN	Neural network
SA	Sensitivity analysis
MLR	Multiple linear regression
DT	Decision tree
GDBT	Gradient boosting decision tree
SVM	Support victor machine learning
RB-NN	Radial basis neural network
MLP-NN	Multilayer perception neural network
RBF	Radial basis function interpolation
R^2	Coefficient of determination
MSE	Mean square error
RMSE	Root mean square error
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MAD	Mean absolute deviation
GPS	Global positioning system
AVL	Automatic vehicle location
APC	Automatic passenger count

CHAPTER ONE

1. INTRODUCTION

Over the last two decades, the transportation sector has accounted for a significant share of Greenhouse Gas (GHG) emissions. Currently, it accounts for 16.2% of worldwide GHG emissions (Ritchie and Roser, 2020) and around 25% of GHG emissions in Canada (Environment-Canada, 2021). On the other hand, the electrification of transportation networks is a superior option for reducing transportation's carbon footprint. Consequently, it became one of the most crucial topics in transportation research and for transportation agencies.

In particular, bus transit systems are suitable for implementing and operating electric powertrain technology due to the nature of fixed routes and timely operation. Using electric Bus (e-Bus) transit systems in our modern communities has gained significant importance in the context of energy-saving and emission-reduction.

e-Buses could be classified into three main classes: Overhead Trolley Electric Buses, which supply the electric motor with the energy continuously using an external power resource such as trolleybuses which works by overhead wires. Fuel Cell Electric Buses (FCEBs), which use hydrogen fuel cells to generate onboard energy for the electric motor. Battery-Electric Buses (BEBs), which store the energy using an onboard battery that supplies the electric motor with the required energy.

Battery-Electric Buses (BEBs) is considered a long-term cost-effective alternative, since it guarantees the optimal use of the charging process during electricity off-peak times

as well as it does not require major infrastructure compared to Overhead Trolly Electric Buses (EESI, 2018). Moreover, it was found that BEBs have a lower energy consumption rate on average compared to other e-Buses. Besides, BEB technology is considered more mature for mass adoption compared to FCEBs that still face several technical challenges due to the lack of global technical regulations for hydrogen vehicles and hydrogen productions (Chen et al., 2007; Trencher and Edianto, 2021).

In this respect, the BEB system is considered a more economical and agile option; and adopting it in transit networks is considered the ultimate solution to benefit from the clean electricity generation in the transportation context. However, before phasing out diesel buses to adopt the BEBs system, governments and transit planners need to understand the impact of an all-electric fleet on the existing energy infrastructure.

Implementing BEBs in transit systems requires meticulous infrastructure planning, where each bus route must have an appropriate charging scheme that adheres to the operation schedule (El-Taweel et al., 2020; Schwurzinger, 2012). The charging stations and the onboard battery have the highest bearing on the total system cost in a BEB transit system (Wellik et al., 2021). Indeed, the state-of-the-art has been focusing on the optimization of BEB systems and the associated infrastructure network, intending to minimize the total cost of ownership related to BEB implementation in transit systems (El-Taweel et al., 2020; He et al., 2019; Rupp et al., 2020). However, minimizing the cost depends solely on an accurate estimation of the consumed energy during operation.

Consequently, predicting accurate rates of the consumed energy during the bus trips (i.e. the energy consumption or E_C) is the cornerstone for all studies related to BEB,

including those focusing on fleet optimization, battery capacity/performance, the spatial distribution of infrastructure, component sizing (battery & charger power), GHG, and impact on the utility grid (Abdelaty et al., 2021; Liu et al., 2017; Qi et al., 2018a; Teoh et al., 2018; Vepsäläinen et al., 2018a).

The accurate estimation of the energy consumption rates is highly uncertain since the operational range of the BEBs in real-world environments relies on the circumstances surrounding the bus trips and the route where they operate. The uncertainties associated with BEB's energy consumption are due to parameters concerning the speed profiles, driver behaviour, route topology, passenger-hour load profile, and weather conditions, among others (Abdelaty et al., 2021; Hjelkrem et al., 2021; Rupp et al., 2020). These parameters are specified in detail in Section 1.1.

1.1. Parameters Impacting BEBs' Energy Consumption

The energy consumption rates vary significantly based on abundant parameters classified into four sets. First, the vehicle parameters that include all the physical parameters related to the bus and the battery. These parameters can be defined as; the bus mass (m), which is the total weight of the bus elements, including the curb weight, motor, gearbox, wheels, battery, and passengers (Gallet et al., 2018; Vepsäläinen et al., 2019); the frontal area (A_F), which depends on the BEB's type (Franca, 2015; Lajunen, 2018); the rolling resistance coefficient (C_r), which varies based on the road surface and the weather conditions (Gallet et al., 2018; Vepsäläinen et al., 2019); the drag coefficient (C_d), which depends on the BEB design, speed, and the mass and (Gallet et al., 2018; Lajunen, 2014); and the initial status

of charge (SoC_i), battery capacity (C_B), and the battery temperature (T_B), which are battery-related parameters (Franca, 2018; Kivekäs et al., 2018; Vepsäläinen et al., 2019).

Second, the operational parameters that include the parameters related to transit operation and driver behaviour. The transit operation parameters used in the literature include the number of stops along the bus route (S_N), spacing between the stops (S_S), stop density (S_D), and passenger loading (P_L) (Gallet et al., 2018; Kivekas et al., 2018). Whilst the driver behaviour (i.e. slow, normal, or aggressive) can be estimated during the bus trips using average speed (V_a), acceleration rates (a), and deceleration rates (d); which directly impacts the consumed energy during the acceleration process, and the recovered energy during the deceleration process (Gao et al., 2017; Lajunen, 2014).

Third, the topological parameters include bus route features such as route length (L) and road gradient (g) (Abdelaty et al., 2021; De Cauwer et al., 2017b).

Fourth, the external parameters that include environmental conditions and auxiliary parameters. The environmental parameters are ambient temperature (T_A) and air density (P_a). Whilst the auxiliary parameters consist of; heating, ventilation, and air conditioning powers (i.e. HVAC power) and it had been predicted as a function of the ambient temperature such as the model developed by (Tammi and Lajunen, 2016) and used by (Vepsäläinen et al., 2019); the regenerative brake (Br_{reg}), which represents the amount of recovered energy during the stops and deceleration rates; the auxiliary power, which includes all other bus elements that consume energy such as bus doors and power steering (Gallet et al., 2018; Vepsäläinen et al., 2018a; Vepsäläinen et al., 2019).

Furthermore, the uncertainty in estimating the E_C rates complicates the research on transit network design and optimal transit operating profile configuration that promotes BEB energy usage. Towards that end, the literature suggests two approaches for estimating and modelling the E_C of the BEBs, as clarified in Section 1.2. (Hjelkrem et al., 2021; Pamuła and Pamuła, 2020; Vepsäläinen et al., 2018b).

1.2. Energy Consumption Estimation

The first approach is collecting real-world BEBs' E_C data through field experiments. However, this approach frequently describes the E_C behaviour within a restricted number of real-world conditions, which restricts the generality of the results owing to the small number of unique trials (De Cauwer et al., 2017b; Qi et al., 2018b). The second approach is using E_C simulation models that mimic the performance of the BEBs in the real world. It depends on vehicular kinematic and dynamic conditions representing the bus motion (Hahn and Valentine, 2019; Rupp et al., 2019). However, it has some limitations since the simulation approach often depends on the vehicular parameters and the dynamic conditions, which leads to a deficiency in studying the impact of the operational and topological parameters on the BEBs' E_C.

Both approaches are viable, but their practical significance and simplicity for transit operators and municipalities are restricted. This is due to the required technical skills and/or expenditures to simulate or test the BEB fleets (Abdelaty et al., 2021; De Cauwer et al., 2017a; Mohamed et al., 2018).

To that end, there is a rising demand from transit service providers and municipalities to develop data-driven models that can present a BEB's E_C prediction without the necessity for complex simulation models or when there is a shortage of access to BEB real-world data. The data-driven model allows predicting the energy consumption based on big amounts of observed data from different energy consumption parameters representing the BEB transit operation. Section 1.3. presents the different techniques used in the literature for developing the E_C data-driven prediction models.

1.3. Data-driven Energy Prediction Models

The aforementioned parameters (i.e. vehicular, operational, topological, and external parameters) are often used, separately or jointly, in developing the data-driven prediction models through four modelling techniques, including:

1) Multiple Regression Analysis (MRA) determines the most significant input parameters to estimate the E_C based on how much each parameter reduces the residual sum of squares (Liu et al., 2017; Teoh et al., 2018). A few studies had employed linear regression models to predict the BEB's E_C and study the impact of the various parameters such as passenger loading, trip time, the spacing between stops, and the difference between route elevations on the E_C (Pamuła and Pamuła, 2020; Teoh et al., 2018; Vepsäläinen et al., 2018b).

Despite their scarcity in the BEB domain, MRA models have been previously used to predict the E_C of electric vehicles (EVs) with promising results. The linear regression models are used to predict the EV's E_C based on several parameters, including the rolling

resistance, aerodynamic drag, average speed, acceleration/deceleration rates, HVAC, ambient temperature, state of charge, and road gradient (De Cauwer et al., 2017b; Liu et al., 2017; Qi et al., 2018b; Yuan et al., 2017). While Galvin et al. (2017) conducted a multivariate linear regression model to study the driver behaviour's impact on the EV's E_C. Moreover, Liu et al. (2018) and Wang et al. (2017) used ordinary least squares regression and multilevel mixed-effects regression models to assess the impact of the ambient temperature, road gradient, and HVAC on the EV's E_C.

The results of the BEB studies show the high significance of the number of stops and the driver aggressiveness on the E_C rates (Vepsäläinen et al., 2018b). Also, Teoh et al. (2018) concluded that the route length and passenger loading have a high impact on the BEB's E_C. At the same time, the EVs studies concluded the high impact of the average speed, HVAC power, and acceleration/deceleration rates on the E_C rates (De Cauwer et al., 2017; Galvin, 2017; Liu et al., 2018). Besides, the road gradient has the highest impact on the E_C (De Cauwer et al., 2017; Liu et al., 2018; Liu et al., 2017; Wang et al., 2017).

2) Deep Learning Neural Network (DLNN) intends to recognize the fundamental relationships in a set of data through a process that mimics how the human brain operates (Dreyfus, 2005). It was used by Pamula and Pamula (2020) to estimate the E_C for BEBs based on route characteristics such as travel time, the distance between bus stops, weather, and elevation differences.

There is a diversity of studies outside the BEB domain that applied DLNN models for the EVs and the internal combustion engine (ICE) vehicles. For example, (Kanarachos et al., 2019; Shankar and Marco, 2013) trained a recurrent neural network (RNN) to predict

the EVs' E_C considering the average speed, average acceleration/deceleration rates, stop density, trip distance, and trip time. While Masikos et al. (2014) applied the general regression neural network (GRNN) to implement the EVs' E_C model based on the state of charge, vehicle mass, road gradient, and ambient temperature. While for ICE vehicles, (Ping et al., 2019; Yamashita et al., 2018) developed a neural network model for predicting the fuel consumption for the vehicles based on the driver behaviour parameters. Moreover, (Du et al., 2017; Wu and Liu, 2011, 2012) used the back-propagation neural network and the radial basis function neural network to develop a predictive model for vehicles' fuel consumption based on speed, engine style, and vehicle weight.

The results of (Pamula and Pamula, 2020) show that the E_C rates change significantly with the variation in the distance and spacing between stops. Moreover, Diaz Alvarez et al. (2014) demonstrated that average speed had the lowest impact on the E_C rates, while Qi et al. (2018a) concluded that average speed is the highest significant parameter in estimating the E_C rates.

3) Interpolation Method (IM) interpolates the value of the dependent parameter as a function of a set of independent parameters (Vepsäläinen et al., 2019). The surrogate modelling technique is used as an interpolation method to perform a sensitivity analysis for E_C rates based on the uncertainties in the weather (i.e. ambient temperature and rolling resistance), the driver behaviour (i.e. average speed and driver aggressiveness), and the state of charge (Vepsäläinen et al., 2019; Vepsäläinen et al., 2018b).

The results show that the ambient temperature has the highest impact on the E_C, while the rolling resistance has the second-highest impact on the E_C (Vepsäläinen et al.,

2019). Moreover, there is a high significance of stop density and driver aggressiveness on the E_C rates (Vepsäläinen et al. 2018b).

4) Machine Learning (ML) provides the model with the ability to automatically analyze and interpret the underlying patterns and structures in the data (Abdelaty et al., 2021; Ma et al., 2021). The ML models in BEB's E_C application are limited, however, Ma et al. (2021) applied a gradient boosting decision tree to evaluate the impact of various parameters on the E_C of the diesel and electric buses. They found that the number of stops and their distance significantly affect the BEB's E_C more than the diesel buses.

The above studies suggest some important modelling techniques in predicting the BEB's E_C, as well as the parameters affecting the E_C rates. However, through this brief review, some research gaps are defined.

1.4. Research Gaps

Battery electric buses (BEBs) represent a promising solution for having public transportation systems working with a clean energy source. Due to the nature of fixed routes, timely operation, and scheduled trips, the implementation of BEBs in the transit context is considered a seamless transition towards a zero greenhouse gases transit system. However, and similar to personal electric vehicles (EVs), range anxiety remains a major hurdle. In the bus transit context, range anxiety, often referred to as energy consumption uncertainty, is a significant deterrent for mainstream implementation of BEBs. Predominantly, several projects and research are conducted globally for better understanding the uncertainty in the energy consumption rates. However, the BEB's energy

consumption varies due to uncertainty in operational, topological, and environmental attributes. This variation is hard to estimate properly, which might affect the BEB's daily operation in transit systems.

Towards that end, the research gaps can be specified through the following points:

- To implement the BEBs in the bus transit systems, there is needs to understand the transit network behaviour in the normal condition and under the potential disruptions. For example, any disruptive events that occur, such as an electricity outage in one of the charging bus stations, will lead to a disruption in the entire transit system. Therefore, understanding and assessing the bus transit network's behaviours will contribute towards avoiding the challenges that can face the BEB's implementations in the future.

- There are valid calls to study the intertwined relationships between the BEB's E_C from one hand and the operational (e.g. driver behaviour and passenger loading), topological (e.g. road gradient and road condition), and external (e.g. weather condition, HVAC power, and road condition) features of transit networks from the other hand. However, the studies that quantify the impact of these features on the BEB's E_C are scarce.

- Although there are different data-driven models are utilized in the literature, their results are not consistent. Besides, there is no information about which technique is better or more accurate in predicting the BEB's E_C rates in the transit systems.

- To the best of our knowledge, there is a lack in the data-driven prediction models that combine the speed profiles, route topology, passenger-hour load profile, and weather conditions through open-source low-resolution data to surmount the difficulties in collecting real-world data. That is, in turn, disregarding the impact of substantial parameters such as road gradient, road condition, and stop density, which will reduce the accuracy of the model estimation.

- A comprehensive validation process is needed to stabilize the data-driven models in the BEBs' E_C research for overcoming the uncertainty in the operational, topological, and external parameters.

1.5. Research Objectives

Toward that end, and considering the defined research gap, this study aims at developing simulation, data-driven, and low-resolution models using big data to quantify the BEB's E_C, with the overarching goal of developing a comprehensive planning framework for BEB implementation in bus transit networks. This aim is achieved through four interwind objectives (as shown in Figure 1.1).

Objective one

Quantifying the operational and topological characteristics of bus transit networks using complex network theory. This objective provides a fundamental base for understanding bus transit networks' behaviour under disruptive events such as electricity outages in the case of BEBs. Although the thesis focuses on vehicle-level energy consumption modelling,

simulation, and prediction, the ultimate goal is to provide decision-makers data-supported recommendation for implementing e-Bus in the transit context.

Objective two

Developing and validating a robust BEB simulation model for energy consumption estimation. The simulation model is utilized to investigate the impacts of the vehicular, operational, topological, and external parameters on the E_C of BEBs, using a generated full-factorial experimental dataset (i.e. training dataset), as well as a generated fractional-factorial experimental dataset (i.e. testing dataset). A regression analysis and sensitivity analysis are used to quantify the parameters with high significant on the BEB's E_C, among all other parameters.

Objective three

Developing, comparing, and assessing the feasibility of big-data analytics and data-driven prediction models to numerically estimate BEB's E_C. These models alleviate the need for energy simulation models and offer practitioners ready-to-implement tools for BEB's E_C.

The selected models belong to four techniques, Multiple Regression Analysis (MRA), Deep Learning Neural Network (DLNN), Interpolation Method (IM), and Machine Learning (ML), and were partially selected as the dominant models applied in previous studies. They comprise the multiple linear regression, interpolation method, decision tree, gradient-boosting decision tree, support vector machine learning, multiple perception neural network, and radial-basis neural network.

Such a comparison between data-driven prediction models will inform transit providers and academia alike on the best data-driven model for BEBs energy consumption prediction, as well as the limitations of each modelling approach.

Objective four

Developing and assessing the open-source low-resolution data-based framework to estimate BEB's E_C. This framework integrates the modelling efforts in objectives 1-3 and offers a practical mobilization of knowledge for transit providers. The framework relies on low-resolution open-source data collected by all transit providers. These include weather and topological data collected using Environment Canada's weather website and CanElevation series, as well as operational/topological data collected using the automatic vehicle location (AVL), automatic passenger count (APC), and topological parameters.

Overall, the thesis provides genuine contributions to BEB research and offers a practical framework for addressing the energy consumption uncertainty associated with BEB operation in the transit context. Further, the results offer transit planners the means to set up the optimum transit operations profile that improves the use of BEB energy.

Chapter Five

Low-Resolution Data Preparation
- Six bus routes in Hamilton, Canada
- Local, arterial, and collector

Locate AVL Data Inventory
- AVL dataset
- Prepare average speed values
- Extract one-minute positioning data

Locate APC Data Inventory
- APC dataset
- Prepare occupancy values
- Extract per-trip occupancy levels

Locate CanElevation Data
- Elevation dataset
- Prepare elevation values
- Estimate constant grade

Locate Weather Data
- Weather dataset
- Prepare weather values
- Estimate HVAC consumption

Full-Fractional Design
- 6,048 operation scenario

Base Model Data

Observed Ec data
- Simulation-Based Ec

Predicted Ec data
- Prediction-Based Ec

Checking Model Performance
- Evaluate the model error through three-step validation method

Trip-Level Validation (Per Parameter)

One-Minute Single Trip Validation

Error Analysis

Network Science

Topological Measurements
- 40 Bus transit networks

Degree Centrality

Betweenness Centrality

Average Path Length

Clustering Coefficient

Static-Robustness
- Three measures

Robustness Metric

Robustness Indicator

Critical Threshold

Cluster I
- Small node-sensitive networks

Cluster II
- Small link-sensitive networks

Cluster III
- Medium less-sensitive networks

Dynamic-Robustness
- Three targeted removal

Bus stations with highest betweenness

Bus stations with highest degree

Bus stations with highest link weight

Losses in total trip kilometers

Losses in total route distance

Losses in total service frequency

Chapter Two

Future Research

Chapter Three

Operational Parameters
- Average speed
- Acceleration rates
- Deceleration rates
- State of charge
- Passenger loading
- Road condition

Vehicular Parameters
- Bus mass
- Rolling resistance
- Drag coefficient

Model Validation
- ADVISOR
- Autonomie
- Real-world Altoona

Data-Driven Prediction Models
- Multiple linear regression
- Radial basis function interpolation
- Decision tree
- Gradient-boosting decision tree
- Support vector machine learning
- Multilayer perception neural network
- Radial-basis neural network

Topological Parameters
- Road gradient
- Route length
- Stop density

External Parameters
- Weather condition
- HVAC power
- Auxiliary power
- Regenerated power
- Air Density
- Ambient Temp.

Full-Factorial Design
- Generating big data
- Training dataset
- 907,199 operation scenarios

Fractional-Factorial Design
- Generating big data
- Testing dataset
- 169,344 operation scenarios

Simulation Model
- Estimating the Ec for training and testing datasets
- MATLAB Simulink
- BEB NewFlyer XE40
- Battery capacity of 200 kWh

Regression Analysis
- Significant parameters
- Parameters' correlations
- Predicting the Ec for training and testing datasets
- Validating the model

Sensitivity Analysis
- Sobol indices
- First-order effect
- Total-order effect

Comparing and evaluating the performance of the models to identify the best data-driven prediction model

Chapter Four

Figure 1-1. A flow chart illustrating the four intertwined objectives of the thesis

1.6. Thesis Outline

The dissertation includes six chapters that are organized as follows:

Chapter one is an introduction to the implementation of the BEBs in the transit context. It includes a comprehensive review of the BEB's E_C and the parameters that affect it. Besides, the methods used to measure the E_C and the data-driven models used in the literature to predict the E_C are presented in detail. Additionally, it defines the research gap and the objectives of this dissertation.

Chapter two is the accomplished study for the fulfillment of objective one in the dissertation. It is published as "Abdelaty, H., Mohamed, M., Ezzeldin, M., & El-Dakhakhni, W. (2020). Quantifying and classifying the robustness of bus transit networks. Transportmetrica A: Transport Science, 16(3), 1176-1216. https://doi.org/10.1080/23249935.2020.1720042".

Author contributions: conceptualization, Mohamed, M.; methodology, Abdelaty, H. and Mohamed, M.; software, Abdelaty, H.; formal analysis, Abdelaty, H.; investigation, Abdelaty, H.; resources, Mohamed, M., Ezzeldin, M., and El-Dakhakhni, W.; data curation, Abdelaty, H. and Mohamed, M.; first draft preparation, Abdelaty, H.; review and editing, Mohamed, M., Ezzeldin, M., and El-Dakhakhni, W.; visualization, Abdelaty, H.; supervision, Mohamed, M.; project administration, Mohamed, M.; funding acquisition, Mohamed, M. All authors have read and agreed to the published version of the manuscript.

Chapter three is the accomplished study for the fulfillment of objective two in the dissertation. It is published as "Abdelaty, H., & Mohamed, M. (2021). A prediction model

for battery electric bus energy consumption in transit. Energies, 14(2824), 1-26. https://doi.org/10.3390/en14102824".

Author contributions: conceptualization, Mohamed, M.; methodology, Abdelaty, H. and Mohamed, M.; software, Abdelaty, H.; validation, Abdelaty, H.; formal analysis, Abdelaty, H.; investigation, Abdelaty, H.; resources, Mohamed, M.; data curation, Abdelaty, H. and Mohamed, M.; first draft preparation, Abdelaty, H.; review and editing, Mohamed, M.; visualization, Abdelaty, H.; supervision, Mohamed, M.; project administration, Mohamed, M.; funding acquisition, Mohamed, M. Both authors have read and agreed to the published version of the manuscript.

Chapter four is the accomplished study for the fulfillment of objective three in the dissertation. It is published as "Abdelaty, H., Al-Obaidi, A., Mohamed, M., & Farag, H. (2021). Machine learning prediction models for battery-electric bus energy consumption in transit. Transportation Research Part D: Transport and Environment, 96(102868), 1-27. https://doi.org/10.1016/j.trd.2021.102868".

Author contributions: conceptualization, Mohamed, M.; methodology, Abdelaty, H., Al-Obaidi, A., and Mohamed, M.; software, Abdelaty, H.; validation, Abdelaty, H. and Al-Obaidi, A.; formal analysis, Abdelaty, H.; investigation, Abdelaty, H.; resources, Mohamed, M. and Farag, H.; data curation, Abdelaty, H. and Mohamed, M.; first draft preparation, Abdelaty, H.; review and editing, Mohamed, M. and Farag, H.; visualization, Abdelaty, H.; supervision, Mohamed, M.; project administration, Mohamed, M.; funding acquisition, Mohamed, M. All authors have read and agreed to the published version of the manuscript.

Chapter five is the accomplished study for the fulfillment of objective four in the dissertation. It is under review "Abdelaty, H., & Mohamed, M. (2021). A Framework for BEB Energy Prediction Using Low-Resolution Open-Source Data-Driven Model. Transportation Research Part D: Transport and Environment.

Author contributions: conceptualization, Mohamed, M.; methodology, Abdelaty, H. and Mohamed, M.; software, Abdelaty, H.; validation, Abdelaty, H.; formal analysis, Abdelaty, H.; investigation, Abdelaty, H.; resources, Mohamed, M.; data curation, Abdelaty, H. and Mohamed, M.; first draft preparation, Abdelaty, H.; review and editing, Mohamed, M.; visualization, Abdelaty, H.; supervision, Mohamed, M.; project administration, Mohamed, M.; funding acquisition, Mohamed, M. Both authors have read and agreed to the published version of the manuscript.

Chapter six includes the conclusion and discussion of the results. As well, the limitation and future work are defined.

1.7. References

Abdelaty, H., Al-Obaidi, A., Mohamed, M., Farag, H.E.Z., 2021. Machine learning prediction models for battery-electric bus energy consumption in transit. Transportation Research Part D: Transport and Environment 96.

Chen, F., Fernandes, T.R.C., Yetano Roche, M., da Graça Carvalho, M., 2007. Investigation of challenges to the utilization of fuel cell buses in the EU vs transition economies. Renewable and Sustainable Energy Reviews 11(2), 357-364.

De Cauwer, C., Verbeke, W., Coosemans, T., Faid, S., Van Mierlo, J., 2017a. A Data-Driven Method for Energy Consumption Prediction and Energy-Efficient Routing of Electric Vehicles in Real-World Conditions. Energies 10, 608.

De Cauwer, C., Verbeke, W., Coosemans, T., Faid, S., Van Mierlo, J., 2017b. A Data-Driven Method for Energy Consumption Prediction and Energy-Efficient Routing of Electric Vehicles in Real-World Conditions. Energies 10(5).

Dreyfus, G., 2005. Neural Networks: Methodology and Applications. Springer.

Du, Y., Wu, J., Yang, S., Zhou, L., 2017. Predicting vehicle fuel consumption patterns using floating vehicle data. J Environ Sci (China) 59, 24-29.

EESI, 2018. Fact Sheet: Battery Electric Buses: Benefits Outweigh Costs, Environmental and Energy Study Institute.

El-Taweel, N.A., Farag, H.E.Z., Mohamed, M., 2020. Integrated Utility-Transit Model for Optimal Configuration of Battery Electric Bus Systems. IEEE Systems Journal 14(1), 738-748.

Environment-Canada, 2021. Greenhouse Gas Emissions, Canadian Environmental Sustainability Indicators.

Franca, A., 2015. Electricity consumption and battery lifespan estimation for transit electric buses: drivetrain simulations and electrochemical modelling, Department of Mechanical Engineering. University of Victoria, p. 162.

Franca, A., 2018. Electricity consumption and battery lifespan estimation for transit electric buses: drivetrain simulations and electrochemical modelling.

Gallet, M., Massier, T., Hamacher, T., 2018. Estimation of the energy demand of electric buses based on real-world data for large-scale public transport networks. Applied Energy 230, 344-356.

Galvin, R., 2017. Energy consumption effects of speed and acceleration in electric vehicles: Laboratory case studies and implications for drivers and policymakers. Transportation Research Part D: Transport and Environment 53, 234-248.

Gao, Z., Lin, Z., LaClair, T.J., Liu, C., Li, J.-M., Birky, A.K., Ward, J., 2017. Battery capacity and recharging needs for electric buses in city transit service. Energy 122, 588-600.

Hahn, B., Valentine, D., 2019. SIMULINK® Toolbox. In Essential MATLAB for Engineers and Scientists.

He, Y., Song, Z., Liu, Z., 2019. Fast-charging station deployment for battery electric bus systems considering electricity demand charges. Sustainable Cities and Society 48.

Hjelkrem, O.A., Lervåg, K.Y., Babri, S., Lu, C., Södersten, C.-J., 2021. A battery electric bus energy consumption model for strategic purposes: Validation of a proposed model structure with data from bus fleets in China and Norway. Transportation Research Part D: Transport and Environment 94.

Kanarachos, S., Mathew, J., Fitzpatrick, M.E., 2019. Instantaneous vehicle fuel consumption estimation using smartphones and recurrent neural networks. Expert Systems with Applications 120, 436-447.

Kivekäs, K., Lajunen, A., Vepsäläinen, J., Tammi, K., 2018. City Bus Powertrain Comparison: Driving Cycle Variation and Passenger Load Sensitivity Analysis. Energies 11(7).

Kivekas, K., Vepsäläinen, J., Tammi, K., 2018. Stochastic Driving Cycle Synthesis for Analyzing the Energy Consumption of a Battery Electric Bus. IEEE Access 6, 55586-55598.

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

Lajunen, A., 2018. Lifecycle costs and charging requirements of electric buses with different charging methods. Journal of Cleaner Production 172, 56-67.

Liu, K., Wang, J., Yamamoto, T., Morikawa, T., 2018. Exploring the interactive effects of ambient temperature and vehicle auxiliary loads on electric vehicle energy consumption. Applied Energy 227, 324-331.

Liu, K., Yamamoto, T., Morikawa, T., 2017. Impact of road gradient on energy consumption of electric vehicles. Transportation Research Part D: Transport and Environment 54, 74-81.

Ma, X., Miao, R., Wu, X., Liu, X., 2021. Examining influential factors on the energy consumption of electric and diesel buses: A data-driven analysis of large-scale public transit network in Beijing. Energy 216.

Masikos, M., Demestichas, K., Adamopoulou, E., Theologou, M., 2014. Mesoscopic forecasting of vehicular consumption using neural networks. Soft Computing 19(1), 145-156.

Mohamed, M., Ferguson, M., Kanaroglou, P., 2018. What hinders adoption of the electric bus in Canadian transit? Perspectives of transit providers. Transportation Research Part D: Transport and Environment 64, 134-149.

Pamula, T., Pamula, W., 2020. Estimation of the Energy Consumption of Battery Electric Buses for Public Transport Networks Using Real-World Data and Deep Learning. Energies 13(9).

Pamuła, T., Pamuła, W., 2020. Estimation of the Energy Consumption of Battery Electric Buses for Public Transport Networks Using Real-World Data and Deep Learning. Energies 13(9).

Ping, P., Qin, W., Xu, Y., Miyajima, C., Takeda, K., 2019. Impact of Driver Behavior on Fuel Consumption: Classification, Evaluation and Prediction Using Machine Learning. IEEE Access 7, 78515-78532.

Qi, X., Wu, G., Boriboonsomsin, K., Barth, M.J., 2018a. Data-driven decomposition analysis and estimation of link-level electric vehicle energy consumption under

real-world traffic conditions. Transportation Research Part D: Transport and Environment 64, 36-52.

Qi, Z., Yang, J., Jia, R., Wang, F., 2018b. Investigating Real-World Energy Consumption of Electric Vehicles: A Case Study of Shanghai. Procedia Computer Science 131, 367-376.

Ritchie, H., Roser, M., 2020. CO_2 and Greenhouse Gas Emissions.

Rupp, M., Handschuh, N., Rieke, C., Kuperjans, I., 2019. Contribution of country-specific electricity mix and charging time to environmental impact of battery electric vehicles: A case study of electric buses in Germany. Applied Energy 237, 618-634.

Rupp, M., Rieke, C., Handschuh, N., Kuperjans, I., 2020. Economic and ecological optimization of electric bus charging considering variable electricity prices and CO2eq intensities. Transportation Research Part D: Transport and Environment 81.

Schwurzinger, 2012. Report on electrified public transport bus system.

Shankar, R., Marco, J., 2013. Method for estimating the energy consumption of electric vehicles and plug-in hybrid electric vehicles under real-world driving conditions. IET Intelligent Transport Systems 7(1), 138-150.

Tammi, K., Lajunen, A., 2016. Energy consumption and carbon dioxide emission analysis for electric city buses. Curran Associates, Inc.

Teoh, L.E., Khoo, H.L., Goh, S.Y., Chong, L.M., 2018. Scenario-based electric bus operation: A case study of Putrajaya, Malaysia. International Journal of Transportation Science and Technology 7(1), 10-25.

Trencher, G., Edianto, A., 2021. Drivers and Barriers to the Adoption of Fuel Cell Passenger Vehicles and Buses in Germany. Energies 14(4).

Vepsäläinen, J., Kivekäs, K., Otto, K., Lajunen, A., Tammi, K., 2018a. Development and validation of energy demand uncertainty model for electric city buses. Transportation Research Part D: Transport and Environment 63, 347-361.

Vepsäläinen, J., Otto, K., Lajunen, A., Tammi, K., 2019. Computationally efficient model for energy demand prediction of electric city bus in varying operating conditions. Energy 169, 433-443.

Vepsäläinen, J., Ritari, A., Lajunen, A., Kivekäs, K., Tammi, K., 2018b. Energy Uncertainty Analysis of Electric Buses. Energies 11(12).

Wang, J.-b., Liu, K., Yamamoto, T., Morikawa, T., 2017. Improving Estimation Accuracy for Electric Vehicle Energy Consumption Considering the Effects of Ambient Temperature. Energy Procedia 105, 2904-2909.

Wellik, T.K., Griffin, J.R., Kockelman, K.M., Mohamed, M., 2021. Utility-transit nexus: Leveraging intelligently charged electrified transit to support a renewable energy grid. Renewable and Sustainable Energy Reviews 139.

Wu, J.-D., Liu, J.-C., 2011. Development of a predictive system for car fuel consumption using an artificial neural network. Expert Systems with Applications 38(5), 4967-4971.

Wu, J.-D., Liu, J.-C., 2012. A forecasting system for car fuel consumption using a radial basis function neural network. Expert Systems with Applications 39(2), 1883-1888.

Yamashita, R.-J., Yao, H.-H., Hung, S.-W., Hackman, A., 2018. Accessing and constructing driving data to develop fuel consumption forecast model. IOP Conference Series: Earth and Environmental Science 113.

Yuan, X., Zhang, C., Hong, G., Huang, X., Li, L., 2017. Method for evaluating the real-world driving energy consumptions of electric vehicles. Energy 141, 1955-1968.

CHAPTER TWO

2. QUANTIFYING AND CLASSIFYING THE ROBUSTNESS OF BUS TRANSIT NETWORKS

Abstract: This study adopts Complex Network Theory in the context of bus transit network. The study aims at quantifying the topological characteristics and assessing the validity of static robustness metrics as expressive measures of transit networks robustness. In addition, dynamic-robustness indices, that consider transit operational profile, are utilized to measure the cascading impacts of disruptive events. The analysis is based on a dataset of 40 bus transit networks. The results indicate that bus networks do not follow any major network types; scale-free, small-world, or random. Furthermore, the static-robustness metrics produced contradictory results, which raises valid concerns on their applicability. The dynamic-robustness indices indicated significant cascading impacts resulting from node removal relative to the removal of links. This was further examined through a two-step cluster analysis, which resulted in three distinct clusters of networks; small node-sensitive; small link-sensitive; and medium less-sensitive networks. These findings are directed to inform a robustness-based design of bus networks.

Keywords: bus transit networks; Complex Network Theory; static/dynamic-robustness

2.1. Introduction

The importance of public transit networks in modern communities cannot be overemphasized. Transit reliability, in particular, is considered the keystone for service quality and integral service component to promote the use of transit (Mahmoud and Hine 2013; Mahmoud and Hine 2016; Mahmoud et al. 2011). In typical operation conditions, transit reliability is measured using various indices such as adherence to schedule, frequency, and magnitude of delays among other measures. However, abnormal external/internal conditions (e.g. disruptive events such as power outage in any of the bus charging stations) could have significant consequences on the reliability and the delivery of service. As such, evaluating the behaviour of bus transit networks under disruptive scenarios is an essential aspect for their quality design and operation (Frappier et al. 2018; Hassan et al. 2013; Qiu et al. 2014; Wei et al. 2017).

In this respect, Complex Network Theory (CNT) advocated by Erdös and Rényi (1960), which aims at evaluating the interdependencies between the components of a system, has been widely used in the transportation context. The applications of CNT in transportation could be classified into two broad research streams: static- and dynamic based assessments of transportation networks (Mattsson and Jenelius 2015). The static based assessment approach is rooted in graph theory. The assessment is grounded on the graphical, often referred to as topological, relationships between the comprising components of the network (e.g. the relationships between link/route and nodes/hubs/stations). The main aim of this approach is to evaluate the topological characteristics of transportation networks. While the dynamic-based assessment approach

incorporates additional attributes to represent the operational features of different transportation systems (e.g. mode, traffic flow, capacity, path length, etc.). This approach assesses the system behaviour under disruption scenarios (Mattsson and Jenelius 2015; Sun et al. 2018). Although both research streams are based on the applications of CNT, their contributions are distinctly different.

These approaches have been applied to different transportation networks such as air traffic (Hassan et al. 2013; Wandelt et al. 2017; Zou et al. 2013), rail (Aldrich et al. 2015; Lozano et al. 2006; Sun et al. 2018; Zhu et al. 2018), metro (subway) (Derrible and Kennedy 2010; Sun and Guan 2016; Wang et al. 2017; Wu et al. 2018), and bus transit networks (Huang et al. 2015; Shanmukhappa et al. 2018; Sui et al. 2012). As a result, several measures, in addition to topological measures, have emerged in the CNT-transportation literature such as robustness, vulnerability, and resilience. In general, robustness refers to the ability of a system to "maintain its basic functions in the presence of internal and external errors" (Barabási 2016) pp-303. The vulnerability could be defined as "the sensitivity of networks" (Sun et al. 2018) pp-13. While resilience is described as "the system's capability to persist when exposed to changes or shocks" (Mattsson and Jenelius 2015) pp-19. The scope of this study focuses only on robustness measures; both vulnerability and resilience measures are beyond the scope of the present study.

Despite previous attempts, there are no universal definitions of network robustness in the transportation context (Mattsson and Jenelius 2015). While we do not aim to resolve this issue, we are using the following definitions in the context of the present study. Static-robustness is defined as a holistic network-level measure that quantifies the overall network

performance as a function of its comprising components. Dynamic-robustness quantifies the performance of the network under different disruption scenarios to the network components.

The present study aims at utilizing CNT measures in the transportation context and focuses on bus transit networks, which surprisingly received less attention in the literature relative to rail, air and metro networks. More specifically, the study contributes to the current literature through 1) the assessment of bus transit network topological characteristics and the associated network model; 2) the development of comparative analysis of static-based measures of bus transit network robustness; and 3) the behaviour of bus transit networks under different disruption scenarios using dynamic-measures of robustness.

The study utilizes a dataset of 40 bus transit networks in Canada. Such a large dataset facilitates a direct comparison between static- and dynamic-measures and their corresponding impacts on bus transit networks. Indeed, this is expected to offer significant contributions to the bus transit design, planning, and operation that could also be mobilized to multiple other contexts.

Following this introduction, a review of the CNT applications in transit literature is presented in Section 2.2. Section 2.3 details the data collection process and explains the utilized methods. Section 2.4 reports the results with emphasis on bus transit network topological characteristics and the robustness, both static and dynamic, of bus transit networks. Section 2.5 discusses a classification approach for bus transit networks robustness. Lastly, Section 2.6 concludes the study and offers directions for policymakers

and planners involved in the design and operation of bus transit networks.

2.2. Applications of Complex Network Theory in Transit

As highlighted in the introduction, CNT has introduced a new spectrum of transit network measures (See Appendix 2.1). First, static-measures have been utilized to assess the topological characteristics of transit networks based on the graphical arrangements of nodes and links. In this respect, four static-measures have been frequently operationalized in the literature including: Degree Centrality, Betweenness Centrality, Clustering Coefficient, and Average Shortest Path measures (Derrible and Kennedy 2009; Hossain and Alam 2017; Sun and Guan 2016; Sun et al. 2018; Wu et al. 2018; Zhu et al. 2018).

Degree Centrality is used to evaluate the importance of nodes in a network as well as the existence of dominant nodes (Hubs) (Hossain and Alam 2017; Sun et al. 2018). While Betweenness Centrality is used to evaluate the connection between a node to all other nodes that are not directly connected to it. This measure is often utilized to evaluate the connectivity of transit networks (Lozano et al. 2006; Pien et al. 2015; Wu et al. 2018; Zhu et al. 2018). Clustering Coefficient and Average Shortest Path measures provide indications on the number of transfers and alternative routes to travel through the network. Both highlight the effectiveness of the network topological profile (Hossain and Alam 2017; Sun and Guan 2016; Wu et al. 2018).

The estimation of the four static-measures depends on two main parameters: link weight and shortest path length, as will be highlighted in the methodology section. In the literature, the path length is frequently utilized as the number of steps (links) between any

pair of nodes and links are typically considered to be equally weighted such as the works of (Demšar et al. 2008; Han and Liu 2009; Roth et al. 2012). Such an approach limits our understanding of the dynamic behaviour of the transportation network under disruption scenarios and subsequently falls short to quantify the extent of cascading effects. In other words, despite the importance of the topological profile of transit networks, static-measures alone are unable to account for several fundamental parameters in transit operation. These include the impact of service frequency and passenger flow on the behaviour of transit networks under disruption scenarios (Lozano et al. 2018; Sun et al. 2018; Wu et al. 2018).

Therefore, additional measures are utilized in the literature to overcome these issues by incorporating transit operational features in the analysis. As such, two families of measures emerge in the literature static-based robustness and dynamic-based robustness.

First, a transit network static-robustness is assessed through a variety of measures. For example, Wang et al. (2017) developed 14 different measures to assess the robustness of metro networks. However, three measures are frequently adopted in the literature. The Robustness Indicator (r^T) developed by Derrible and Kennedy (2010) which was further modified by Wang et al. (2017) as the Robustness Metric (R^t), and the Critical Threshold (f_c) incepted by Barabási (2016). These measures provide a holistic (one value) indication on the network robustness.

The Robustness Indicator (r^T) assess the network robustness through the classification of routes and stations. In this respect, routes are classified to single and multiple, while stations are classified to end and transfer. Derrible and Kennedy (2010) argued that the relationships between these measures quantify the network robustness. This

approach was slightly modified in Wang et al. (2017), yet they followed the same concept. Barabási (2016) has advocated for a different measure, namely the Critical Threshold (f_c). The measure is derived from the degree of each node in the network. This approach is based on quantifying the fraction of non-operational nodes that result in a completely disconnected network.

Second, dynamic-based network robustness measures are based on an iterative assessment procedure. The procedure typically starts by removing a node/link from the network followed by assessing the impact of this removal on the behaviour of the entire network, as well as quantifying the cascading effects (Bababeik et al. 2017; M'cleod et al. 2017; Yang and Liu 2018). As such, a network could be assessed at different fractions of non-operational components (i.e. 10% non-operational nodes). The removal procedure is subject to several criteria ranging from random (M'cleod et al. 2017) to targeted (; Guo and Su 2017; von Ferber et al. 2012) removals. The latter is often utilized to target nodes/links with the highest attributes (e.g. betweenness, degree, or link weight) (Zhou and Liu 2017). The network robustness is assessed as the ratio between the number of non-operational nodes/links and the total number of nodes/links. These measures are detailed in the methodology section.

Aside from the measures, the utilization of CNT models in bus transit is scarce, as highlighted in Table 2.1. However, several observations could be drawn from the literature of rail and metro networks. First previous research efforts have focused on transit networks with, to a great extent, similar sizes and characteristics. With respect to topological measures, transit networks in general, do not fit a specific network model (e.g. small world

and random), instead, the model emerges from the size of the underlying transit network and the geographical distribution of nodes and links. That said, Huang et al. (2015) and Sui et al. (2012) have identified that public transit networks follow a small-world network model in their analysis of Beijing transit network. This contradicts the results presented in Shanmukhappa et al. (2018), where they have analyzed three bus networks and argued that Hong-Kong transit network behaves as a scale-free network, while networks in London and Bengaluru follow a random network model.

Table 2-1. Applications of CNT in the transportation literature

Authors	Context	Network	Assessment Type	Sample Size
Lozano et al. (2006)	-	Rail	Topological measures & Robustness (Static)	8 Networks
Scott et al. (2006)	Hypothetical	Highway	Robustness (Dynamic)	3 Networks
Berche et al. (2009)	Global	Urban Transit	Robustness (Dynamic)	14 Networks
Duan et al. (2009)	China	Bus	Robustness (Dynamic)	1 Network
Shanmukhappa et al. (2018)	Global	Bus	Topological measures	3 Networks
von Ferber et al. (2009)	Global	Urban Transit	Topological measures	14 Cities
Derrible and Kennedy (2010)	Global	Metro	Robustness (Static)	33 Networks
Sui et al. (2012)	China	Bus	Topological measures	9 Cities
Zou et al. (2013)	China	Bus	Robustness (Static & Dynamic)	1 Network
Aldrich et al. (2015)	USA	Rail and Highway	Topological measures	1 Network
Huang et al. (2015)	China	Bus and Rail	Robustness & Vulnerability (Dynamic)	1 City
Sun and Guan (2016)	China	Metro	Topological measures & Vulnerability	1 Networks
Hossain and Alam (2017)	Australia	Airport	Topological measures	20 Cities
Zhu et al. (2018)	China & Global	Airport and Rail	Topological measures	23 Chinese Cities and 13 International Cities
Wang et al. (2017)	Global	Metro	Topological measures	33 Networks
Derrible and Kennedy (2009)	Global	Subway	Topological measures	19 Networks
Wu et al. (2018)	Global	Metro	Topological measures & Robustness (Dynamic)	6 Networks
Shanmukhappa et al. (2018)	Global	Bus	Topological measures	3 Cities
Sun et al. (2018)	China	Rail	Topological measures & Vulnerability	1 Network
Saidi et al. (2017)	Global	Rail	Topological measures	6 Networks

The literature also shows that the Clustering Coefficient is very sensitive to the number of transfer stations that offer alternative paths. However, from a connectivity

perspective, rail networks could be seen as more connected relative to air networks based on different network topological measures (Berche et al. 2009; Sun and Guan 2016; von Ferber et al. 2012; Wang et al. 2017; Zhu et al. 2018). In general, metro and rail networks also exhibit a relatively higher level of robustness under random failures when compared to targeted removal (Berche et al. 2009; Sun and Guan 2016; von Ferber et al. 2012; Wang et al. 2017; Zhu et al. 2018).

With respect to the robustness of bus transit networks, Duan et al. (2009) demonstrated that bus networks are sensitive to targeted removal of stations with high Betweenness Centrality values. While in contrast, random failures have insignificant influences on bus networks. Shanmukhappa et al. (2018) echoed similar conclusions and argued that the weight associated with nodes has a significant impact on the topological efficiency of transit networks. However, it should be noted that node weight was estimated as a function of Points of Interests (POIs) and the population density. In addition, the results of Shanmukhappa et al. (2018) are based on a super-node network structure, where spatially close by nodes are clustered together in the network.

Furthermore, there is no consensus in the literature on how to quantify a transit network robustness. In general, the dominant approach in estimating robustness is grounded on static relationships between the number of transfer/end stations and the number of links in the network. However, there are several limitations associated with this approach arising from the fact that operational characteristics (i.e. frequency and passenger flow) are not accounted for in the analysis.

In this respect, the literature highlighted that the dynamic behaviour of transit

network robustness could be measured through the cascading effect following a node/link removal. After the removal of node/link, the functional loss (the degree of the network disfunction due to a node/link removal) of a network is considered an effective measure of the dynamic-robustness of networks as well as accommodates the operational characteristics of transit networks (Guo and Su 2017; Mattsson and Jenelius 2015; Nian et al. 2019; Wang and Liu 2016; Xing et al. 2017).

It should be noted that different assessment approaches are carried out to quantify the functional loss resulting from different node/link removals. The utilization of the Giant Component is the most frequent approach in evaluating such effects (Wang et al. 2017). The Giant Component is the largest operational connected component of the network following a node/link removal. However, in bus transit networks, this approach might be limited as isolated components of the same network could provide partial independent operations. Therefore, we argue that the effect of disruptive scenarios is best evaluated as the operational losses in the network. This is estimated independently as a function of the number of non-operational nodes (frequency loss), the number of canceled trips (distance loss), and the reduction in the total trip kilometer (frequency X distance) in the entire network.

Some research gaps could be depicted in the literature review. First, there are contradictory results on the identification of the network type (e.g. small world or scale-free) associated with bus transit network. Second, and although there are several static-based robustness measures, these measures are not compared in the context of bus transit networks despite their frequent utilization. Third, previous studies focused on a small

sample of bus networks that share almost the same attributes (e.g. geography and operation), while the significant variation between bus transit networks is not considered in previous work.

Therefore, the present study stands to provide several contributions. First, we quantify the topological profile, static-robustness, and dynamic-robustness of a wide range of bus transit networks spanning from simple fixed-route hub-and-spoke networks to complex interlining networks using a dataset of 40 bus transit networks. Second, we compare a number of static-robustness measures proposed in the literature in the context of the bus transit network. Third, we analyse the dynamic-robustness of transit network following a node/link removal on bus transit networks based on operational frequency, travel distance, and the total trip kilometers. Lastly, we provide a statistical-based classification of bus transit networks based on their corresponding robustness indices. Overall, we debate that the study provides original contributions to the design and operation of bus transit networks through assessing transit service at component- and network-levels.

2.3. Methodology

2.3.1. Data Processing and Measures

The data processing of bus transit networks for CNT applications is based on the works of Derrible and Kennedy (2010) and Sun et al. (2018). Each bus route is modeled as a link that passes through numerous bus stations modeled as nodes. As illustrated in Figure 2.1, stations are classified as follows: transfer stations (N_{Tr}) that represent interchange stations between routes, start/end stations (N_E) in the bus route; and the total number of stations

(N_T) which include intermediate stations that are neither a transfer nor a start/end. It should

be noted that transfer/end stations (N_{TE}) are included in the N_T.
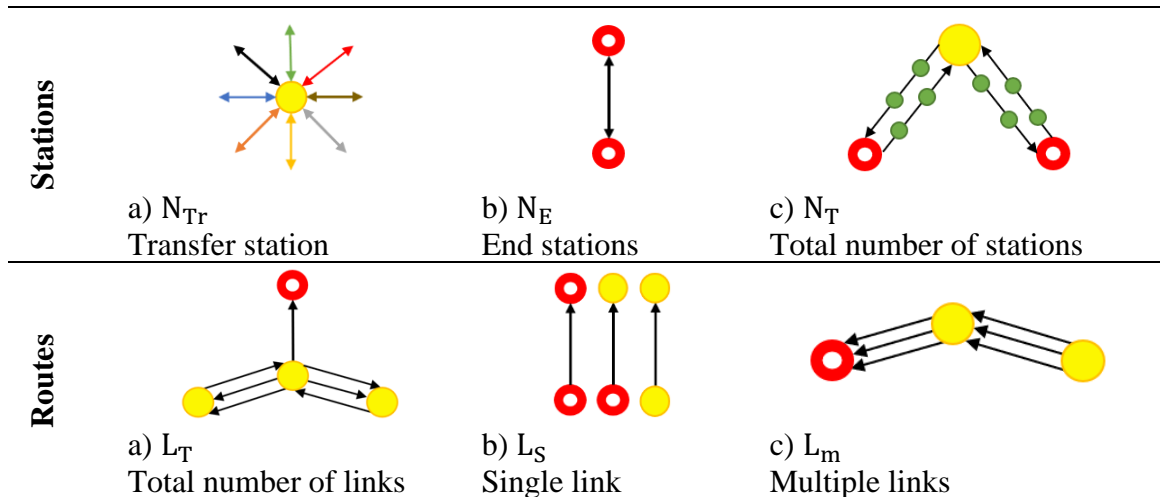
Figure 2-1. Schematic representation of bus transit data for CNT models.
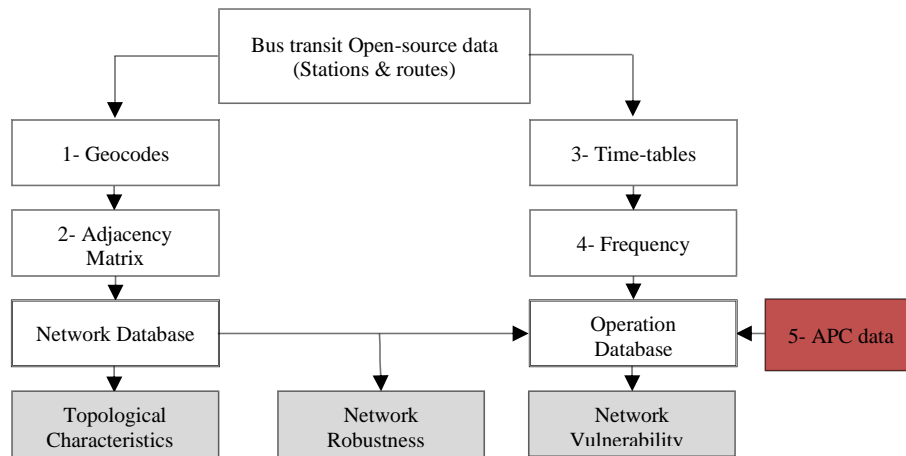
Figure 2-2. Data processing flow chart.

Similarly, as shown in Figure 2.1, transit routes are classified into three categories.

The total number of links (L_T), a single link (L_S) between two stations that represents a

single bus route, and multiple links (L_m) between two stations that represent multiple bus

routes on the same link.

Two datasets were created following a five-step process detailed in Figure 2.2. First, we extracted an open-source bus transit data that includes identifiers for stations and routes. Stations are geocoded, and routes are mapped to extract the travel distance (Manhattan distance) between stations. Together, these formed the adjacency matrix developed in step 2, which represents a bus transit network database. Operational time-tables are used to estimate the frequency of service for each bus route in steps 3 and 4, which are integrated with the Automatic Passenger Count (APC) data from step 5. Together, these steps represent a bus transit operation database. It should be noted that APC data was not available for almost all networks and was therefore excluded from the analysis.

In order to account for the unique characteristics of bus transit networks, the travel distance between stations and the route frequency are incorporated in the analysis. First, the travel distance between each pair of stations is utilized to estimate the shortest path in the network. Second, the route frequency is used to inform the weight of each link in the networks. Both metrics are jointly utilized to estimate the total trip kilometers for each network that are subsequently used to estimate the robustness of transit networks, as detailed in the following subsections. It should be noted that the estimation of the topological measures was carried out without considering the link weight or transit frequency.

2.3.1.1. Topological Measures

Although several topological measures exist in the literature (Sun et al. 2016), four topological measures are frequently utilized for the application of CNT within the transit context. Each measure represents a specific feature of the transit network as follows:

The Network Average Degree (Eq. 2.1), as highlighted in (Albert and Barabasi 2002), is the average number of links connecting all nodes. This measure could be used to estimate the number of dominant stations in the transit network when the degree of a station is higher than the Average Degree ($K_i > K_{av}$).

Average Degree $$K_{av} = \frac{\sum_{i=1}^{N} K_i}{N_{TE}}$$ Eq. 2.1

The Betweenness Centrality of a node (Eq. 2.2) is calculated based on the model developed by Freeman (1978). It is evaluated as a ratio between the number of shortest paths that pass through the node to the total number of shortest paths in the entire network. This measure could be seen as an important indicator of a station, which falls on the shortest paths in the network, whereas the global Betweenness refers to the average for the entire network.

Betweenness $$C_B(i) = \sum_{k \neq i \neq j} \frac{\sigma_{ij}(i)}{\sigma_{kj}}$$ Eq. 2.2

The Clustering Coefficient measure (Eq. 2.3) was introduced by Watts and Strogatz (1998) to represent the degree to which a node in the network tends to cluster with other nodes. Therefore, it provides an indication of the possibility of using transfer stations and alternative routes between bus stations. The Global Clustering Coefficient refers to the average for each network.

Clustering $$C(i) = \frac{2L_i}{K_i(K_i-1)}$$ Eq. 2.3

The Average Path Length (Eq. 2.4) is the average value of the minimum number of steps between each pair of nodes in the same network (Seoane et al. 2013). This measure

provides a clear indication of the reachability of transit stations in the network.

$$\text{Average Path Length} \quad PL_{av} = \frac{1}{N_{TE}(N_{TE}-1)} \sum_{i,j=1,N;i \neq j} d_{ij} \qquad \text{Eq. 2.4}$$

2.3.1.2. Robustness Measures

With respect to robustness measures, three measures are frequently utilized in the literature as follows. The Robustness Indicator (R^t) (Eq. 2.5) developed by Derrible and Kennedy (2010) is based on the number of alternative routes between stations, as an indication of the network integrity.

$$\text{Robustness Indicator} \quad R^t = \frac{L_T - N_{TE} - L_m + 1}{N_{Total}} \qquad \text{Eq. 2.5}$$

The Robustness Metric (r^T) (Eq. 2.6) was proposed by Wang et al. (2017) based on the work of Derrible and Kennedy (2010). This measure does not consider the number of multiple routes operating on the same link in the estimation of network robustness.

$$\text{Robustness Metric} \quad r^T = \frac{ln\,(L_T - N_{TE} + 2)}{N_{Total}} \qquad \text{Eq. 2.6}$$

The Critical Threshold (f_c) (Barabási 2016) refers to the fraction of non-operational stations that yields a completely disconnected network (Eq. 2.7).

$$\text{Critical Threshold} \quad f_c = 1 - \frac{1}{\frac{K_{av.}^2}{K_{av.}} - 1} \qquad \text{Eq. 2.7}$$

2.3.1.3. Dynamic-Robustness Assessment

The dynamic-robustness of bus transit networks is often estimated based on two strategies of node/link removal. Random removal strategy, based on removing nodes/links randomly and targeted removal strategy, based on removing nodes with specific attributes. After each removal, the non-operational portion of the network is assessed and compared against the

original network to provide a quantified measure of the network robustness. In this study, we utilize targeted removal. This is attributed to the fact that transit networks, in general, are more sensitive to targeted than to random attacks as argued by (Duan et al. 2009; Sun, Zhao, and Lu 2015; Shanmukhappa, Ho, and Tse 2018; Sun et al. 2018).

The present study employs three approaches for targeted component removal, as illustrated in Figure 2.3. First, we removed nodes with the highest Degree (Figure 2.3.b). Second, nodes with the highest Betweenness (Figure 2.3.c) are removed. Third, we targeted links with the highest link weight (Figure 2.3.d).

The removal process was stratified and starts with the removal of 10% of the component with the highest attribute (Degree, Betweenness, or link weight). This is followed by targeting the component with the highest 30%, 50%, 70%, and 90% in sequential steps. After each removal stage, the robustness index of the network is estimated.

Three robustness indices are developed in this study. First, a distance robustness index represents the ratio between the operational distance after a node/link removal to the total distance in the entire network. Second, an index for service frequency, which is the ratio of the total frequency of buses after node/link removal to the total frequency in the network. Third, a robustness index for trip kilometers, which is the ratio of the operational trip kilometers after node/link removal to the total operational trip kilometers in the entire network.

It should be noted that the estimation of the network operational component does not follow the concept of the Giant Component. In this study, an operational component is

defined as the sum of all operational nodes and links in the network. This is because bus transit networks will always be able to partially operate on the residual components after node(s)/link(s) removal. This could be easily depicted in Figure 2.3.c, whereby the two disconnected components are able to provide partial services. Overall, the integration of the three removal strategies and the three robustness indices resulted in nine new measures for each bus transit network.

a)   Hypothetical bus transit network

b) Removal of node (highest Degree)

c) Removal of node (highest Betweenness)

d) Removal of highest link weight
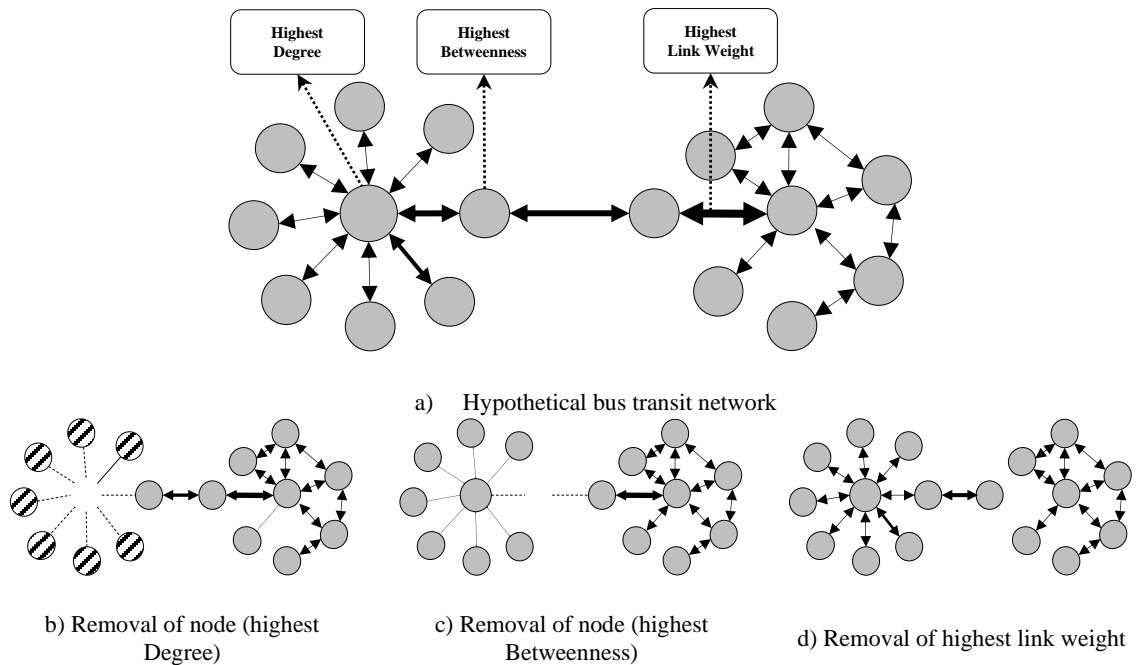
Figure 2-3. An example of the removal strategies applied in the study

2.3.2. Application for Canadian Bus Transit Networks

Transit service in Canada has delivered a total of 3.05 billion passenger trips in 2015 (CUTA 2016). The primary mode of the mobility of a significant portion (62.3%) of these trips is bus transit. While all other transit systems (e.g. ferry, streetcars, light rail, and heavy rail) provided the remaining 37.7% (CUTA 2016). This significant utilization of bus transit,

relative to all other transit systems, is achieved through 102 bus transit networks operating approximately 15,500 buses on 2,300 routes throughout Canadian cities (Mohamed et al. 2017).

The Canadian Urban Transit Association (CUTA) classifies bus transit services into five groups based on service area coverage and the corresponding population size. Group 5 (42% of all networks) represents bus transit networks operating in small cities (population less than 50,000). While Group 1 (3% of all networks) represents bus transit networks for the large metropolitan areas with a population of more than 2,000,000 individuals (CUTA 2016). Other groups represent intermediate values as they relate to the population served and fleet size.

A dataset of 40 bus transit networks was collected for Groups 2 to 5 based on the CUTA classification (CUTA 2016). The data collection was based on publicly available data in May 2018. It should be noted that the total number of stops in each network was estimated based on the number of stops listed on the service timetables.

A stratified sampling approach was implemented, whereby each group was represented by at least five transit networks as follows: Group 5 (14 networks), Group 4 (13 networks), Group 3 (8 networks), and Group 2 (5 networks). That said, and due to a data access restriction, Group 1, which contains 4 transit networks, was not included in the current study. It should be noted that the selection of the transit network in each group was subject to the availability of GIS data for bus stops and routes.

Table 2-2. Descriptive statistics of Canadian bus transit networks

ID	Network	N_{Total}*	N_{TE}	L_T	L_m	L_s	Route length (Km)		Frequency (Bus Unit)	
							Min	Max	Min	Max
Group 5										
1	Belleville	37	10	25	21	4	1.5	17.2	14	31
2	Cornwall	30	6	14	10	4	1.6	7.4	27	32
3	Cranbrook	25	9	17	16	1	1.4	6.3	5	25
4	Fort St. John	17	10	25	19	6	0.6	7.5	22	23
5	Kitimat	36	12	29	21	8	0.9	64.4	1	30
6	Leduc	17	5	13	8	5	2.0	12.5	4	7
7	North Bay	23	4	16	6	10	1.8	14.6	25	37
8	Port Alberni	11	7	21	14	7	1.6	12.2	3	18
9	Squamish	18	10	38	20	18	0.5	9.9	12	21
10	Sunshine Coast	21	8	26	14	12	2.1	30.4	9	30
11	Terrace	43	13	43	27	16	0.9	64.3	1	15
12	Vernon	29	9	20	14	6	3.2	10.8	12	29
13	Welland	29	8	23	16	7	2.2	12.5	1	28
14	Whitehorse	50	18	51	31	20	1.3	11.6	3	23
15	Yellowknife	38	11	21	16	5	0.2	5.3	18	18
Mean, (St. d)		28.8 (10.9)	9.3 (3.7)	25.5 (10.9)	16.9 (6.7)	8.6 (5.6)	1.5 (0.8)	19.1 (19.3)	10.5 (9.1)	24.5 (7.8)
Group 4										
16	Barrie	57	43	96	87	9	1.1	21.6	5	35
17	Brandon	40	12	34	28	6	1.0	8.9	12	30
18	Brantford	48	12	31	28	3	1.7	24.3	25	30
19	Chilliwack	49	17	50	41	9	0.4	56.9	2	47
20	Kamloops	62	19	57	44	13	1.0	25.5	1	54
21	Kingston	52	23	88	66	22	0.4	13.7	14	70
22	Lethbridge	45	13	34	26	8	1.2	15.1	25	46
23	Milton	47	10	26	20	6	1.2	13.1	9	36
24	Nanaimo	83	14	70	41	29	0.8	36.2	1	52
25	Prince George	43	21	85	62	23	0.5	14.9	2	31
26	**Saint John**	102	32	99	73	26	0.4	22.7	2	110
27	Thunder Bay	71	20	73	52	21	0.7	18.7	3	57
28	Whistler	25	11	28	20	8	0.8	11.8	13	68
Mean, (St. d)		55.7 (20.5)	19.0 (9.5)	59.3 (27.3)	45.2 (21.5)	14.1 (8.8)	0.9 (0.4)	21.8 (12.8)	8.8 (8.6)	51.2 (22.3)

Group 3										
29	Burlington	143	28	111	79	32	0.7	15.9	2	52
30	London	241	65	217	179	38	0.3	11.7	6	89
31	Niagara Region	41	10	31	22	9	1.2	29.6	6	16
32	Oakville	179	29	130	82	48	0.8	14.7	3	45
33	Regina	94	38	147	108	39	0.7	21.0	10	44
34	Saskatoon	132	52	161	121	40	0.2	13.7	2	99
35	Victoria	223	103	449	280	169	0.4	59.4	1	123
36	Windsor	96	22	76	63	13	0.8	21.8	5	73
Mean, (St. d)		143.6 (68.1)	43.4 (29.6)	165.3 (127.5)	116.8 (80.3)	48.5 (50.5)	0.6 (0.3)	23.5 (15.6)	4.4 (3.0)	67.6 (34.9)
Group 2										
37	Brampton	410	115	407	308	99	0.2	30.4	4	131
38	Durham Region	336	85	273	237	36	0.4	80.8	2	92
39	Hamilton	318	56	152	119	33	0.4	20.9	11	135
40	Waterloo	244	65	249	193	56	0.5	20.8	6	150
Mean, (St. d)		327.0 (68.2)	80.3 (26.1)	270.3 (105.1)	214.3 (79.2)	56.0 (30.4)	0.38 (0.13)	38.2 (28.7)	5.8 (3.9)	127.0 (24.7)
Grand Mean, (St. d)		**90.1 (97.8)**	**26.4 (27.0)**	**88.9 (102.2)**	**65.8 (75.2)**	**23.1 (30.3)**	**1.0 (0.7)**	**22.8 (17.9)**	**8.2 (7.9)**	**52.1 (36.9)**

*Please note that N_T includes all stations in the networks as listed in the timetables.

Table 2.2 provides some descriptive statistics of bus transit networks in Canada. The table highlights the significant variation of the characteristics of the Canadian bus transit networks. Although descriptive data of each group seems to be consistent, some anomalies are depicted in the CUTA classification. For example, Saint John Transit network in Group 4 has 102 bus stations while the average number of stations for all bus transit networks in Group 4 is 55.70. Furthermore, the maximum bus frequency in Saint John is about 110 bus, while the average bus frequency for the entire bus transit networks in the same group is 51.20.

2.4. Results

2.4.1. Bus Transit Network Topological Profile

The probability distributions of the Network Average Degree (K_{av}) (considering all transit networks in the dataset) for the unweighted and weighted degree are plotted in Figure 2.4. Based on the weighted K_{av}, the number of frequent buses that pass between each pair of bus stations has been considered in the calculations. The probability distribution of the Network Average Degree value fits a logarithmic distribution with R^2 ranging from 0.909 to 0.949 for both unweighted and weighted K_{av}, respectively (plotted in Figure 2.4).

The results of other topological measures are presented in Table 2.3. First, the Global Betweenness Centrality C_B (G) ranges from 0 to 1, and the minimum and maximum values represent the Local Betweenness Centrality for stations. In this respect, there is an apparent variation between networks. The results show that small networks have the highest Global Betweenness Centrality value as well as the largest range for the Local

Betweenness. This indicates the existence of dominant stations in the network and that lie on the largest number of shortest paths in the network. In contrast, the results of relatively larger networks show a lower value of Global Betweenness and smaller variation on the Local Betweenness of individual stations. This indicates the existence of a wide range of shortest paths in the network.

Following the same analogy, a high Global Clustering Coefficient C(G) value indicates that stations in the network are relatively more interconnected, which provide a larger number of alternative routes to travel through between stations. In contrast, a lower value indicates weak interconnectivity of the network. The results show a significant variation between the considered networks, which does not follow any specific trend. The Average Path Length (PL$_{av}$) is affected directly by the size of the bus network, where larger networks have higher PL$_{av}$, while smaller networks possess lower PL$_{av}$ values. Although these results provide details on the topological characteristics of transit networks, they fall short in providing a tangible indication of transit network performance and robustness.
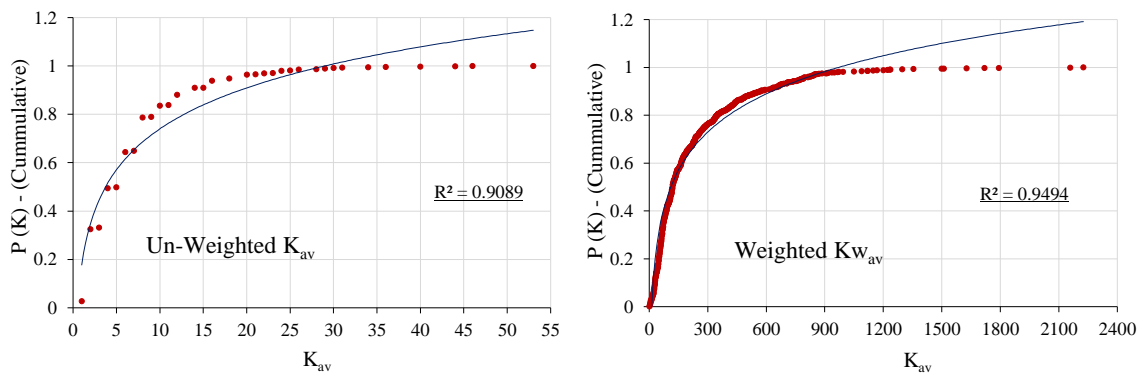
Figure 2-4. Statistical unweighted and weighted distributions for 40 bus transit networks

46

Table 2-3. Results of the topological measures

ID	Transit Network (City)	Hubs H_N	Average Degree K_{av} (max, min, St. d)	Average Weighted Degree K_{Wav} (max, min, St. d)	Betweenness $C_B(G)$(max, min, St. d)	Clustering $C(G)$(max, min, St. d)	Av. Path Length PL_{av}
1	Cornwall	1	4.67 (14, 2, 4.676)	141.00 (1111.4, 137.0, 371.215)	0.17 (1, 0, 0.408)	0.00 (0, 0, 0)	1.67
2	North Bay	1	8.00 (16, 2, 5.889)	254.00 (3581.2, 765.0, 1247.964)	0.25 (1, 0, 0.5)	0.00 (0, 0, 0)	1.50
3	Vernon	1	4.44 (16, 2, 4.447)	92.44 (1630.2, 192.0, 459.708)	0.18 (0.95, 0, 0.308)	0.00 (0, 0, 0)	2.28
4	Cranbrook	2	3.78 (14, 2, 4.055)	61.11 (727.1, 43.0, 216.344)	0.14 (0.93, 0, 0.317)	0.13 (1, 0, 0.330)	2.00
5	Leduc	2	5.20 (11, 2, 3.834)	28.80 (379.5, 81.5, 131.181)	0.27 (0.83, 0, 0.384)	0.00 (0, 0, 0)	1.80
6	Welland	2	5.75 (20, 2, 5.898)	114.75 (2214.7, 232.8, 673.236)	0.13 (0.93, 0, 0.323)	0.25 (0.75, 0, 0.371)	1.79
7	Belleville	3	5.00 (16,2, 4.643)	124.60 (1860.7,180.9, 656.700)	0.14 (0.78,0, 0.249)	0.21 (0.67,0, 0.308)	2.09
8	Port Alberni	3	6.00 (12, 2, 3.464)	73.71 (446.1, 112.5, 128.697)	0.20 (0.70, 0, 0.255)	0.31 (0.67, 0, 0.312)	1.98
9	Brantford	3	5.17 (22, 2, 2.906)	140.33 (4956.6, 228.8, 1330.434)	0.09 (0.88, 0, 0.250)	0.27 (1, 0, 0.390)	1.89
10	Milton	3	5.20 (20, 2, 5.827)	151.60 (2996, 97.2, 858.894)	0.13 (0.83, 0, 0.265)	0.07 (0.33, 0, 0.139)	2.02
11	Niagara Region	3	6.20 (16, 2, 4.158)	79.60 (2390.8, 16.8, 685.901)	0.14 (0.69, 0, 0.221)	0.13 (1, 0, 0.312)	2.12
12	Fort St. John	4	5.00 (10, 2, 2.160)	114.40 (896.0, 133.4, 224.650)	0.18 (0.55, 0, 0.154)	0.12 (0.33, 0, 0.152)	2.41
13	Kitimat	4	4.83 (12, 2, 3.129)	38.83 (863.1, 30.0, 243.317)	0.18 (0.76, 0, 0.247)	0.10 (0.33, 0, 0.208)	2.83
14	Terrace	4	6.62 (16, 2, 4.350)	42.00 (856.6, 45.5, 232.702)	0.18 (0.57, 0, 0.189)	0.16 (0.6, 0, 0.264)	2.93
15	Brandon	4	5.67 (18, 2, 4.163)	154.00 (1244.9, 216.0, 282.660)	0.12 (0.66, 0, 0.173)	0.23 (0.67, 0, 0.224)	2.18
16	Lethbridge	4	5.23 (18, 2, 5.450)	171.23 (3968.8, 93.6, 1315.673)	0.13 (0.68, 0, 0.260)	0.07 (0.5, 0, 0.276)	2.42
17	Sunshine Coast	5	6.50 (10, 0, 2.777)	136.25 (3864.6, 74.7, 1234.915)	0.30 (0.57, 0, 0.267)	0.00 (0, 0, 0)	2.82
18	Nanaimo	5	10.00 (24, 2, 7.442)	236.57 (4699.0, 230.4, 1271.420)	0.11 (0.40, 0, 0.146)	0.29 (0.83, 0, 0.346)	2.26
19	Whistler	5	5.09 (14, 2, 4.036)	170.00 (2191.0, 59.8, 745.113)	0.21 (0.76, 0, 0.276)	0.00 (0, 0, 0)	2.91
20	Squamish	6	7.60 (16, 2, 5.317)	114.00 (978.5, 56.4, 328.540)	0.21 (0.56, 0, 0.238)	0.07 (0.33, 0, 0.142)	2.69
21	Kamloops	7	6.00 (18, 2, 5.333)	170.63 (3829.6, 124.7, 974.978)	0.12 (0.54, 0, 0.165)	0.21 (1, 0, 0.315)	3.01
22	Kingston	7	7.65 (30, 2, 7.643)	299.91 (4757.2, 32.4, 1251.025)	0.08 (0.46, 0, 0.123)	0.14 (1, 0, 0.267)	2.59
23	Prince George	7	8.10 (25, 3, 5.319)	176.38 (3020.8, 214.6, 655.248)	0.10 (0.48, 0, 0.123)	0.26 (1, 0, 0.302)	2.83
24	Whitehorse	8	5.67 (10, 2, 2.497)	84.44 (691.2, 71.4, 184.298)	0.18 (0.4, 0.048, 0.115)	0.05 (0.17, 0, 0.072)	3.79
25	Yellowknife	8	3.82 (6, 2, 1.401)	68.72 (161.1, 30.6, 47.217)	0.34 (0.48, 0, 0.171)	0.18 (1, 0, 0.300)	4.07
26	Chilliwack	9	5.88 (16, 2, 4.328)	101.765 (2109.4, 39.6, 651.208)	0.12 (0.48, 0, 0.150)	0.11 (0.67, 0, 0.241)	2.84
27	Saint John	9	6.19 (23, 2, 5.019)	142.50 (1957.6, 2.8, 416.187)	0.10 (0.64, 0, 0.153)	0.07 (0.4, 0, 0.144)	3.97
28	Thunder Bay	9	7.30 (20, 2, 5.362)	204.10 (2991.3, 38.4, 914.756)	0.13 (0.63, 0, 0.154)	0.20 (0.83, 0, 0.259)	3.27
29	Oakville	9	8.97 (44, 2, 8.686)	243.10 (5529.0, 19.8, 1129.820)	0.07 (0.64, 0, 0.124)	0.16 (0.67, 0, 0.255)	2.87
30	Barrie	10	4.47 (31,1, 5.347)	124.05 (3277.3,69.9, 601.534)	0.04 (0.47,0, 0.083)	0.14 (1,0, 0.260)	3.27
31	Saskatoon	10	6.19 (53, 1, 8.745)	182.85 (7416.5, 3.0, 1261.914)	0.03 (0.55, 0, 0.087)	0.15 (1, 0, 0.286)	3.12
32	Windsor	10	6.91 (20, 2, 5.371)	270.36 (4077.9, 129.0, 1287.365)	0.09 (0.40, 0, 0.120)	0.21 (0.5, 0, 0.216)	2.76
33	Burlington	11	7.93 (30, 2, 6.981)	178.50 (5882.9, 13.6, 1250.233)	0.09 (0.48,0, 0.129)	0.22 (1, 0, 0.275)	3.01
34	Regina	14	7.74 (46, 2, 9.394)	227.21 (8424.0, 31.1, 1819.504)	0.05 (0.47, 0, 0.102)	0.17 (1, 0, 0.319)	2.82
35	London	24	6.68 (30, 1, 6.268)	273.29 (4494.1, 52.6, 1104.507)	0.04 (0.29, 0, 0.057)	0.08 (0.67, 0, 0.190)	3.73
36	Hamilton	24	5.43 (24, 2, 4.728)	247.29 (5677.5, 66.0, 1358.558)	0.06 (0.38, 0, 0.092)	0.12 (1, 0, 0.250)	5.19
37	Durham Region	25	6.43 (26, 1, 5.084)	161.91 (4023.5, 19.5, 831.484)	0.04 (0.24, 0, 0.059)	0.14 (0.67, 0, 0.185)	4.49

ID	Transit Network (City)	Hubs H_N	Average Degree K_{av} (max, min, St. d)	Average Weighted Degree K_{Wav} (max, min, St. d)	Betweenness $C_B(G)$(max, min, St. d)	Clustering $C(G)$(max, min, St. d)	Av. Path Length PL_{av}
38	Waterloo	29	7.66 (34, 1, 6.683)	332.83 (8458.5, 109.0, 1727.870)	0.04 (0.45, 0, 0.076)	0.13 (0.67, 0, 0.203)	3.98
39	Victoria	34	8.72 (40, 1, 6.954)	305.07 (13535.8, 2.6, 1821.622)	0.04 (0.36, 0, 0.062)	0.10 (0.67, 0, 0.168)	4.66
40	Brampton	55	7.08 (40, 2, 6.048)	390.42 (9175.3, 21.0, 1327.978)	0.04 (0.48, 0, 0.065)	0.10 (1, 0, 0.226)	5.37
	Grand Mean, (St. d)	**22.03 (10.90)**	**6.27 (1.44)**	**165.61 (85.71)**	**0.131 (0.076)**	**0.133 (0.85)**	**2.91 (0.95)**

These topological measures provide an opportunity to identify the network type. In general, there are three main classes of networks: scale-free, small-world, and random. Each class possesses unique characteristics with respect to the network behaviour against disruptive events. A network is identified as a scale-free if the P(K) distribution follows a power-law distribution with gamma (G) value between 2 and 3 (Eq. 2.8) (Choromański, Matuszak, and Mie¸kisz 2013; Barabási 2016). However, it is hard to attribute clear scale free properties since the power-law decay tendency of the transit networks (de Regt et al. 2019). Similarly, a random network is identified from the distribution of P(K), and if the distribution follows a Poisson distribution, the network is identified as a random network (Eq. 2.9) (Newman, Strogatz, and Watts 2001; Barabási 2016). A small-world network is identified through the Average Path Length, the total number of transfer stations, and the Average Degree of the network (Eq. 2.10) (Barabási 2016).

$$P(K) \simeq K^{-\gamma} \qquad\qquad 2 < \gamma < 3, \text{Scale-free} \qquad\qquad \text{Eq. 2.8}$$

$$PL_{av} \sim {ln\,N}/{ln\,K_{av}} \qquad\qquad PL_{av} \simeq {ln\,N}/{ln\,K_{av}}, \text{Small world} \qquad\qquad \text{Eq. 2.9}$$

$$P(K) = e^{-K_{av}} * \frac{(K_{av})^{K_i}}{K_i!} \qquad P(K) \simeq \text{Poison Distribution, Random Network} \qquad \text{Eq. 2.10}$$

The results indicate that bus transit networks in Canada do not follow any of the aforementioned network classes. For example, the results of Hamilton and Port Alberni networks indicated that G = 1.264 & 0.006, PL_{av} = 5.185 & 1.976, and the ln N/ln (K_{av})= 2.38 & 1.086 for both networks, respectively. In addition, they did not fit a Poison distribution with very low R^2 values of (0.270, 0.095, respectively). These results echo the findings of Shanmukhappa, Ho, and Tse (2018). Nonetheless, the results do not follow their

proposed super-node network structure (i.e. grouping stations within a spatial buffer), nor the works of (Sui et al. 2012; Huang et al. 2015) in their analysis of multimodal transit network.

Another approach to investigate the impacts of the topological measures on the network behaviour is through the relationships between the number of Hub stations (H) and the topological measures. However, Figure 2.5 indicates the lack of significant association between the number of Hubs (H) and the network topological measures, which is evident by the low R^2 values (based on power-law distribution). In addition, the relationship between the number of Hubs and the Clustering coefficient did not fit any distribution (Figure 2.5.b).
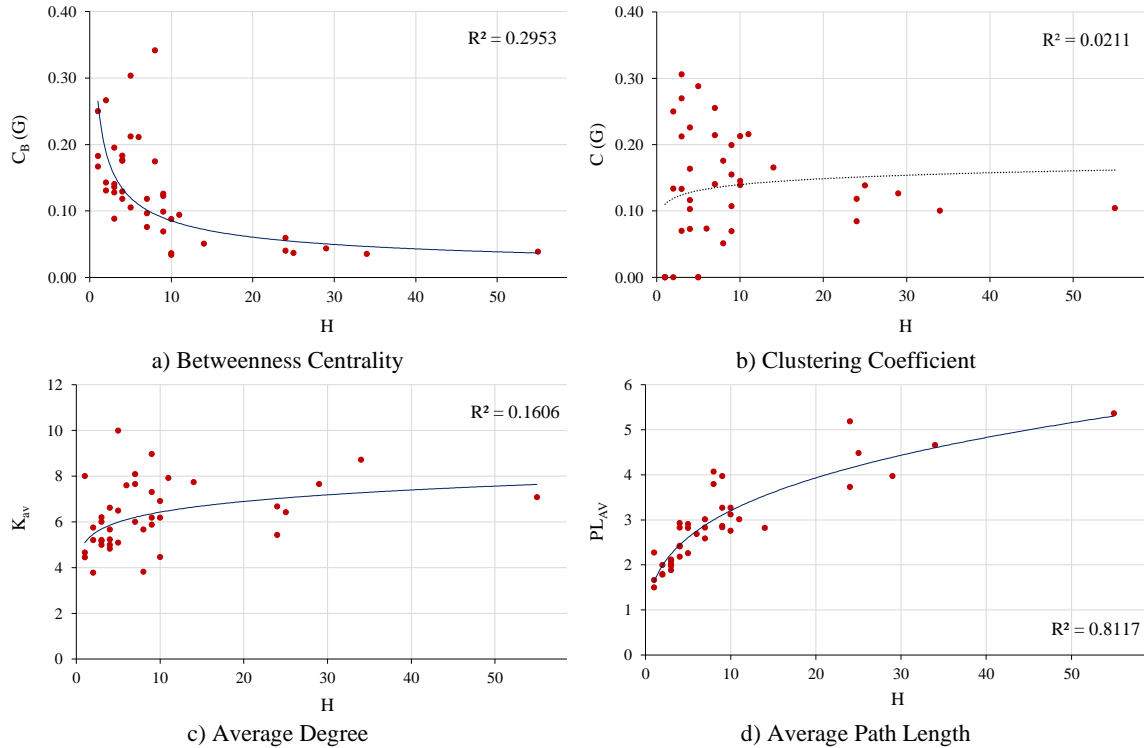
Figure 2-5. The association between the number of Hubs and the topological measures

2.4.2. Static-Robustness of Bus Transit Networks

The lack of clear indications on the network model from the topological measures motivates the investigation of network robustness. As highlighted in the literature section, three main measures are frequently utilized to provide a holistic numerical value of network robustness. The results of the three robustness measures are detailed in Table 2.4. These include the Critical Threshold (f_c), Robustness Indicator (R^t), and Robustness Metric (r^T). It should be noted that the values in Table 2.4 are not normalized between the three measures. However, all measures range between zero and one; with the latter corresponds to the maximum robustness value.

The Critical Threshold (f_c) results range between 0.919 and 0.095, with a standard deviation of 0.0238. The results indicate that, in general, transit networks with several hubs are more robust relative to networks with fewer hubs. This might be attributed to the existence of several alternative routes between origin and destinations. However, it should be noted the estimation of the critical threshold depends solely on the ratio between the Average Degree and its second moment. Therefore, networks with more hubs will be favoured in this regard. Second, for the Robustness Metric (r^T), the highest value is 0.252 while the lowest value is 0.014, with a standard deviation of 0.05. Generally, there is no apparent pattern on the relationship between bus networks and the Robustness Metric. This is due to the dependency on the difference between the numbers of links (L_T) and the number of transfer and end stations (N_{TE}). Third, the results of the Robustness Indicator (R^t) range between 0.977 and 0.130, with a standard deviation of 0.202. Since the Robustness Indicator depends on the number of multiple links (L_m) between bus stations,

there are clear indications that relatively smaller networks with fewer hubs and fewer

multiple links are reported less robust.

Table 2-4. Static-Robustness results for Canadian bus transit networks

ID	Network	H	R^t	r^T	f_c
1	**Cornwall***	1	0.167	0.077	0.681
2	North Bay	1	0.130	0.115	0.556
3	Vernon	1	0.207	0.088	0.763
4	Cranbrook	2	0.320	0.092	0.582
5	Leduc	2	0.235	0.135	0.131
6	Welland	2	0.310	0.098	0.810
7	Belleville	3	0.324	0.077	0.554
8	**Port Alberni***	3	0.727	0.252	0.212
9	Brantford	3	0.354	0.063	0.789
10	**Milton***	3	0.234	0.061	0.739
11	Niagara Region	3	0.317	0.076	0.556
12	Fort St. John	4	0.588	0.167	0.231
13	Kitimat	4	0.278	0.082	0.103
14	**Terrace***	4	0.349	0.081	0.386
15	**Brandon***	4	0.425	0.079	0.676
16	Lethbridge	4	0.311	0.070	0.583
17	**Sunshine Coast***	5	0.333	0.143	0.116
18	**Nanaimo***	5	0.337	0.049	0.707
19	**Whistler***	5	0.400	0.118	0.333
20	**Squamish***	6	0.611	0.189	0.449
21	**Kamloops***	7	0.419	0.059	0.663
22	Kingston	7	0.846	0.081	0.812
23	**Prince George***	7	0.977	0.097	0.704
24	Whitehorse	8	0.280	0.071	0.095
25	**Yellowknife***	8	0.158	0.065	0.550
26	Chilliwack	9	0.510	0.073	0.525
27	Saint John	9	0.412	0.042	0.630
28	Thunder Bay	9	0.465	0.056	0.615
29	Oakville	9	0.302	0.026	0.896
30	Barrie	10	0.789	0.070	0.864
31	Saskatoon	10	0.530	0.036	0.919
32	Windsor	10	0.438	0.042	0.499
33	Burlington	11	0.364	0.031	0.800
34	Regina	14	0.755	0.050	0.882
35	London	24	0.477	0.021	0.756
36	Hamilton	24	0.201	0.014	0.692
37	Durham Region	25	0.455	0.016	0.748
38	Waterloo	29	0.529	0.021	0.799
39	Victoria	34	0.798	0.026	0.829
40	Brampton	55	0.473	0.014	0.865
Grand Mean, (St.d)		**22.03 (10.90)**	**0.428 (0.202)**	**0.076 (0.050)**	**0.602 (0.238)**

* refers to network that have contradictory robustness values across the three measures (detailed in Figure 2.6)

The normalized robustness values of 40 Canadian bus transit networks are shown

in Figure 2.6. This illustrates the variation among robustness measures. These results show

significant discrepancies as some networks exhibit relatively high robustness values in

some measures and low values in others, as illustrated in Figure 2.6.
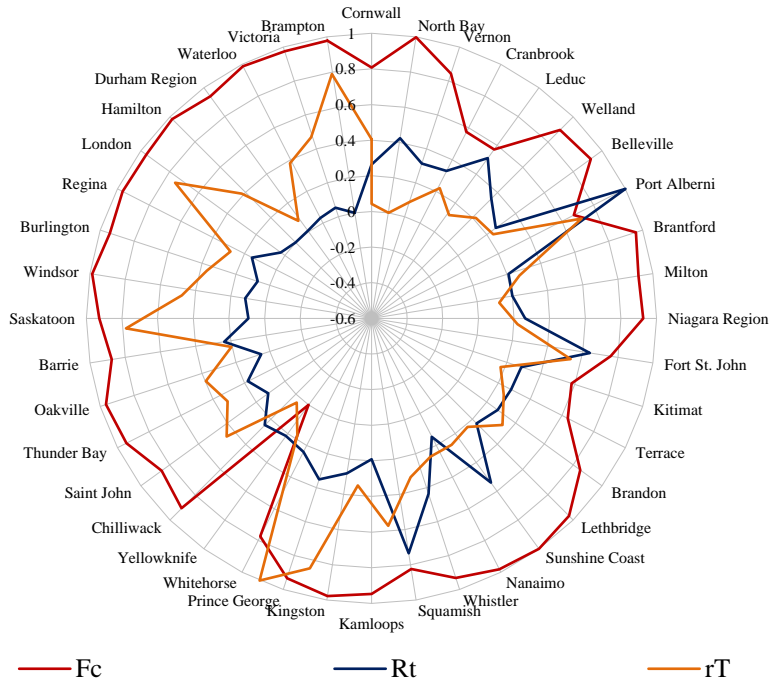
Figure 2-6. Normalized distribution of Static-Robustness measures for 40 considered transit networks

It can thus be argued that static measures of robustness fall short in providing tangible evidence on the behaviour of the transit network. Even the results of a correlation analysis between the three measures indicate either the lack of significant correlation or the existence of a negative correlation in some cases as detailed in Table 2.5.

Table 2-5. Correlation between Static-Robustness measures

	H	R^t	r^T	f_c
H	1			
R^t	0.270	1		
r^T	-0.543**	0.128	1	
f_c	0.343*	0.389*	-0.192	1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

2.4.3. Dynamic-Robustness of Bus Transit Networks

The dynamic-robustness indices of Canadian bus transit networks are listed in Tables 2.6, 2.7, and 2.8. The results are presented following the three targeted removal strategies of: nodes with the highest Degree (Table 2.6); nodes with the highest Betweenness (Table 2.7); and links with the highest link weight (Table 2.8). Each removal strategy was assessed using three indices representing the losses in the total trip kilometers, route distances, and service frequency. Results of the latter two indices are presented in Appendices 2.2–2.9.

Table 2-6. The losses in trip kilometers due to the removal of nodes with the highest Degree

						Percentage removal of nodes with the highest Degree				
ID	Network	Hubs	K_{max}/L_T	N_{TE}	N_L	10% (number of nodes)	30% (number of nodes)	50% (number of nodes)	70% (number of nodes)	90% (number of nodes)
1	Cornwall	1	100%	6	14	0.000 (1)	0.000 (2)	0.000 (3)	0.000 (4)	0.000 (5)
2	North Bay	1	100%	4	16	0.000 (1)	0.000 (1)	0.000 (2)	0.000 (3)	0.000 (4)
3	Vernon	1	80%	9	20	0.171 (1)	0.000 (3)	0.000 (5)	0.000 (6)	0.000 (8)
4	Cranbrook	2	82%	9	17	0.140 (1)	0.000 (3)	0.000 (5)	0.000 (6)	0.000 (8)
5	Leduc	2	85%	5	13	0.190 (1)	0.000 (2)	0.000 (3)	0.000 (4)	0.000 (5)
6	Welland	2	87%	8	23	0.060 (1)	0.001 (2)	0.000 (4)	0.000 (6)	0.000 (7)
7	Niagara Region	3	64%	10	25	0.345 (1)	0.091 (3)	0.000 (5)	0.000 (7)	0.000 (9)
8	Brantford	3	57%	7	21	0.205 (1)	0.000 (4)	0.000 (6)	0.000 (8)	0.000 (11)
9	Milton	3	71%	12	31	0.128 (1)	0.000 (3)	0.000 (5)	0.000 (7)	0.000 (9)
10	Belleville	3	77%	10	26	0.474 (1)	0.019 (3)	0.000 (5)	0.000 (7)	0.000 (9)
11	Port Alberni	3	52%	10	31	0.539 (1)	0.098 (2)	0.015 (4)	0.000 (5)	0.000 (6)
12	Brandon	4	40%	10	25	0.539 (1)	0.264 (4)	0.114 (6)	0.028 (8)	0.000 (11)
13	Lethbridge	4	41%	12	29	0.648 (1)	0.000 (4)	0.000 (7)	0.000 (9)	0.000 (12)
14	Fort St. John	4	37%	13	43	0.420 (1)	0.211 (3)	0.108 (5)	0.000 (7)	0.000 (9)
15	Kitimat	4	53%	12	34	0.247 (1)	0.066 (4)	0.020 (6)	0.000 (8)	0.000 (11)
16	Terrace	4	53%	13	34	0.390 (1)	0.268 (4)	0.093 (7)	0.010 (9)	0.000 (12)
17	Nanaimo	5	38%	8	26	0.496 (1)	0.103 (4)	0.048 (7)	0.000 (10)	0.000 (13)
18	Whistler	5	34%	14	70	0.487 (1)	0.130 (3)	0.000 (6)	0.000 (8)	0.000 (10)
19	Sunshine Coast	5	50%	11	28	0.679 (1)	0.559 (2)	0.300 (4)	0.000 (6)	0.000 (7)
20	Squamish	6	42%	10	38	0.736 (1)	0.516 (3)	0.000 (5)	0.000 (7)	0.000 (9)
21	Kamloops	7	32%	19	57	0.519 (2)	0.087 (6)	0.000 (10)	0.000 (13)	0.000 (17)
22	Kingston	7	34%	23	88	0.404 (2)	0.109 (7)	0.000 (12)	0.000 (16)	0.000 (21)
23	Prince George	7	29%	21	85	0.554 (2)	0.165 (6)	0.002 (11)	0.000 (15)	0.000 (19)
24	Whitehorse	8	20%	18	51	0.697 (2)	0.264 (5)	0.104 (9)	0.066 (13)	0.000 (16)
25	Yellowknife	8	29%	11	21	0.746 (1)	0.277 (3)	0.166 (6)	0.000 (8)	0.000 (10)
26	Oakville	9	32%	17	50	0.424 (3)	0.148 (9)	0.004 (15)	0.000 (20)	0.000 (26)
27	Chilliwack	9	23%	32	99	0.651 (2)	0.473 (5)	0.013 (9)	0.000 (12)	0.000 (15)
28	Saint John	9	27%	20	73	0.502 (3)	0.127 (10)	0.090 (16)	0.000 (22)	0.000 (29)
29	Thunder Bay	9	34%	29	130	0.385 (2)	0.062 (6)	0.002 (10)	0.000 (14)	0.000 (18)
30	Saskatoon	10	32%	43	96	0.254 (5)	0.021 (16)	0.009 (26)	0.000 (36)	0.000 (47)
31	Windsor	10	33%	52	161	0.582 (2)	0.123 (7)	0.023 (11)	0.000 (15)	0.000 (20)
32	Barrie	10	26%	22	76	0.397 (4)	0.082 (13)	0.030 (22)	0.000 (30)	0.000 (39)

33	Burlington	11	27%	28	111	0.287 (3)	0.159 (8)	0.106 (14)	0.000 (20)	0.000 (25)
34	Regina	14	31%	38	147	0.193 (4)	0.003 (11)	0.000 (19)	0.000 (27)	0.000 (34)
35	Hamilton	24	14%	65	217	0.543 (6)	0.076 (17)	0.042 (28)	0.037 (39)	0.037 (50)
36	London	24	16%	56	152	0.446 (7)	0.097 (20)	0.014 (33)	0.000 (46)	0.000 (59)
37	Durham Region	25	10%	85	273	0.432 (9)	0.164 (26)	0.053 (43)	0.013 (60)	0.000 (77)
38	Waterloo	29	14%	65	249	0.403 (7)	0.054 (20)	0.009 (33)	0.002 (46)	0.000 (59)
39	Victoria	34	9%	103	449	0.419 (10)	0.167 (31)	0.051 (52)	0.001 (72)	0.000 (93)
40	Brampton	35	10%	115	407	0.540 (12)	0.235 (35)	0.097 (58)	0.000 (81)	0.000 (104)

Table 2-7. The losses in trip kilometre due to the removal of nodes with the highest Betweenness

| ID | Network | Hubs | K_{max}/L_T | N_{TE} | N_L | Percentage removal of nodes with the highest Betweenness | | | | |
						10% (number of nodes)	30% (number of nodes)	50% (number of nodes)	70% (number of nodes)	90% (number of nodes)
1	Cornwall	1	100%	6	14	0.000 (1)	0.000 (2)	0.000 (3)	0.000 (4)	0.000 (5)
2	North Bay	1	100%	4	16	0.000 (1)	0.000 (1)	0.000 (2)	0.000 (3)	0.000 (4)
3	Vernon	1	80%	9	20	0.171 (1)	0.000 (3)	0.000 (5)	0.000 (6)	0.000 (8)
4	Cranbrook	2	82%	9	17	0.140 (1)	0.000 (3)	0.000 (5)	0.000 (6)	0.000 (8)
5	Leduc	2	85%	5	13	0.190 (1)	0.000 (2)	0.000 (3)	0.000 (4)	0.000 (5)
6	Welland	2	87%	8	23	0.060 (1)	0.001 (2)	0.000 (4)	0.000 (6)	0.000 (7)
7	Niagara Region	3	64%	10	25	0.345 (1)	0.106 (3)	0.000 (5)	0.000 (7)	0.000 (9)
8	Brantford	3	57%	7	21	0.205 (1)	0.000 (4)	0.000 (6)	0.000 (8)	0.000 (11)
9	Milton	3	71%	12	31	0.128 (1)	0.000 (3)	0.000 (5)	0.000 (7)	0.000 (9)
10	Belleville	3	77%	10	26	0.474 (1)	0.019 (3)	0.000 (5)	0.000 (7)	0.000 (9)
11	Port Alberni	3	52%	10	31	0.539 (1)	0.098 (2)	0.000 (4)	0.000 (5)	0.000 (6)
12	Brandon	4	40%	10	25	0.539 (1)	0.143 (4)	0.114 (6)	0.028 (8)	0.000 (11)
13	Lethbridge	4	41%	12	29	0.648 (1)	0.000 (4)	0.000 (7)	0.000 (9)	0.000 (12)
14	Fort St. John	4	37%	13	43	0.420 (1)	0.211 (3)	0.108 (5)	0.000 (7)	0.000 (9)
15	Kitimat	4	53%	12	34	0.247 (1)	0.066 (4)	0.020 (6)	0.000 (8)	0.000 (11)
16	Terrace	4	53%	13	34	0.825 (1)	0.231 (4)	0.024 (7)	0.000 (9)	0.000 (12)
17	Nanaimo	5	38%	8	26	0.496 (1)	0.091 (4)	0.005 (7)	0.000 (10)	0.000 (13)
18	Whistler	5	34%	14	70	0.487 (1)	0.130 (3)	0.000 (6)	0.000 (8)	0.000 (10)
19	Sunshine Coast	5	50%	11	28	0.636 (1)	0.587 (2)	0.300 (4)	0.000 (6)	0.000 (7)
20	Squamish	6	42%	10	38	0.779 (1)	0.516 (3)	0.000 (5)	0.000 (7)	0.000 (9)
21	Kamloops	7	32%	19	57	0.519 (2)	0.087 (6)	0.000 (10)	0.000 (13)	0.000 (17)
22	Kingston	7	34%	23	88	0.404 (2)	0.085 (7)	0.000 (12)	0.000 (16)	0.000 (21)
23	Prince George	7	29%	21	85	0.398 (2)	0.146 (6)	0.002 (11)	0.000 (15)	0.000 (19)
24	Whitehorse	8	20%	18	51	0.697 (2)	0.264 (5)	0.104 (9)	0.066 (13)	0.040 (16)
25	Yellowknife	8	29%	11	21	0.746 (1)	0.669 (3)	0.283 (6)	0.000 (8)	0.000 (10)
26	Oakville	9	32%	17	50	0.424 (3)	0.148 (9)	0.004 (15)	0.000 (20)	0.000 (26)
27	Chilliwack	9	23%	32	99	0.595 (2)	0.411 (5)	0.000 (9)	0.000 (12)	0.000 (15)
28	Saint John	9	27%	20	73	0.588 (3)	0.219 (10)	0.032 (16)	0.000 (22)	0.000 (29)
29	Thunder Bay	9	34%	29	130	0.422 (2)	0.112 (6)	0.002 (10)	0.000 (14)	0.000 (18)
30	Saskatoon	10	32%	43	96	0.196 (5)	0.021 (16)	0.000 (26)	0.000 (36)	0.000 (47)
31	Windsor	10	33%	52	161	0.507 (2)	0.072 (7)	0.009 (11)	0.000 (15)	0.000 (20)
32	Barrie	10	26%	22	76	0.391 (4)	0.116 (13)	0.005 (22)	0.000 (30)	0.000 (39)
33	Burlington	11	27%	28	111	0.287 (3)	0.194 (8)	0.119 (14)	0.066 (20)	0.000 (25)
34	Regina	14	31%	38	147	0.210 (4)	0.041 (11)	0.000 (19)	0.000 (27)	0.000 (34)
35	Hamilton	24	14%	65	217	0.669 (6)	0.156 (17)	0.075 (28)	0.037 (39)	0.037 (50)
36	London	24	16%	56	152	0.446 (7)	0.097 (20)	0.014 (33)	0.000 (46)	0.000 (59)
37	Durham Region	25	10%	85	273	0.479 (9)	0.146 (26)	0.048 (43)	0.013 (60)	0.000 (77)
38	Waterloo	29	14%	65	249	0.450 (7)	0.107 (20)	0.005 (33)	0.002 (46)	0.000 (59)
39	Victoria	34	9%	103	449	0.414 (10)	0.164 (31)	0.070 (52)	0.005 (72)	0.000 (93)
40	Brampton	35	10%	115	407	0.540 (12)	0.235 (35)	0.099 (58)	0.002 (81)	0.000 (104)

Table 2-8. The losses in trip kilometre due to the removal of links with the highest Link Weight

ID	Network	Hubs	K_{max}/L_T	N_{TE}	N_L	Percentage removal of links with the highest Link Weight				
						10% (number of nodes)	30% (number of nodes)	50% (number of nodes)	70% (number of nodes)	90% (number of nodes)
1	Cornwall	1	100%	6	14	0.787 (1)	0.524 (4)	0.350 (7)	0.182 (10)	0.082 (13)
2	North Bay	1	100%	4	16	0.769 (2)	0.480 (5)	0.252 (8)	0.097 (11)	0.016 (14)
3	Vernon	1	80%	9	20	0.825 (2)	0.569 (6)	0.360 (10)	0.178 (14)	0.046 (18)
4	Cranbrook	2	82%	9	17	0.673 (2)	0.440 (5)	0.276 (9)	0.164 (12)	0.047 (15)
5	Leduc	2	85%	5	13	0.840 (1)	0.445 (4)	0.246 (7)	0.095 (9)	0.026 (12)
6	Welland	2	87%	8	23	0.794 (2)	0.460 (7)	0.194 (12)	0.014 (16)	0.002 (21)
7	Niagara Region	3	64%	10	25	0.714 (3)	0.331 (9)	0.160 (16)	0.067 (22)	0.018 (28)
8	Brantford	3	57%	7	21	0.714 (3)	0.401 (9)	0.236 (16)	0.124 (22)	0.033 (28)
9	Milton	3	71%	12	31	0.770 (3)	0.512 (8)	0.291 (13)	0.151 (18)	0.028 (23)
10	Belleville	3	77%	10	26	0.661 (3)	0.409 (8)	0.225 (13)	0.099 (18)	0.008 (23)
11	Port Alberni	3	52%	10	31	0.718 (2)	0.421 (6)	0.221 (11)	0.072 (15)	0.019 (19)
12	Brandon	4	40%	10	25	0.794 (3)	0.487 (10)	0.275 (17)	0.125 (24)	0.040 (31)
13	Lethbridge	4	41%	12	29	0.836 (3)	0.544 (10)	0.313 (17)	0.129 (24)	0.023 (31)
14	Fort St. John	4	37%	13	43	0.687 (3)	0.391 (8)	0.235 (13)	0.104 (18)	0.009 (23)
15	Kitimat	4	53%	12	34	0.595 (3)	0.275 (9)	0.054 (15)	0.016 (20)	0.003 (26)
16	Terrace	4	53%	13	34	0.544 (4)	0.234 (13)	0.097 (22)	0.027 (30)	0.006 (39)
17	Nanaimo	5	38%	8	26	0.660 (7)	0.326 (21)	0.142 (35)	0.043 (49)	0.006 (63)
18	Whistler	5	34%	14	70	0.601 (3)	0.250 (8)	0.110 (14)	0.047 (20)	0.005 (25)
19	Sunshine Coast	5	50%	11	28	0.759 (3)	0.425 (8)	0.206 (13)	0.072 (18)	0.008 (23)
20	Squamish	6	42%	10	38	0.671 (4)	0.371 (11)	0.217 (19)	0.089 (27)	0.006 (34)
21	Kamloops	7	32%	19	57	0.704 (6)	0.396 (17)	0.202 (29)	0.099 (40)	0.021 (51)
22	Kingston	7	34%	23	88	0.782 (9)	0.488 (26)	0.287 (44)	0.121 (62)	0.012 (79)
23	Prince George	7	29%	21	85	0.712 (9)	0.417 (26)	0.225 (43)	0.083 (60)	0.007 (77)
24	Whitehorse	8	20%	18	51	0.650 (5)	0.360 (15)	0.225 (26)	0.116 (36)	0.034 (46)
25	Yellowknife	8	29%	11	21	0.680 (2)	0.360 (6)	0.179 (11)	0.096 (15)	0.027 (19)
26	Oakville	9	32%	17	50	0.689 (13)	0.330 (39)	0.143 (65)	0.046 (91)	0.007 (117)
27	Chilliwack	9	23%	32	99	0.518 (5)	0.281 (15)	0.116 (25)	0.032 (35)	0.007 (45)
28	Saint John	9	27%	20	73	0.667 (10)	0.386 (30)	0.204 (50)	0.073 (69)	0.009 (89)
29	Thunder Bay	9	34%	29	130	0.732 (7)	0.429 (22)	0.218 (37)	0.067 (51)	0.011 (66)
30	Saskatoon	10	32%	43	96	0.697 (16)	0.401 (48)	0.213 (81)	0.086 (113)	0.010 (145)
31	Windsor	10	33%	52	161	0.724 (8)	0.412 (23)	0.204 (38)	0.077 (53)	0.010 (68)
32	Barrie	10	26%	22	76	0.738 (10)	0.454 (29)	0.270 (48)	0.133 (67)	0.027 (86)
33	Burlington	11	27%	28	111	0.616 (11)	0.305 (33)	0.140 (56)	0.045 (78)	0.006 (100)
34	Regina	14	31%	38	147	0.677 (15)	0.377 (44)	0.195 (74)	0.082 (103)	0.015 (132)
35	Hamilton	24	14%	65	217	0.676 (15)	0.344 (46)	0.184 (76)	0.082 (106)	0.016 (137)
36	London	24	16%	56	152	0.731 (22)	0.413 (65)	0.205 (109)	0.077 (152)	0.014 (195)
37	Durham Region	25	10%	85	273	0.675 (27)	0.357 (82)	0.170 (137)	0.056 (191)	0.006 (246)
38	Waterloo	29	14%	65	249	0.709 (25)	0.406 (75)	0.219 (125)	0.097 (174)	0.020 (224)
39	Victoria	34	9%	103	449	0.598 (45)	0.286 (135)	0.123 (225)	0.038 (314)	0.004 (404)
40	Brampton	35	10%	115	407	0.645 (41)	0.370 (122)	0.196 (204)	0.079 (285)	0.012 (366)

Results of the removal of nodes with the highest Degree and the highest Betweenness and their impacts on the total lose trip kilometers are also illustrated in Figures 2.7 and 2.8, respectively. The results demonstrate the significant impact of removing merely 10% of nodes with the highest Degree and/or Betweenness (x-axis). Almost all

considered transit networks exhibit a significant loss in operation. Some networks even experience complete loss of operation following the removal of only 10% of nodes. The results highlight that, in general, removing 50% of nodes (highest Betweenness or Degree) would result in cancelling more than 85% of the total trip kilometers for all transit networks with the exception of Squamish, Yellowknife, and Whitehorse transit networks.

The probability of node removal in transit networks does not represent a real-world scenario. Therefore, the behaviour of transit networks could be better depicted through the removal of links with the highest weights, which represents the case of congestion in the road network.

In this respect, Figure 2.9 illustrates the results of the 40 considered transit networks. The results demonstrate a distinct relationship between the proportional removal of links and the associated losses in the total trip kilometers. It is clear that all networks lose trip kilometers in a slower, yet consistent, rerates relative to the effect of node removal. This relationship is further examined through a linear regression model. The regression model indicates a significant relationship with R^2 =0.977 and the losses in the total trip kilometer are estimated as (-0.989 (% of link removed) + 0.92) as depicted in Figure 2.9. However, the results of the linear regression model should be carefully interpreted due to the shape of the relationship.
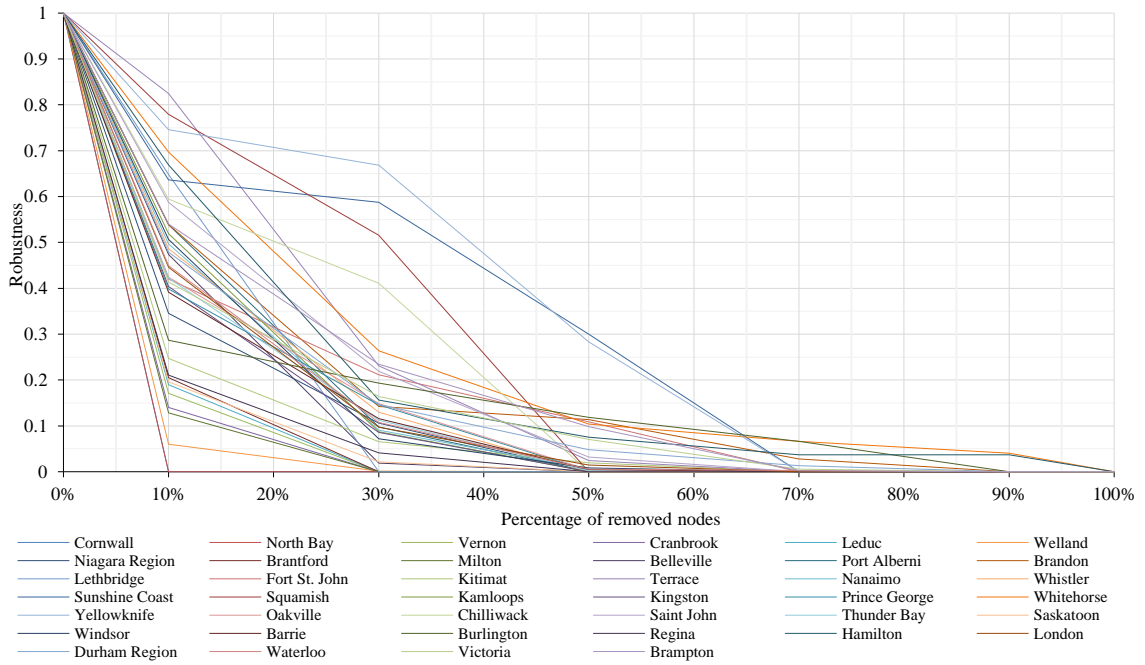
Figure 2-7. The robustness index for trip kilometre due to the removal of nodes with the highest Degree Centrality
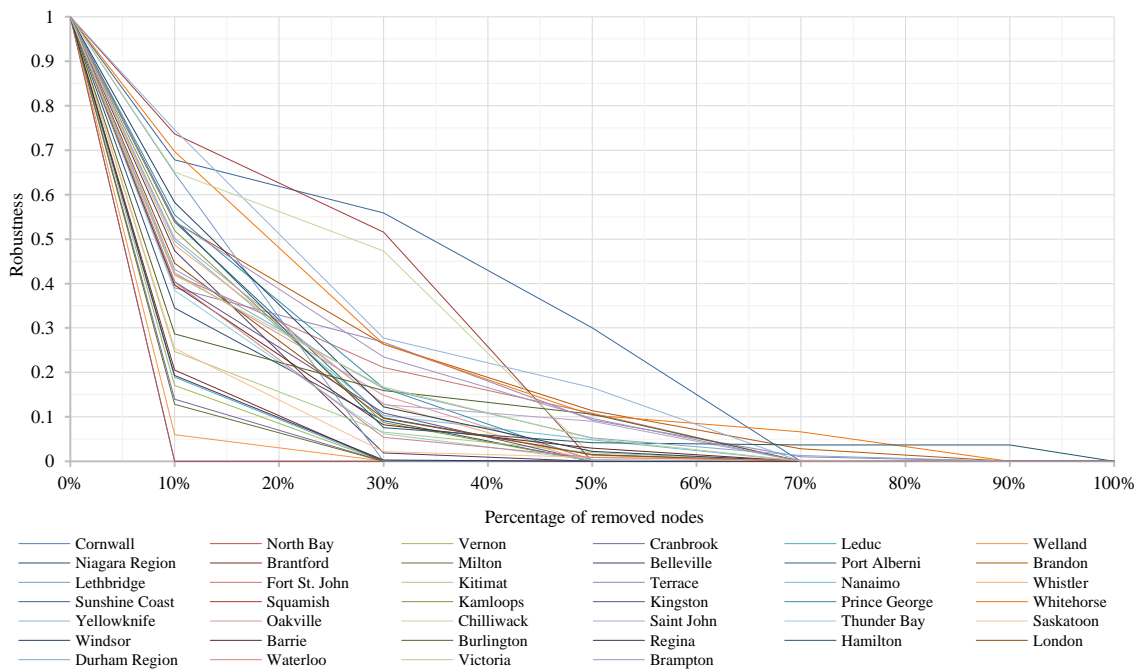
Figure 2-8. The robustness index for trip kilometre due to the removal of nodes with the highest Betweenness Centrality
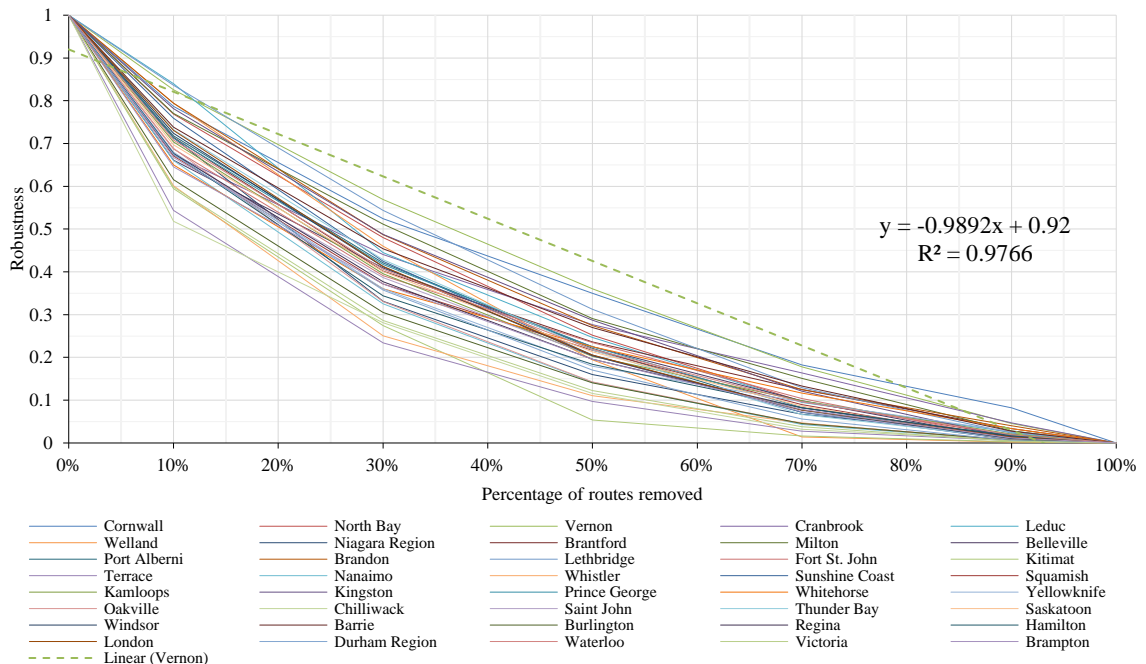
Figure 2-9. The robustness index for trip kilometre due to the removal of links with the highest weight

Overall, the dynamic-robustness measures offer various advantages compared to the static-based measures, in particular in quantifying the cascading impacts of different disruptive scenarios. In this respect, the cascading impacts of nodes removal have more bearing on the network behaviour relative to the removal of links. Some might argue that such results are intuitive. However, the cascading service interruption behaviour of transit network should be assessed based on the joint impact of the removal of nodes and links.

Therefore, we developed a statistical classification model, namely a two-step cluster analysis to classify transit networks based on the combination of network structure and the results of the developed nine robustness metrics. Two-step cluster analysis is performed using IBM SPSS 21. The analysis combines hierarchical and non-hierarchical clustering algorithms in a single output model. First, a distance measure is applied to establish a

hierarchical cluster solution, which is further refined using agglomerative hierarchical cluster method that maximizes the log-likelihood. The number of clusters is determined based on the change in the Bayesian Inference Criterion (BIC) (Mohamed et al. 2016). This method is often recommended to overcome the limitations of performing two separate clustering models (i.e. hierarchical and non-hierarchical).

All 40 networks were clustered based on their operational characteristics (9 variables) and the Robustness Index for total trip kilometers resulting from the removal of 10%, 30%, 50%, 70%, and 90% of links with the highest link weight (5 variables referred to as LW 0.1 – LW 0.9), and 10%, 30%, and 50% of nodes with the highest betweenness and degree (6 variables referred to as HB 0.1 – HB 0.5 and HD 0.1 – HD 0.5). The impacts of removing 70% and 90% of the nodes with the highest betweenness and degree were eliminated due to the high frequency of zero values that might impact the clustering model. The results indicate that the data fits three distinct clusters informed by the change in the BIC value, as illustrated in Figure 2.10. Each cluster explains a distinct profile of bus transit networks as follows. It should be noted though that the data in Figure 2.10 represents the standardized values of the 20 variables. The mean values of cluster centers are detailed in Table 2.9.
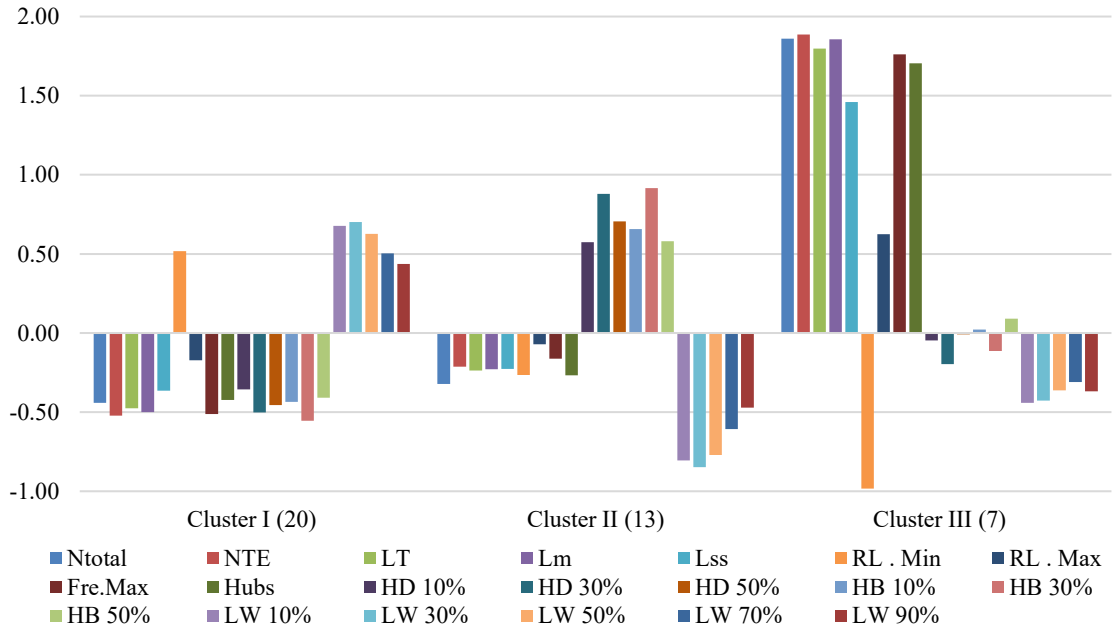
Figure 2-10. Profiling bus transit network based on cascading behaviour and topological measures

Table 2-9. Cluster centres (mean values)

Variables	Cluster I (20 Networks)	Cluster II (13 Networks)	Cluster III (7 Networks)
N_{Total}	46.95	58.62	272.00
N_{TE}	12.30	20.62	77.29
L_T	40.30	64.77	272.57
L_m	28.20	48.54	205.29
L_S	12.10	16.23	67.29
$RL._{Min}$	1.33	0.82	0.34
$RL._{Max}$	19.70	21.49	33.96
$Fre._{Max}$	33.20	46.08	117.00
Hubs	5.30	6.69	24.43
HD 10%	0.34	0.52	0.40
HD 30%	0.06	0.25	0.10
HD 50%	0.01	0.08	0.04
HB 10%	0.33	0.56	0.42
HB 30%	0.05	0.29	0.12
HB 50%	0.01	0.08	0.04
LW 10%	0.75	0.64	0.67
LW 30%	0.45	0.33	0.36
LW 50%	0.25	0.16	0.19
LW 70%	0.11	0.06	0.07
LW 90%	0.02	0.01	0.01

(HD) Highest Degree, (HB) Highest Betweenness, (LW) Link Weight

2.4.3.1. Cluster I – Small Node-Sensitive Networks

The first cluster includes 20 transit networks that share similar profiles. These 20 networks are featured with relatively: small number of single and transfer stops, fewer multiple links, less frequent buses, and fewer hubs.

This profile is very sensitive to the removal of nodes with the highest Degree or Betweenness. Networks in Cluster I exhibit a 75% reduction in the total trip kilometers with the removal of 10% of nodes with the highest attributes, whereas the removal of 30% of nodes causes a complete loss of operation. The networks in Cluster I are, relatively, the most robust against the removal of links. The removal of 10% and 30% of links with the highest weight reduce the operational total trip kilometer to 75% and 45%, respectively.

The contradictory behaviours associated with node and links removals of transit networks in Cluster I might be attributed to the fact that these transit networks feature fewer hubs and fewer multiple routes. Therefore, the impacts associated with node removal significantly halt transit operation

2.4.3.2. Cluster II – Small Link-Sensitive Networks

The second cluster includes 13 transit networks. Similar to Cluster I, networks in Cluster II operate relatively: small number of single and transfer stops, frequent buses, and few hubs. However, these 13 networks feature multiple transit routes operating on the same links.

Cluster II is relatively less sensitive to the removal of nodes with the highest Degree or Betweenness. Networks in cluster two exhibit 45% reduction in the total trip kilometers with the removal of 10% of nodes with the highest attributes, whereas the removal of 30%

of nodes reduces the operational capacity to 10%. Similarly, the removal of 10% and 30% of links with the highest weights reduce the operational total trip kilometers to 64% and 33%, respectively.

The distinct behaviour transit networks in Cluster II might be attributed to the fact that these transit networks feature a few hubs and a large number of multiple routes. And therefore, there are significant operational impacts associated with the removal of links; however, not as severe as the case of node removal.

2.4.3.3. Cluster III – Medium Less-Sensitive Networks

Seven networks form Cluster III, which represents large transit networks. Networks in Cluster III feature: large number of single and transfer stops, very frequent buses, multiple hubs, and a large number of multiple links.

The behaviour of networks in Cluster III could be seen as a mix of Clusters I and II. Networks are equally sensitive to the removal of nodes with the highest Degree or Betweenness and the removal of links with the highest weights. Networks in Cluster III exhibit 60% reduction in the total trip kilometers with the removal of 10% of nodes with the highest attributes, whereas the removal of 30% of nodes reduces the operational capacity to 25%. In addition, networks in Cluster III are very sensitive to the removal of links. The removal of 10% and 30% of links with the highest weight reduce the operational total trip kilometers to 67% and 36%, respectively.

Overall, the cluster analysis provides a very clear profile on the behaviour of different types of networks against disruption events impacting nodes and links with the

highest attributes. Such results prove that there is no single network model that can fit transit networks, given their varied operation features. In addition, the results highlight that the dynamic-robustness could be assessed based on the robustness of network nodes and links. It should be noted that the statistical comparison, using an Analysis of Variance (ANOVA) model, of the three clusters with respect to the static-robustness measures did not yield significant variation. However, the opposite is true for dynamic-robustness measures. However, given the sensitivity of ANOVA to sample size, results of the MANOVA models should be carefully interpreted.

2.5. Discussion and Conclusion

Motivated by the scarceness of studies on bus transit network robustness analysis in the literature, this study utilizes Complex Network Theory (CNT) to quantify the topological profile and the dynamic robustness of bus transit networks. This is mainly carried out to better inform transit planners and policymakers on avenues to enhance the robustness of bus transit networks. Towards that end, we first quantified four topological measures, namely: Degree Centrality, Betweenness Centrality, Clustering Coefficient, and Average Path Length for 40 Canadian bus transit networks. Subsequently, we evaluated both the static- and the dynamic-robustness of bus transit networks in Canada.

The topological results demonstrate that there is no general network class (e.g. small world, random, or scale-free) observed from the results. Unlike previous findings in the literature pertaining to the current bus network type, which to some degree focused on larger networks, our study provides quantified evidence on the lack of a network class that

could be holistically applied to all bus transit networks. This is mainly attributed to the significant variation between bus transit networks and necessitates the investigation of the cascading loss of service behaviour of each bus transit network under disruptions.

The static-robustness results highlighted a significant level of contradictions between the three measures in the literature: Robustness Metric, Critical Threshold, and the Robustness Indicator. In addition, given the negative correlation between some of these measures, there is a concern on the validity of some of these measures. This, in itself, is a sound contribution to future research and echo the findings of Wang et al. (2017) in the context of metro networks. Although static-measures of robustness are very appealing due to their simplicity, we strongly argue that their utilization leads to misleading outcomes as proved in this study.

Furthermore, we argue that these measures fall short in accommodating the unique operational characteristics associated with different bus transit networks. Future research studies are encouraged to develop more representative static-measures to overcome these fundamental gaps.

The developed dynamic-robustness assessment provided two fundamental findings. First, targeted removal of nodes with the highest attribute (Degree and/or Betweenness) has severe cascading impacts on bus transit network relative to the removal of links. The behaviour of bus transit networks after the removal of links is, to some extent, similar with a negative steady relationship.

Second, there are three distinct clusters of bus transit networks. First, small node

sensitive networks demonstrated a significant degree of sensitivity associated with the removal of nodes with the highest attribute. These networks feature relatively fewer hubs and fewer multiple links. Therefore, the impact of node removal is severe. Second, small link-sensitive networks feature relatively a higher number of multiple transit routes operating on the same links. Therefore, there is a significant impact on operation associated with the removal of links with the highest weight. Third, the considered medium-size networks are relatively less sensitive to the removal of links and/or nodes. This is mainly attributed to the distribution of hubs and multiple links.

Aside from the scientific contributions, we provided practical indications directed to policymakers and planners. The results of the cluster analysis provide a vivid picture of the holistic behaviour of transit networks in Canada, which could be easily assessed and interpreted. The three bus transit network behaviours emerged from the analysis provide directions to enhance their robustness, especially in area of rapid utilization of disruptive technologies such as electric buses (Mahmoud et al. 2016; Mohamed et al. 2017).

Lastly, it is important to note the limitations of our study. Most notable is that the present study did not represent large transit networks in Group 1 due to data limitation. In addition, the present study did not account for the variation in time-tables and passenger capacity. Such variation will indeed impose temporal variation in link weight and station degree. Nonetheless, several recommendations could be inferred from the present study as discussed earlier. Furthermore, we recommend the inclusion of network-level measures in the design process of the bus transit networks, which proved to be effective in capturing the dynamic-robustness of transit services.

2.6. Acknowledgement

2.7. Appendices

Appendix 2-1. Applications of complex Network measures in transit

Measures	Mathematical Equation	Author(s)
Closeness Centrality	$C_c(i) = \dfrac{N-1}{\sum_{v_{j \in n, i \neq j}} d_{ij}}$	Hossain and Alam (2017)
	$C_c(i) = \dfrac{1}{\sum_{v_{j \in n, i \neq j}} d_{ij}}$	von Ferber et al. (2009) Berche et al. (2009)
Betweenness Centrality	$C_B(i) = \displaystyle\sum_{k \neq i \neq j} \dfrac{\sigma_{ij}(i)}{\sigma_{kj}}$	Hossain and Alam (2017) Sun and Guan (2016) von Ferber et al. (2009) Berche et al. (2009)
Clustering Coefficient	$C_i = \dfrac{1}{K_i(K_i-1)} \displaystyle\sum_{j,k j} a_{ij}\, a_{jk} a_{ik}$	Hossain and Alam (2017)
	$C_i = \dfrac{2L_i}{K_i(K_i-1)}$	von Ferber et al. (2009) Berche et al. (2009)
Weighted Clustering Coefficient	$C_i^w = \dfrac{1}{K_i(K_i-1)} \displaystyle\sum_{j,k} \dfrac{1}{w_i} \dfrac{w_{ij}+w_{jk}}{2} a_{ij}\, a_{jk} a_{ik}$	Hossain and Alam (2017)
	$C_i = \dfrac{2}{K_i(K_i-1)} \displaystyle\sum_{j,k}^{N} (W_{ij} W_{jk} W_{ki})^{\frac{1}{3}}$	Aldrich et al. (2015)
Efficiency	$E(G) = \dfrac{1}{N(N-1)} \displaystyle\sum_{i \neq j} \dfrac{1}{d_{ij}}$	Wu et al. (2018) Sun and Guan (2016) Zou et al. (2013)
Average Shortest Path Length	$L(G) = \dfrac{1}{N(N-1)} \displaystyle\sum_{i \neq j} d_{ij}$	Hossain and Alam (2017) Sun and Guan (2016) Lozano et al. (2006) von Ferber et al. (2009)
	$L(G) = \dfrac{2}{N(N-1)} \displaystyle\sum_{i \neq j} d_{ij}$	Berche et al. (2009)
Connectivity	$Cn_{ij} = D_{ij}^V \times (W_{ij}^{Air} \times Cn_{ij}^{Air} + W_{ij}^{Train} \times Cn_{ij}^{Train})$	Zhu et al. (2018)
	$\rho = \dfrac{v_c^t - e^m}{v^t}$	Derrible and Kennedy (2009)
	$M(G) = L_T - \dfrac{1}{L_T!} \displaystyle\sum_{\varepsilon \in \varepsilon(G)} \xi(\varepsilon)$	Lozano et al. (2006)
Robustness	$r^T = \dfrac{L_T - N_{TE} + 1}{N}$	Wang et al. (2017)
	$E(1/H) = \dfrac{2}{N(N-1)} \displaystyle\sum_{i,j=1}^{N} \dfrac{1}{H_{ij}}$	

Measures	Mathematical Equation	Author(s)
	$CC_G = \dfrac{1}{N} \sum\limits_{i=1}^{N} \dfrac{2L_i}{K_i(K_i - 1)}$	
	$E(D) = \dfrac{\sum_{i=1}^{N} k_i}{N}$	
	$\bar{\lambda} = \ln\left[\dfrac{1}{N} \sum\limits_{i=1}^{N} e^{\lambda_i}\right]$	
	$K = \dfrac{\sum_{i=1}^{N} k_i^2}{\sum_{i=1}^{N} k_i}$	
	$M_G = \dfrac{L_T - N_{TE} + 1}{2N - 5}$	
	k (Degree Diversity)	
	μ_{N-1} (Algebraic Connectivity)	
	F (Critical Threshold by simulation)	
	$r^T = \dfrac{L_T - N_{TE} - L_m + 1}{N}$	Derrible and Kennedy (2010)
	$fc = f\,(APL = max)$	Duan et al. (2009)
Network Robustness Index	$NRI_a = \dfrac{C}{C_a}$	Scott et al. (2006)
Small World Coefficient	$S = \dfrac{C/C_R}{L/L_R}$	Aldrich et al. (2015)
Connection Indicator	$\gamma(G) = \dfrac{L_T}{3(N - 2)}$	Lozano et al. (2006)
Toughness	$t(G) = \min\limits_{NCV, K(G-N) > 1} \left(\dfrac{|S|}{K(G - S)} \right)$	
Metro Topological Efficiency	$MTE = \dfrac{1}{N(N - 1)} \sum\limits_{i \neq j} \dfrac{1}{r_{ij}}$	
Station Centrality	$I(i) = \dfrac{\sum_{u \subset s} \sum_{w \subset s} \rho_{uw}(i)}{N(N - 1)} u \neq w$	Wu et al. (2018)
Intuitive Routing Decision Probability	$P_{MTP} = (1 - \dfrac{\varepsilon^2}{\lambda^2})^2 \, (1 - \dfrac{(\gamma - \xi)^2}{\xi^2})$	
Capacity of a Node	$C_i = (1 + \alpha)l_i$	Zou et al. (2013)

Please note that the original notations have been modified in the current study to maintain consistency.

Appendix 2-2. Losses in operating distance based on the removal of nodes with the highest Betweenness

ID	Network	Hubs	K_{max}/L_T	N_{TE}	N_L	Percentage removal of nodes with the highest Betweenness									
						10% (number of nodes)		30% (number of nodes)		50% (number of nodes)		70% (number of nodes)		90% (number of nodes)	
1	Cornwall	1	100%	6	14	0.000	(1)	0.000	(2)	0.000	(3)	0.000	(4)	0.000	(5)
2	North Bay	1	100%	4	16	0.000	(1)	0.000	(1)	0.000	(2)	0.000	(3)	0.000	(4)
3	Vernon	1	80%	9	20	0.187	(1)	0.000	(3)	0.000	(5)	0.000	(6)	0.000	(8)
4	Cranbrook	2	82%	9	17	0.117	(1)	0.000	(3)	0.000	(5)	0.000	(6)	0.000	(8)
5	Leduc	2	85%	5	13	0.184	(1)	0.000	(2)	0.000	(3)	0.000	(4)	0.000	(5)
6	Welland	2	87%	8	23	0.086	(1)	0.021	(2)	0.000	(4)	0.000	(6)	0.000	(7)
7	Niagara Region	3	64%	10	25	0.357	(1)	0.123	(3)	0.000	(5)	0.000	(7)	0.000	(9)
8	Brantford	3	57%	7	21	0.199	(1)	0.000	(4)	0.000	(6)	0.000	(8)	0.000	(11)
9	Milton	3	71%	12	31	0.110	(1)	0.000	(3)	0.000	(5)	0.000	(7)	0.000	(9)
10	Belleville	3	77%	10	26	0.542	(1)	0.016	(3)	0.000	(5)	0.000	(7)	0.000	(9)
11	Port Alberni	3	52%	10	31	0.589	(1)	0.147	(2)	0.000	(4)	0.000	(5)	0.000	(6)
12	Brandon	4	40%	10	25	0.581	(1)	0.139	(4)	0.114	(6)	0.028	(8)	0.000	(11)
13	Lethbridge	4	41%	12	29	0.697	(1)	0.000	(4)	0.000	(7)	0.000	(9)	0.000	(12)
14	Fort St. John	4	37%	13	43	0.421	(1)	0.212	(3)	0.110	(5)	0.000	(7)	0.000	(9)
15	Kitimat	4	53%	12	34	0.305	(1)	0.034	(4)	0.013	(6)	0.000	(8)	0.000	(11)
16	Terrace	4	53%	13	34	0.907	(1)	0.282	(4)	0.016	(7)	0.000	(9)	0.000	(12)
17	Nanaimo	5	38%	8	26	0.560	(1)	0.097	(4)	0.004	(7)	0.000	(10)	0.000	(13)
18	Whistler	5	34%	14	70	0.526	(1)	0.155	(3)	0.000	(6)	0.000	(8)	0.000	(10)
19	Sunshine Coast	5	50%	11	28	0.672	(1)	0.622	(2)	0.284	(4)	0.000	(6)	0.000	(7)
20	Squamish	6	42%	10	38	0.746	(1)	0.490	(3)	0.000	(5)	0.000	(7)	0.000	(9)
21	Kamloops	7	32%	19	57	0.588	(2)	0.084	(6)	0.000	(10)	0.000	(13)	0.000	(17)
22	Kingston	7	34%	23	88	0.417	(2)	0.056	(7)	0.000	(12)	0.000	(16)	0.000	(21)
23	Prince George	7	29%	21	85	0.452	(2)	0.210	(6)	0.004	(11)	0.000	(15)	0.000	(19)
24	Whitehorse	8	20%	18	51	0.728	(2)	0.334	(5)	0.119	(9)	0.075	(13)	0.045	(16)
25	Yellowknife	8	29%	11	21	0.746	(1)	0.669	(3)	0.283	(6)	0.000	(8)	0.000	(10)
26	Oakville	9	32%	17	50	0.467	(3)	0.203	(9)	0.022	(15)	0.000	(20)	0.000	(26)
27	Chilliwack	9	23%	32	99	0.708	(2)	0.407	(5)	0.000	(9)	0.000	(12)	0.000	(15)
28	Saint John	9	27%	20	73	0.560	(3)	0.221	(10)	0.017	(16)	0.000	(22)	0.000	(29)
29	Thunder Bay	9	34%	29	130	0.528	(2)	0.144	(6)	0.006	(10)	0.000	(14)	0.000	(18)
30	Saskatoon	10	32%	43	96	0.248	(5)	0.065	(16)	0.000	(26)	0.000	(36)	0.000	(47)
31	Windsor	10	33%	52	161	0.576	(2)	0.113	(7)	0.017	(11)	0.000	(15)	0.000	(20)
32	Barrie	10	26%	22	76	0.380	(4)	0.098	(13)	0.003	(22)	0.000	(30)	0.000	(39)
33	Burlington	11	27%	28	111	0.331	(3)	0.167	(8)	0.096	(14)	0.036	(20)	0.000	(25)
34	Regina	14	31%	38	147	0.206	(4)	0.037	(11)	0.000	(19)	0.000	(27)	0.000	(34)
35	Hamilton	24	14%	65	217	0.637	(6)	0.194	(17)	0.112	(28)	0.048	(39)	0.048	(50)
36	London	24	16%	56	152	0.465	(7)	0.104	(20)	0.012	(33)	0.000	(46)	0.000	(59)
37	Durham Region	25	10%	85	273	0.613	(9)	0.165	(26)	0.081	(43)	0.050	(60)	0.000	(77)
38	Waterloo	29	14%	65	249	0.462	(7)	0.127	(20)	0.010	(33)	0.007	(46)	0.000	(59)
39	Victoria	34	9%	103	449	0.444	(10)	0.211	(31)	0.081	(52)	0.018	(72)	0.000	(93)
40	Brampton	35	10%	115	407	0.511	(12)	0.241	(35)	0.106	(58)	0.003	(81)	0.000	(104)

Appendix 2-3. Losses in operating distance based on the removal of nodes with the highest Degree

ID	Network	Hubs	K_{max}/L_T	N_{TE}	N_L	Percentage removal of nodes with the highest Degree									
						10% (number of nodes)		30% (number of nodes)		50% (number of nodes)		70% (number of nodes)		90% (number of nodes)	
1	Cornwall	1	100%	6	14	0.000	(1)	0.000	(2)	0.000	(3)	0.000	(4)	0.000	(5)
2	North Bay	1	100%	4	16	0.000	(1)	0.000	(1)	0.000	(2)	0.000	(3)	0.000	(4)
3	Vernon	1	80%	9	20	0.187	(1)	0.000	(3)	0.000	(5)	0.000	(6)	0.000	(8)
4	Cranbrook	2	82%	9	17	0.117	(1)	0.000	(3)	0.000	(5)	0.000	(6)	0.000	(8)
5	Leduc	2	85%	5	13	0.184	(1)	0.000	(2)	0.000	(3)	0.000	(4)	0.000	(5)
6	Welland	2	87%	8	23	0.086	(1)	0.021	(2)	0.000	(4)	0.000	(6)	0.000	(7)
7	Niagara Region	3	64%	10	25	0.357	(1)	0.149	(3)	0.000	(5)	0.000	(7)	0.000	(9)
8	Brantford	3	57%	7	21	0.199	(1)	0.000	(4)	0.000	(6)	0.000	(8)	0.000	(11)
9	Milton	3	71%	12	31	0.110	(1)	0.000	(3)	0.000	(5)	0.000	(7)	0.000	(9)
10	Belleville	3	77%	10	26	0.542	(1)	0.016	(3)	0.000	(5)	0.000	(7)	0.000	(9)
11	Port Alberni	3	52%	10	31	0.589	(1)	0.147	(2)	0.057	(4)	0.000	(5)	0.000	(6)
12	Brandon	4	40%	10	25	0.581	(1)	0.317	(4)	0.114	(6)	0.028	(8)	0.000	(11)
13	Lethbridge	4	41%	12	29	0.697	(1)	0.000	(4)	0.000	(7)	0.000	(9)	0.000	(12)
14	Fort St. John	4	37%	13	43	0.421	(1)	0.212	(4)	0.110	(5)	0.000	(7)	0.000	(9)
15	Kitimat	4	53%	12	34	0.305	(1)	0.034	(4)	0.013	(6)	0.000	(8)	0.000	(11)
16	Terrace	4	53%	13	34	0.401	(1)	0.310	(4)	0.107	(7)	0.007	(9)	0.000	(12)
17	Nanaimo	5	38%	8	26	0.560	(1)	0.096	(4)	0.053	(7)	0.000	(10)	0.000	(13)
18	Whistler	5	34%	14	70	0.526	(1)	0.155	(3)	0.000	(6)	0.000	(8)	0.000	(10)
19	Sunshine Coast	5	50%	11	28	0.638	(1)	0.529	(2)	0.284	(4)	0.000	(6)	0.000	(7)
20	Squamish	6	42%	10	38	0.744	(1)	0.490	(3)	0.000	(5)	0.000	(7)	0.000	(9)
21	Kamloops	7	32%	19	57	0.588	(2)	0.084	(6)	0.000	(10)	0.000	(13)	0.000	(17)
22	Kingston	7	34%	23	88	0.417	(2)	0.061	(7)	0.000	(12)	0.000	(16)	0.000	(21)
23	Prince George	7	29%	21	85	0.561	(2)	0.215	(6)	0.004	(11)	0.000	(15)	0.000	(19)
24	Whitehorse	8	20%	18	51	0.728	(2)	0.334	(5)	0.119	(9)	0.075	(13)	0.000	(16)
25	Yellowknife	8	29%	11	21	0.746	(1)	0.277	(3)	0.166	(6)	0.000	(8)	0.000	(10)
26	Oakville	9	32%	17	50	0.467	(3)	0.203	(9)	0.022	(15)	0.000	(20)	0.000	(26)
27	Chilliwack	9	23%	32	99	0.659	(2)	0.554	(5)	0.043	(9)	0.000	(12)	0.000	(15)
28	Saint John	9	27%	20	73	0.423	(3)	0.138	(10)	0.050	(16)	0.000	(22)	0.000	(29)
29	Thunder Bay	9	34%	29	130	0.471	(2)	0.079	(6)	0.006	(10)	0.000	(14)	0.000	(18)
30	Saskatoon	10	32%	43	96	0.307	(5)	0.065	(16)	0.051	(26)	0.000	(36)	0.000	(47)
31	Windsor	10	33%	52	161	0.623	(2)	0.118	(7)	0.026	(11)	0.000	(15)	0.000	(20)
32	Barrie	10	26%	22	76	0.388	(4)	0.067	(13)	0.025	(22)	0.000	(30)	0.000	(39)
33	Burlington	11	27%	28	111	0.331	(3)	0.171	(8)	0.097	(14)	0.000	(20)	0.000	(25)
34	Regina	14	31%	38	147	0.215	(4)	0.003	(11)	0.000	(19)	0.000	(27)	0.000	(34)
35	Hamilton	24	14%	65	217	0.589	(6)	0.109	(17)	0.064	(28)	0.048	(39)	0.048	(50)
36	London	24	16%	56	152	0.465	(7)	0.104	(20)	0.012	(33)	0.000	(46)	0.000	(59)
37	Durham Region	25	10%	85	273	0.536	(9)	0.194	(26)	0.091	(43)	0.050	(60)	0.000	(77)
38	Waterloo	29	14%	65	249	0.448	(7)	0.069	(20)	0.011	(33)	0.007	(46)	0.000	(59)
39	Victoria	34	9%	103	449	0.436	(10)	0.214	(31)	0.063	(52)	0.006	(72)	0.000	(93)
40	Brampton	35	10%	115	407	0.511	(12)	0.241	(35)	0.103	(58)	0.000	(81)	0.000	(104)

Appendix 2-4. Losses in operation distance based on the removal of links with the highest Link Weight

ID	Network	Hubs	K_{max}/L_T	N_{TE}	N_L	Percentage removal of links with the highest Link Weights									
						10% (number of links)		30% (number of links)		50% (number of links)		70% (number of links)		90% (number of links)	
1	Cornwall	1	100%	6	14	0.797	(1)	0.536	(4)	0.352	(7)	0.187	(10)	0.088	(13)
2	North Bay	1	100%	4	16	0.767	(2)	0.479	(5)	0.250	(8)	0.100	(11)	0.020	(14)
3	Vernon	1	80%	9	20	0.857	(2)	0.582	(6)	0.401	(10)	0.253	(14)	0.071	(18)
4	Cranbrook	2	82%	9	17	0.781	(2)	0.581	(5)	0.425	(9)	0.307	(12)	0.111	(15)
5	Leduc	2	85%	5	13	0.851	(1)	0.485	(4)	0.278	(7)	0.096	(9)	0.024	(12)
6	Welland	2	87%	8	23	0.855	(2)	0.605	(7)	0.407	(12)	0.280	(16)	0.045	(21)
7	Niagara Region	3	64%	10	25	0.756	(3)	0.347	(9)	0.145	(16)	0.066	(22)	0.019	(28)
8	Brantford	3	57%	7	21	0.714	(3)	0.394	(9)	0.235	(16)	0.126	(22)	0.033	(28)
9	Milton	3	71%	12	31	0.798	(3)	0.558	(8)	0.290	(13)	0.160	(18)	0.022	(23)
10	Belleville	3	77%	10	26	0.736	(3)	0.471	(8)	0.246	(13)	0.090	(18)	0.013	(23)
11	Port Alberni	3	52%	10	31	0.821	(2)	0.633	(6)	0.506	(11)	0.273	(15)	0.074	(19)
12	Brandon	4	40%	10	25	0.809	(3)	0.530	(10)	0.265	(17)	0.134	(24)	0.059	(31)
13	Lethbridge	4	41%	12	29	0.841	(3)	0.523	(10)	0.328	(17)	0.153	(24)	0.024	(31)
14	Fort St. John	4	37%	13	43	0.688	(3)	0.392	(8)	0.234	(13)	0.104	(18)	0.009	(23)
15	Kitimat	4	53%	12	34	0.493	(3)	0.309	(9)	0.088	(15)	0.043	(20)	0.013	(26)
16	Terrace	4	53%	13	34	0.502	(4)	0.201	(13)	0.130	(22)	0.058	(30)	0.020	(39)
17	Nanaimo	5	38%	8	26	0.680	(7)	0.398	(21)	0.226	(35)	0.113	(49)	0.044	(63)
18	Whistler	5	34%	14	70	0.770	(3)	0.459	(8)	0.234	(14)	0.100	(20)	0.007	(25)
19	Sunshine Coast	5	50%	11	28	0.831	(3)	0.523	(8)	0.218	(13)	0.091	(18)	0.018	(23)
20	Squamish	6	42%	10	38	0.708	(4)	0.431	(11)	0.250	(19)	0.108	(27)	0.008	(34)
21	Kamloops	7	32%	19	57	0.734	(6)	0.511	(17)	0.321	(29)	0.167	(40)	0.050	(51)
22	Kingston	7	34%	23	88	0.810	(9)	0.523	(26)	0.331	(44)	0.153	(62)	0.022	(79)
23	Prince George	7	29%	21	85	0.781	(9)	0.473	(26)	0.248	(43)	0.102	(60)	0.018	(77)
24	Whitehorse	8	20%	18	51	0.722	(5)	0.423	15)	0.280	(26)	0.151	(36)	0.066	(46)
25	Yellowknife	8	29%	11	21	0.680	(2)	0.360	(6)	0.179	(11)	0.096	(15)	0.027	(19)
26	Oakville	9	32%	17	50	0.775	(13)	0.464	(39)	0.307	(65)	0.166	(91)	0.052	(117)
27	Chilliwack	9	23%	32	99	0.635	(5)	0.314	(15)	0.196	(25)	0.084	(35)	0.031	(45)
28	Saint John	9	27%	20	73	0.863	(10)	0.424	(30)	0.287	(50)	0.100	(69)	0.015	(89)
29	Thunder Bay	9	34%	29	130	0.816	(7)	0.522	(22)	0.342	(37)	0.136	(51)	0.033	(66)
30	Saskatoon	10	32%	43	96	0.801	(16)	0.514	(48)	0.320	(81)	0.172	(113)	0.055	(145)
31	Windsor	10	33%	52	161	0.823	(8)	0.512	(23)	0.287	(38)	0.157	(53)	0.048	(68)
32	Barrie	10	26%	22	76	0.793	(10)	0.569	(29)	0.342	(48)	0.171	(67)	0.089	(86)
33	Burlington	11	27%	28	111	0.782	(11)	0.561	(33)	0.406	(56)	0.235	(78)	0.038	(100)
34	Regina	14	31%	38	147	0.737	(15)	0.424	(44)	0.252	(74)	0.119	(103)	0.026	(132)
35	Hamilton	24	14%	65	217	0.756	(15)	0.417	(46)	0.255	(76)	0.134	(106)	0.023	(137)
36	London	24	16%	56	152	0.782	(22)	0.480	(65)	0.285	(109)	0.134	(152)	0.031	(195)
37	Durham Region	25	10%	85	273	0.800	(27)	0.469	(82)	0.285	(137)	0.124	(191)	0.035	(246)
38	Waterloo	29	14%	65	249	0.801	(25)	0.510	(75)	0.307	(125)	0.145	(174)	0.039	(224)
39	Victoria	34	9%	103	449	0.788	(45)	0.616	(135)	0.398	(225)	0.180	(314)	0.050	(404)
40	Brampton	35	10%	115	407	0.778	(41)	0.491	(122)	0.284	(204)	0.151	(285)	0.044	(366)

Appendix 2-5. Losses in service frequency based on the removal of nodes with the highest Betweenness

ID	Network	Hubs	K_{max}/L_T	N_{TE}	N_L	Percentage removal of nodes with the highest Betweenness									
						10% (number of nodes)		30% (number of nodes)		50% (number of nodes)		70% (number of nodes)		90% (number of nodes)	
1	Cornwall	1	100%	6	14	0.000	(1)	0.000	(2)	0.000	(3)	0.000	(4)	0.000	(5)
2	North Bay	1	100%	4	16	0.000	(1)	0.000	(1)	0.000	(2)	0.000	(3)	0.000	(4)
3	Vernon	1	80%	9	20	0.178	(1)	0.000	(3)	0.000	(5)	0.000	(6)	0.000	(8)
4	Cranbrook	2	82%	9	17	0.204	(1)	0.000	(3)	0.000	(5)	0.000	(6)	0.000	(8)
5	Leduc	2	85%	5	13	0.153	(1)	0.000	(2)	0.000	(3)	0.000	(4)	0.000	(5)
6	Welland	2	87%	8	23	0.063	(1)	0.002	(2)	0.000	(4)	0.000	(6)	0.000	(7)
7	Niagara Region	3	64%	10	25	0.487	(1)	0.166	(3)	0.000	(5)	0.000	(7)	0.000	(9)
8	Brantford	3	57%	7	21	0.298	(1)	0.000	(4)	0.000	(6)	0.000	(8)	0.000	(11)
9	Milton	3	71%	12	31	0.261	(1)	0.000	(3)	0.000	(5)	0.000	(7)	0.000	(9)
10	Belleville	3	77%	10	26	0.281	(1)	0.045	(3)	0.000	(5)	0.000	(7)	0.000	(9)
11	Port Alberni	3	52%	10	31	0.395	(1)	0.093	(2)	0.000	(4)	0.000	(5)	0.000	(6)
12	Brandon	4	40%	10	25	0.433	(1)	0.089	(4)	0.056	(6)	0.028	(8)	0.000	(11)
13	Lethbridge	4	41%	12	29	0.595	(1)	0.000	(4)	0.000	(7)	0.000	(9)	0.000	(12)
14	Fort St. John	4	37%	13	43	0.600	(1)	0.280	(3)	0.119	(5)	0.000	(7)	0.000	(9)
15	Kitimat	4	53%	12	34	0.348	(1)	0.193	(4)	0.060	(6)	0.000	(8)	0.000	(11)
16	Terrace	4	53%	13	34	0.597	(1)	0.198	(4)	0.073	(7)	0.000	(9)	0.000	(12)
17	Nanaimo	5	38%	8	26	0.652	(1)	0.133	(4)	0.031	(7)	0.000	(10)	0.000	(13)
18	Whistler	5	34%	14	70	0.468	(1)	0.175	(3)	0.000	(6)	0.000	(8)	0.000	(10)
19	Sunshine Coast	5	50%	11	28	0.655	(1)	0.483	(2)	0.172	(4)	0.000	(6)	0.000	(7)
20	Squamish	6	42%	10	38	0.712	(1)	0.274	(3)	0.000	(5)	0.000	(7)	0.000	(9)
21	Kamloops	7	32%	19	57	0.418	(2)	0.101	(6)	0.000	(10)	0.000	(13)	0.000	(17)
22	Kingston	7	34%	23	88	0.366	(2)	0.118	(7)	0.000	(12)	0.000	(16)	0.000	(21)
23	Prince George	7	29%	21	85	0.523	(2)	0.178	(6)	0.021	(11)	0.000	(15)	0.000	(19)
24	Whitehorse	8	20%	18	51	0.674	(2)	0.316	(5)	0.121	(9)	0.054	(13)	0.037	(16)
25	Yellowknife	8	29%	11	21	0.810	(1)	0.619	(3)	0.238	(6)	0.000	(8)	0.000	(10)
26	Oakville	9	32%	17	50	0.430	(3)	0.113	(9)	0.007	(15)	0.000	(20)	0.000	(26)
27	Chilliwack	9	23%	32	99	0.309	(2)	0.147	(5)	0.000	(9)	0.000	(12)	0.000	(15)
28	Saint John	9	27%	20	73	0.568	(3)	0.228	(10)	0.059	(16)	0.000	(22)	0.000	(29)
29	Thunder Bay	9	34%	29	130	0.492	(2)	0.204	(6)	0.005	(10)	0.000	(14)	0.000	(18)
30	Saskatoon	10	32%	43	96	0.237	(5)	0.023	(16)	0.000	(26)	0.000	(36)	0.000	(47)
31	Windsor	10	33%	52	161	0.509	(2)	0.065	(7)	0.012	(11)	0.000	(15)	0.000	(20)
32	Barrie	10	26%	22	76	0.399	(4)	0.097	(13)	0.013	(22)	0.000	(30)	0.000	(39)
33	Burlington	11	27%	28	111	0.428	(3)	0.286	(8)	0.128	(14)	0.031	(20)	0.000	(25)
34	Regina	14	31%	38	147	0.331	(4)	0.081	(11)	0.000	(19)	0.000	(27)	0.000	(34)
35	Hamilton	24	14%	65	217	0.710	(6)	0.161	(17)	0.048	(28)	0.010	(39)	0.010	(50)
36	London	24	16%	56	152	0.415	(7)	0.088	(20)	0.006	(33)	0.000	(46)	0.000	(59)
37	Durham Region	25	10%	85	273	0.489	(9)	0.165	(26)	0.050	(43)	0.007	(60)	0.000	(77)
38	Waterloo	29	14%	65	249	0.519	(7)	0.131	(20)	0.006	(33)	0.002	(46)	0.000	(59)
39	Victoria	34	9%	103	449	0.508	(10)	0.232	(31)	0.113	(52)	0.005	(72)	0.000	(93)
40	Brampton	35	10%	115	407	0.526	(12)	0.255	(35)	0.070	(58)	0.003	(81)	0.000	(104)

Appendix 2-6. Losses in service frequency based on the removal of nodes with the highest Degree

ID	Network	Hubs	K_{max}/L_T	N_{TE}	N_L	Percentage removal of nodes with the highest Degree									
						10% (number of nodes)		30% (number of nodes)		50% (number of nodes)		70% (number of nodes)		90% (number of nodes)	
1	Cornwall	1	100%	6	14	0.000	(1)	0.000	(2)	0.000	(3)	0.000	(4)	0.000	(5)
2	North Bay	1	100%	4	16	0.000	(1)	0.000	(1)	0.000	(2)	0.000	(3)	0.000	(4)
3	Vernon	1	80%	9	20	0.178	(1)	0.000	(3)	0.000	(5)	0.000	(6)	0.000	(8)
4	Cranbrook	2	82%	9	17	0.204	(1)	0.000	(3)	0.000	(5)	0.000	(6)	0.000	(8)
5	Leduc	2	85%	5	13	0.153	(1)	0.000	(2)	0.000	(3)	0.000	(4)	0.000	(5)
6	Welland	2	87%	8	23	0.063	(1)	0.002	(2)	0.000	(4)	0.000	(6)	0.000	(7)
7	Niagara Region	3	64%	10	25	0.487	(1)	0.116	(3)	0.000	(5)	0.000	(7)	0.000	(9)
8	Brantford	3	57%	7	21	0.298	(1)	0.000	(4)	0.000	(6)	0.000	(8)	0.000	(11)
9	Milton	3	71%	12	31	0.261	(1)	0.000	(3)	0.000	(5)	0.000	(7)	0.000	(9)
10	Belleville	3	77%	10	26	0.281	(1)	0.045	(3)	0.000	(5)	0.000	(7)	0.000	(9)
11	Port Alberni	3	52%	10	31	0.395	(1)	0.093	(2)	0.012	(4)	0.000	(5)	0.000	(6)
12	Brandon	4	40%	10	25	0.433	(1)	0.147	(4)	0.056	(6)	0.028	(8)	0.000	(11)
13	Lethbridge	4	41%	12	29	0.595	(1)	0.000	(4)	0.000	(7)	0.000	(9)	0.000	(12)
14	Fort St. John	4	37%	13	43	0.600	(1)	0.280	(3)	0.119	(5)	0.000	(7)	0.000	(9)
15	Kitimat	4	53%	12	34	0.348	(1)	0.193	(4)	0.060	(6)	0.000	(8)	0.000	(11)
16	Terrace	4	53%	13	34	0.560	(1)	0.271	(4)	0.099	(7)	0.026	(9)	0.000	(12)
17	Nanaimo	5	38%	8	26	0.652	(1)	0.173	(4)	0.035	(7)	0.000	(10)	0.000	(13)
18	Whistler	5	34%	14	70	0.468	(1)	0.175	(3)	0.000	(6)	0.000	(8)	0.000	(10)
19	Sunshine Coast	5	50%	11	28	0.690	(1)	0.517	(2)	0.172	(4)	0.000	(6)	0.000	(7)
20	Squamish	6	42%	10	38	0.561	(1)	0.274	(3)	0.000	(5)	0.000	(7)	0.000	(9)
21	Kamloops	7	32%	19	57	0.418	(2)	0.101	(6)	0.000	(10)	0.000	(13)	0.000	(17)
22	Kingston	7	34%	23	88	0.366	(2)	0.148	(7)	0.000	(12)	0.000	(16)	0.000	(21)
23	Prince George	7	29%	21	85	0.563	(2)	0.134	(6)	0.021	(11)	0.000	(15)	0.000	(19)
24	Whitehorse	8	20%	18	51	0.674	(2)	0.316	(5)	0.121	(9)	0.054	(13)	0.000	(16)
25	Yellowknife	8	29%	11	21	0.810	(1)	0.381	(3)	0.095	(6)	0.000	(8)	0.000	(10)
26	Oakville	9	32%	17	50	0.430	(3)	0.113	(9)	0.007	(15)	0.000	(20)	0.000	(26)
27	Chilliwack	9	23%	32	99	0.481	(2)	0.099	(5)	0.009	(9)	0.000	(12)	0.000	(15)
28	Saint John	9	27%	20	73	0.512	(3)	0.086	(10)	0.054	(16)	0.000	(22)	0.000	(29)
29	Thunder Bay	9	34%	29	130	0.424	(2)	0.094	(6)	0.005	(10)	0.000	(14)	0.000	(18)
30	Saskatoon	10	32%	43	96	0.300	(5)	0.023	(16)	0.005	(26)	0.000	(36)	0.000	(47)
31	Windsor	10	33%	52	161	0.533	(2)	0.109	(7)	0.010	(11)	0.000	(15)	0.000	(20)
32	Barrie	10	26%	22	76	0.361	(4)	0.090	(13)	0.034	(22)	0.000	(30)	0.000	(39)
33	Burlington	11	27%	28	111	0.428	(3)	0.207	(8)	0.108	(14)	0.000	(20)	0.000	(25)
34	Regina	14	31%	38	147	0.278	(4)	0.014	(11)	0.000	(19)	0.000	(27)	0.000	(34)
35	Hamilton	24	14%	65	217	0.482	(6)	0.094	(17)	0.016	(28)	0.010	(39)	0.010	(50)
36	London	24	16%	56	152	0.415	(7)	0.088	(20)	0.006	(33)	0.000	(46)	0.000	(59)
37	Durham Region	25	10%	85	273	0.407	(9)	0.172	(26)	0.046	(43)	0.007	(60)	0.000	(77)
38	Waterloo	29	14%	65	249	0.436	(7)	0.076	(20)	0.018	(33)	0.002	(46)	0.000	(59)
39	Victoria	34	9%	103	449	0.501	(10)	0.208	(31)	0.047	(52)	0.001	(72)	0.000	(93)
40	Brampton	35	10%	115	407	0.526	(12)	0.255	(35)	0.068	(58)	0.000	(81)	0.000	(104)

Appendix 2-7. Losses in service frequency based on the removal of links with the highest Link Weight

ID	Network	Hubs	K_{max}/L_T	N_{TE}	N_L	Percentage removal of links with the highest Link Weights									
						10% (number of links)		30% (number of links)		50% (number of links)		70% (number of links)		90% (number of links)	
1	Cornwall	1	100%	6	14	0.924	(1)	0.707	(4)	0.501	(7)	0.281	(10)	0.135	(13)
2	North Bay	1	100%	4	16	0.876	(2)	0.689	(5)	0.494	(8)	0.301	(11)	0.049	(14)
3	Vernon	1	80%	9	20	0.880	(2)	0.680	(6)	0.454	(10)	0.212	(14)	0.063	(18)
4	Cranbrook	2	82%	9	17	0.822	(2)	0.607	(5)	0.411	(9)	0.236	(12)	0.065	(15)
5	Leduc	2	85%	5	13	0.917	(1)	0.667	(4)	0.431	(7)	0.236	(9)	0.083	(12)
6	Welland	2	87%	8	23	0.880	(2)	0.597	(7)	0.312	(12)	0.013	(16)	0.004	(21)
7	Niagara Region	3	64%	10	25	0.894	(3)	0.709	(9)	0.523	(16)	0.312	(22)	0.106	(28)
8	Brantford	3	57%	7	21	0.904	(3)	0.715	(9)	0.515	(16)	0.318	(22)	0.129	(28)
9	Milton	3	71%	12	31	0.877	(3)	0.685	(8)	0.512	(13)	0.318	(18)	0.091	(23)
10	Belleville	3	77%	10	26	0.851	(3)	0.650	(8)	0.474	(13)	0.300	(18)	0.022	(23)
11	Port Alberni	3	52%	10	31	0.860	(2)	0.581	(6)	0.302	(11)	0.081	(15)	0.035	(19)
12	Brandon	4	40%	10	25	0.907	(3)	0.688	(10)	0.509	(17)	0.281	(24)	0.091	(31)
13	Lethbridge	4	41%	12	29	0.910	(3)	0.720	(10)	0.477	(17)	0.261	(24)	0.113	(31)
14	Fort St. John	4	37%	13	43	0.879	(3)	0.680	(8)	0.481	(13)	0.280	(18)	0.040	(23)
15	Kitimat	4	53%	12	34	0.863	(3)	0.412	(9)	0.180	(15)	0.069	(20)	0.009	(26)
16	Terrace	4	53%	13	34	0.890	(4)	0.623	(13)	0.282	(22)	0.106	(30)	0.022	(39)
17	Nanaimo	5	38%	8	26	0.881	(7)	0.640	(21)	0.409	(35)	0.161	(49)	0.018	(63)
18	Whistler	5	34%	14	70	0.787	(3)	0.477	(8)	0.297	(14)	0.166	(20)	0.028	(25)
19	Sunshine Coast	5	50%	11	28	0.835	(3)	0.607	(8)	0.464	(13)	0.261	(18)	0.035	(23)
20	Squamish	6	42%	10	38	0.875	(4)	0.639	(11)	0.447	(19)	0.244	(27)	0.046	(34)
21	Kamloops	7	32%	19	57	0.874	(6)	0.601	(17)	0.381	(29)	0.225	(40)	0.075	(51)
22	Kingston	7	34%	23	88	0.878	(9)	0.641	(26)	0.421	(44)	0.235	(62)	0.050	(79)
23	Prince George	7	29%	21	85	0.865	(9)	0.651	(26)	0.466	(43)	0.258	(60)	0.044	(77)
24	Whitehorse	8	20%	18	51	0.871	(5)	0.674	15)	0.454	(26)	0.274	(36)	0.076	(46)
25	Yellowknife	8	29%	11	21	0.905	(2)	0.714	(6)	0.524	(11)	0.333	(15)	0.143	(19)
26	Oakville	9	32%	17	50	0.856	(13)	0.601	(39)	0.341	(65)	0.124	(91)	0.015	(117)
27	Chilliwack	9	23%	32	99	0.837	(5)	0.582	(15)	0.292	(25)	0.150	(35)	0.043	(45)
28	Saint John	9	27%	20	73	0.813	(10)	0.661	(30)	0.389	(50)	0.172	(69)	0.041	(89)
29	Thunder Bay	9	34%	29	130	0.856	(7)	0.611	(22)	0.363	(37)	0.178	(51)	0.048	(66)
30	Saskatoon	10	32%	43	96	0.837	(16)	0.617	(48)	0.414	(81)	0.218	(113)	0.037	(145)
31	Windsor	10	33%	52	161	0.839	(8)	0.611	(23)	0.375	(38)	0.180	(53)	0.020	(68)
32	Barrie	10	26%	22	76	0.885	(10)	0.666	(29)	0.475	(48)	0.277	(67)	0.044	(86)
33	Burlington	11	27%	28	111	0.833	(11)	0.532	(33)	0.269	(56)	0.077	(78)	0.018	(100)
34	Regina	14	31%	38	147	0.869	(15)	0.652	(44)	0.429	(74)	0.240	(103)	0.065	(132)
35	Hamilton	24	14%	65	217	0.853	(15)	0.648	(46)	0.427	(76)	0.251	(106)	0.101	(137)
36	London	24	16%	56	152	0.869	(22)	0.637	(65)	0.407	(109)	0.211	(152)	0.059	(195)
37	Durham Region	25	10%	85	273	0.841	(27)	0.576	(82)	0.360	(137)	0.145	(191)	0.014	(246)
38	Waterloo	29	14%	65	249	0.843	(25)	0.627	(75)	0.404	(125)	0.218	(174)	0.050	(224)
39	Victoria	34	9%	103	449	0.809	(45)	0.429	(135)	0.209	(225)	0.087	(314)	0.015	(404)
40	Brampton	35	10%	115	407	0.839	(41)	0.617	(122)	0.385	(204)	0.175	(285)	0.034	(366)
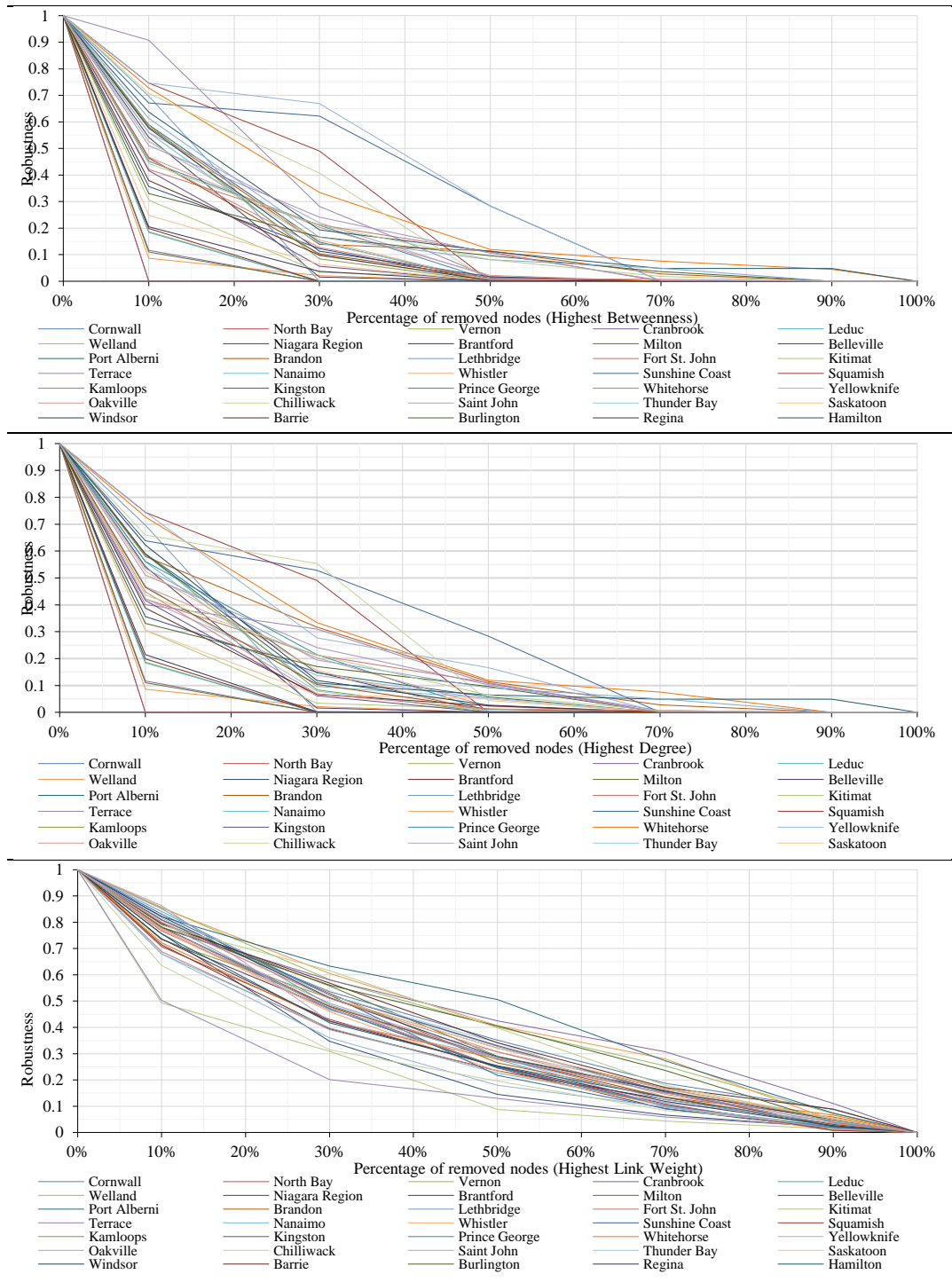
Appendix 2-8. The robustness indices for operating distance for all removal approaches
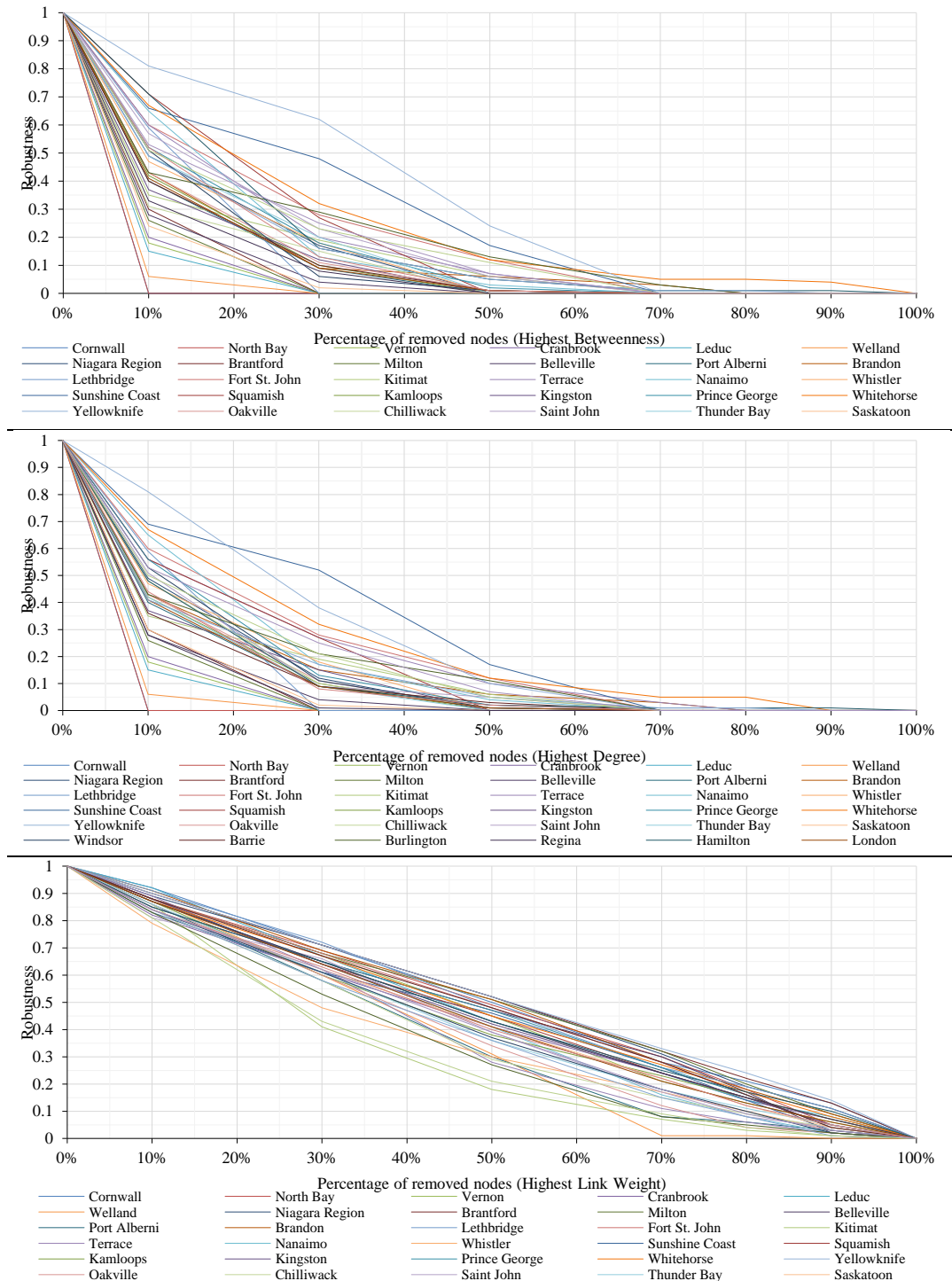
Appendix 2-9. The robustness indices for service frequency for all removal approaches

2.8. References

Albert, R., and A. L. Barabasi. 2002. "Statistical Mechanics of Complex Networks." Reviews of Modern Physics 74 (1): 47–97. doi:10.1103/RevModPhys.74.47.

Aldrich, P. R., J. El-Zabet, S. Hassan, J. Briguglio, E. Aliaj, M. Radcliffe, T. Mirza, T. Comar, J. Nadolski, and C. D. Huebner. 2015. "Monte Carlo Tests of Small-World Architecture for Coarse-Grained Networks of the United States Railroad and Highway Transportation Systems." Physica a-Statistical Mechanics and Its Applications 438: 32–39. doi: 10.1016/j.physa.2015.06.013.

Bababeik, M., N. Khademi, A. Chen, and M. M. Nasiri. 2017. "Vulnerability Analysis of Railway Networks in Case of Multi-Link Blockage." In 19th Euro Working Group on Transportation Meeting (Ewgt2016), 275-84.

Barabási, Albert- László. 2016. Network Science. Cambridge University Press.

Berche, B., C. von Ferber, T. Holovatch, and Y. Holovatch. 2009. "Resilience of Public Transport NetworksAgainstAttacks."EuropeanPhysical JournalB71 (1): 125–137. doi:10.1140/epjb/e2009-00291-3.

Choromański, Krzysztof, Michał Matuszak, and Jacek Mie̜kisz. 2013. "Scale-Free Graph with Preferential Attachment and Evolving Internal Vertex Structure." Journal of Statistical Physics 151 (6): 1175–1183. doi:10.1007/s10955-013-0749-1.

CUTA. 2016. Canadian Transit Fact Book. Toronto, Canada: Canadian Urban Transit Association.

De Regt, Robin, Christian von Ferber, Yurij Holovatch, and Mykola Lebovka. 2019. "Public Transportation in Great Britain Viewed as a Complex Network." Transportmetrica A: Transport Science 15 (2): 722–748.

Demšar, Urška, Olga Špatenková, and Kirsi Virrantaus. 2008. "Identifying Critical Locations in a Spatial NetworkwithGraph Theory."Transactions inGIS12 (1): 61–82. doi:10.1111/j.1467-9671.2008.01086. x.

Derrible, S., and C. Kennedy. 2009. "Network Analysis of World Subway Systems Using Updated Graph Theory." Transportation Research Record 2112 (2112): 17–25. doi:10.3141/2112-03.

Derrible, S., and C. Kennedy. 2010. "The Complexity and Robustness of Metro Networks." Physica a- Statistical Mechanics and Its Applications 389 (17): 3678–3691. doi: 10.1016/j.physa.2010.04.008.

Duan, Houli, Zhiheng Li, Yi Zhang, Zuo Zhang, Danya Yao, Renjie Teng, and Yi Zhang. 2009. "Robustness Analysis of Urban Transit Networks Based on Bipartite Graph Model."IFAC Proceedings Volumes 42 (20): 246–251. doi:10.3182/20090924-3-it-4005.00042.

Erdös, P., and A. Rényi. 1960. "On the Evolution of Random Graphs." The Mathematical Institute of the Hungarian Academy of Science 5: 17–61.

Frappier, A., C. Morency, and M. Trepanier. 2018. "Measuring the Quality and Diversity of Transit Alternatives." Transport Policy 61: 51–59. doi: 10.1016/j.tranpol.2017.10.007.

Freeman, Linton C. 1978. "Centrality in Social Networks Conceptual Clarification." Social Networks 1 (3): 215–239. doi:10.1016/0378-8733(78)90021-7.

Guo, Lulu, and Guofeng Su. 2017. "Topology Vulnerability Analysis of Several Urban Metro Networks." In Proceedings of the 3rdACM SIGSPATIAL Workshop on Emergency Management Using, edited by The Association for Computing Machinery (ACM), 1–4. Redondo Beach, CA, USA: ACM.

Han, C., and L. Liu. 2009. Topological Vulnerability of Subway Networks in China. Paper presented at the 2009 International Conference on Management and Service Science, 20-22 Sept. 2009.

Hassan, M. N., Y. E. Hawas, and K. Ahmed. 2013. "A Multi-Dimensional Framework for Evaluating the Transit Service Performance." Transportation Research Part a-Policy and Practice 50: 47–61. doi: 10.1016/j.tra.2013.01.041.

Hossain, M. M., and S. Alam. 2017. "A Complex Network Approach Towards Modeling and Analysis of the Australian Airport Network." Journal of Air Transport Management 60: 1–9. doi: 10.1016/j.jairtraman.2016.12.008.

Huang, A. L., H. M. Zhang, W. Guan, Y. Yang, and G. Q. Zong. 2015. "Cascading Failures in Weighted Complex Networks of Transit Systems Based on Coupled Map Lattices." Mathematical Problems in Engineering 2015: 1–16. doi:Artn 94079510.1155/2015/940795.

Lozano, Antonio J., Juan A. Mesa, and Francisco A. Ortega. 2006. "Rapid Transit Network Robustness Indices: From the Topological to the Operational Viewpoint." In

ARRIVAL project technical report 53.

Lozano, A. J., Juan Mesa, and Francisco Ortega. 2018. Rapid Transit Network Robustness Indices: From the Topological to the Operational Viewpoint.

Mahmoud, Moataz, Ryan Garnett, Mark Ferguson, and Pavlos Kanaroglou. 2016. "Electric Buses: A Review of Alternative Powertrains." Renewable and Sustainable Energy Reviews 62: 673–684. doi: 10.1016/j.rser.2016.05.019.

Mahmoud, M., and J. Hine. 2013. "Using AHP to Measure the Perception gap Between Current and Potential Users of bus Services." Transportation Planning and Technology 36 (1): 4–23. doi:10.1080/03081060.2012.745316.

Mahmoud, M., and J. Hine. 2016. "Measuring the Influence of bus Service Quality on the Perception of Users." Transportation Planning and Technology 39 (3): 284–299. doi:10.1080/03081060.2016.1142 224.

Mahmoud, M., Julian Hine, and Anil Kashyap. 2011. "Bus transit service quality monitoring in UK: A methodological framework." Proceedings of the ITRN.

Mattsson, L. G., and E. Jenelius. 2015. "Vulnerability and Resilience of Transport Systems – A Discussion of Recent Research." Transportation Research Part a-Policy and Practice 81: 16–34. doi: 10.1016/j.tra.2015.06.002.

M'cleod, Lani, Richard Vecsler, Yuan Shi, Ekaterina Levitskaya, Sunny Kulkarni, Sergey Malinchik, and Stanislav Sobolevsky. 2017. "Vulnerability of Transportation Networks: The New York City Subway System Under Simultaneous Disruptive Events." Procedia Computer Science 119: 42–50. doi:

10.1016/j.procs.2017.11.158.

Mohamed, M., H. Farag, N. El-Taweel, and M. Ferguson. 2017. "Simulation of Electric Buses on a Full Transit Network: Operational Feasibility and Grid Impact Analysis." Electric Power Systems Research 142: 163–175. doi: 10.1016/j.epsr.2016.09.032.

Mohamed, M., C. Higgins, M. Ferguson, and P. Kanaroglou. 2016. "Identifying and Characterizing Potential Electric Vehicle Adopters in Canada: A Two-Stage Modelling Approach." Transport Policy 52: 100–112. doi: 10.1016/j.tranpol.2016.07.006.

Newman, M. E., S. H. Strogatz, and D. J. Watts. 2001. "Random Graphs with Arbitrary Degree Distributions and Their Applications." Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics 64 (2 Pt 2): 026118. doi: 10.1103/PhysRevE.64.026118.

Nian, Guangyue, Fangxi Chen, Zhe Li, Yi Zhu, and Daniel Sun. 2019. "Evaluating the Alignment of new Metro Line Considering Network Vulnerability with Passenger Ridership." Transportmetrica A: Transport Science 15 (2): 1402–1418. doi: 10.1080/23249935.2019.1599080.

Pien, K. C., K. Han, W. L. Shang, A. Majumdar, and W. Ochieng. 2015. "Robustness Analysis of the European air Traffic Network." Transportmetrica A: Transport Science 11 (9): 772–792.

Qiu, F., W. Q. Li, and J. Zhang. 2014. "A Dynamic Station Strategy to Improve the

Performance of Flex-Route Transit Services." Transportation Research Part C-Emerging Technologies 48: 229–240. doi: 10.1016/j.trc.2014.09.003.

Roth, C., S. M. Kang, M. Batty, and M. Barthelemy. 2012. "A Long-Time Limit for World Subway Networks." Journal of The Royal Society Interface 9 (75): 2540–2550. doi:10.1098/rsif.2012.0259.

Saidi, S., S. C. Wirasinghe, L. Kattan, and S. Esmaeilnejad. 2017. "A Generalized Framework for Complex Urban Rail Transit Network Analysis." Transportmetrica A: Transport Science 13 (10): 874–892. doi:10.1080/23249935.2017.1348401.

Seoane, Freire, Maria Jesus, Fernando González Laxe, and Carlos Pais Montes. 2013. "Foreland Determination for Containership and General Cargo Ports in Europe (2007 2011)." Journal of Transport Geography 30: 56–67. doi: 10.1016/j.jtrangeo.2013.03.003.

Scott, D. M., D. C. Novak, L. Aultman-Hall, and F. Guo. 2006."Network Robustness Index: A New Method for Identifying Critical Links and Evaluating the Performance of Transportation Networks." Journal of Transport Geography 14: 215–227. doi: 10.1016/j.jtrangeo.2005.10.003.

Shanmukhappa, T., I. W. H. Ho, and C. K. Tse. 2018. "Spatial Analysis of bus Transport Networks Using Network Theory." Physica a-Statistical Mechanics and Its Applications 502: 295–314. doi: 10.1016/j.physa.2018.02.111.

Sui, Y., F. J. Shao, R. C. Sun, and S. J. Li. 2012. "Space Evolution Model and Empirical Analysis of an Urban Public Transport Network." Physica a-Statistical Mechanics

and Its Applications 391 (14): 3708–3717. doi: 10.1016/j.physa.2012.01.011.

Sun, Daniel, Shukai Chen, Chun Zhang, and Suwan Shen. 2016. "A bus Route Evaluation Model Based on GIS and Super-Efficient Data Envelopment Analysis." Transportation Planning and Technology 39 (4): 407–423. doi:10.1080/03081060.2016.1160582.

Sun, D., and S. T. Guan. 2016. "Measuring Vulnerability of Urban Metro Network From Line Operation Perspective." Transportation Research Part a-Policy and Practice 94: 348–359. doi: 10.1016/j.tra.2016. 09.024.

Sun, L. S., Y. C. Huang, Y. Y. Chen, and L. Y. Yao. 2018. "Vulnerability Assessment of Urban Rail Transit Based on Multi-Static Weighted Method in Beijing, China." Transportation Research Part a-Policy and Practice 108: 12–24. doi: 10.1016/j.tra.2017.12.008.

Sun, Daniel, Yuhan Zhao, and Qing-Chang Lu. 2015. "Vulnerability Analysis of Urban Rail Transit Networks: A Case Study of Shanghai, China." Sustainability 7 (6): 6919–6936. doi:10.3390/su7066919.

von Ferber, Christian, Bertrand Berche, Taras Holovatch, and Yurij Holovatch. 2012. "A Tale of two Cities." Journal of Transportation Security 5 (3): 199–216. doi:10.1007/s12198-012-0092-9.

Wandelt, Sebastian, Xiaoqian Sun, and Jun Zhang. 2019. "Evolution of Domestic Airport Networks: a Review and Comparative Analysis." Transportmetrica B: Transport Dynamics 7 (1): 1–17. doi:10.1080/21680566.2017.1301274.

Wang, X. R., Y. Koc, S. Derrible, S. N. Ahmad, W. J. A. Pino, and R. E. Kooij. 2017. "Multi-criteria Robustness Analysis of Metro Networks." Physica a-Statistical Mechanics and Its Applications 474: 19–31. doi: 10.1016/j.physa.2017.01.072.

Wang, Shuai, and Jing Liu. 2016."Robustness of Single and Interdependent Scale-Free Interaction Networks with Various Parameters." Physica A: Statistical Mechanics and its Applications 460: 139–151. doi: 10.1016/j.physa.2016.04.035.

Watts, Duncan J., and Steven H. Strogatz. 1998. "Collective Dynamics of 'Small-World' Networks." Nature 393: 440–442. doi:10.1038/30918.

Wei, R., X. Y. Liu, Y. J. Mu, L. M. Wang, A. Golub, and S. Farber. 2017. "Evaluating Public Transit Services for Operational Efficiency and Access Equity." Journal of Transport Geography 65: 70–79. doi: 10.1016/j.jtrangeo.2017.10.010.

Wu, X. T., H. R. Dong, C. K. Tse, I. W. H. Ho, and F. C. M. Lau. 2018. "Analysis of Metro Network Performance From a Complex Network Perspective." Physica a-Statistical Mechanics and Its Applications 492: 553–563. doi: 10.1016/j.physa.2017.08.074.

Xing, Y. Y., J. Lu, S. D. Chen, and S. Dissanayake. 2017."Vulnerability Analysis of Urban Rail Transit Based on Complex Network Theory: a Case Study of Shanghai Metro." Public Transport 9 (3): 501–525. doi:10.1007/s12469-017-0170-2.

Yang, Zhirou, and Jing Liu. 2018. "Robustness of Scale-Free Networks with Various Parameters Against Cascading Failures." Physica A: Statistical Mechanics and its Applications 492: 628–638. doi: 10.1016/j.physa.2017.09.093.

Yassien, Y., M. Ezzeldin, M. Mohamed, and W. El-Dakhakhni. 2020. "Air Transportation

Infrastructure Robustness Assessment for Proactive Systemic Risk Management." Journal of Management in Engineering. doi:10.1061/(ASCE)ME.1943-5479.0000789.

Zhou, M., and J. Liu. 2017. "A Two-Phase Multi objective Evolutionary Algorithm for Enhancing the Robustness of Scale-Free Networks Against Multiple Malicious Attacks." IEEE Transactions. on Cybernetics 47 (2): 539–552. doi:10.1109/TCYB.2016.2520477.

Zhu, Z. R., A. M. Zhang, and Y. H. Zhang. 2018. "Connectivity of Intercity Passenger Transportation in China: A Multi-Modal and Network Approach." Journal of Transport Geography 71: 263–276. doi: 10.1016/j.jtrangeo.2017.05.009.

Zou, Z. Y., Y. Xiao, and J. Z. Gao. 2013. "Robustness Analysis of Urban Transit Network Based on Complex Networks Theory." Kybernetes 42 (3): 383–399. doi:10.1108/03684921311323644.

CHAPTER THREE

3. A PREDICTION MODEL FOR BATTERY ELECTRIC BUS ENERGY CONSUMPTION IN TRANSIT

Abstract: This study investigates the impacts of vehicular, operational, topological, and external parameters on the energy consumption (E_C) of battery-electric buses (BEBs) in transit operation. Furthermore, the study develops a data-driven prediction model using big data for BEB energy consumption in transit operation that considers these four parameters. A Simulink energy model is developed to estimate the E_C rates and validated using the Altoona's test real-world data. A full-factorial experiment is used to generate 907,199 scenarios for BEB operation informed by 120 real-world drive cycles. A multivariate multiple regression model was developed to predict BEB's E_C. The regression model explained more than 96% of the variation in the E_C of the BEBs. The results show the significant impacts of road grade, the initial state of charge, road condition, passenger loading, driver aggressiveness, average speed, HVAC, and stop density on BEB's energy consumption, each with a different magnitude. The study concluded that the optimal transit profile for BEB operation is associated with rolling grade and relatively lower stop density (one to two stops/km).

Keywords: *energy consumption; battery-electric buses; simulation model; full-factorial design; multiple linear regression; operational; topological; external parameters*

3.1. Introduction

The transportation system has been contributing to a high share of greenhouse gas (GHG) emissions over the past two decades (Perrotta et al. 2014; Markel et al. 2002). Currently, transportation contributes to approximately 20% of global GHG emissions. In contrast, electricity is a clean source of energy (depending on generation), with the potential of net-zero GHG emissions (Ferguson et al. 2019; Mohamed et al. 2016). In most developed countries, electricity generation is being shifted to more clean and renewable sources (Kennedy 2015). Therefore, electric mobility is considered a better alternative to reduce the transportation carbon footprint.

In this respect, the electrification of transportation systems is at the forefront of transportation researchers and transportation agencies. In particular, the bus transit system represents a very proper context to operate electric powertrain technology due to the nature of fixed routes and timely operation (Ferguson et al. 2019; Mahmoud et al. 2016). Electric bus (e-Bus) transit systems provide avenues to reduce GHG emissions and other advantages such as reducing noise and increasing energy efficiency (Borén 2020; Kühne 2010). However, before phasing out diesel buses, it is crucial for governments, transit providers, and utility companies to understand the impact of an all-electric fleet on the existing energy infrastructure (Mohamed et al. 2017; Quarles et al. 2020; Wellik et al. 2021; Mohamed et al. 2018).

In general, e-buses are classified into two main types: fuel cell e-bus (FCEB), which uses hydrogen fuel cells to generate energy on board for the electric motor; and battery e-

bus (BEB), which stores the energy on an onboard battery that supplies the electric motor with the required energy (Mahmoud et al. 2016). The latter is argued to be more economically feasible due to the relatively lower premium cost and the potential for optimal utilization of the charging process during off-peak times. Besides, BEB technology is mature compared to FCEB, which still faces several technical challenges (Mahmoud et al. 2016; Chen et al. 2007).

Implementing BEBs in transit systems requires meticulous infrastructure planning, where each bus route must have an appropriate charging scheme that adheres to the operation schedule (El-Taweel et al. 2020). In a BEB transit system, the charging stations and the on-board battery have the highest bearing on the total system cost. The cost of the on-board battery draws about one-third of the overall system cost (Quarles et al. 2020; Wellik et al. 2021; Liu et al. 2019; Abdelaty et al. 2020). Indeed, the state-of-the-art has been focusing on optimizing BEB systems and the associated infrastructure network, to minimize the total cost of ownership associated with BEB implementation in transit systems through the optimal allocation of resources (El-Taweel et al. 2020; El-Taweel et al. 2017; He et al. 2019; Rupp et al. 2020). However, this process depends on an accurate estimation of the consumed energy during operation.

Consequently, predicting an accurate energy consumption (E_C) rate is the cornerstone for all BEB studies, including those focusing on fleet optimization, battery capacity/performance, the spatial distribution of infrastructure, component sizing (battery and charger), GHG emissions, and utility grid impact analysis (Mohamed et al. 2017; Qi et al. 2018; Teoh et al. 2018; Vepsäläinen et al. 2018a; Wang et al. 2020; Zhou et al. 2016).

BEB energy consumption is sensitive to several external operational and topological parameters such as stop density, road grade, driving behaviour, traffic condition, road condition, and temperature, to name just a few (Liu et al. 2017; Basso et al. 2019; De Cauwer et al. 2015; Kivekäs et al. 2018; Vepsäläinen et al. 2019; Abdelaty and Mohamed 2020). Furthermore, each transit route exhibits a unique operational and topological profile that contributes to energy demand, and the impact of these external parameters on energy consumption requires detailed modelling.

Although models have been developed in the electric vehicles (EVs) domain to account for such impact, surprisingly, there are a limited number of studies in transit application that relates the energy consumption of BEBs to the operational, topological, and external parameters of transit networks (Abdelaty et al. 2021). To better depict the value of such an analysis, one should consider that adjusting the spacing between bus stops or route speed might yield a significant reduction in energy consumption (Vepsäläinen et al. 2018a).

Therefore, there are valid calls to study the intertwined relationships between the energy consumption of BEBs on the one hand and the vehicular, operational, topological, and external features of transit networks on the other hand. Although some models have been developed to quantify the impact of several parameters of energy model (e.g. speed, rolling resistance, mass, temperature, and frontal area) on the energy consumption (Rupp et al. 2020; Vepsäläinen et al. 2019; Kunith et al. 2017), our study contributes to understanding, and predicting, the relationship between the operational, topological, and external features of transit networks and the energy consumption of BEBs. This is curial

for policymakers and transit providers to better plan BEB adoption in transit.

Toward that end, the present study aims at quantifying the association between the vehicular, operational, topological, and external parameters of the transit network and the energy consumption of BEBs. More specifically, our goals are to (1) develop a prediction model for BEB energy consumption from the features of transit networks, which in turn (2) is used to inform the design of the optimal transit operation profile that enhances BEB energy utilization.

Following the introduction, a comprehensive review of BEB energy consumption studies is presented in Section 3.2. Section 3.3 provides details on the energy consumption simulation model and the model validation. Furthermore, it illustrates the experimental design and the specification of the proposed prediction model. The results of the proposed prediction model are presented in Section 3.4, followed by a discussion highlighting the practical implications of the proposed model. Lastly, Section 3.6 concludes the study.

3.2. BEB Energy Consumption Models

BEB energy consumption varies significantly based on numerous parameters, classified into four groups vehicular, operational, topological, and external, as illustrated in Figure 3.1.
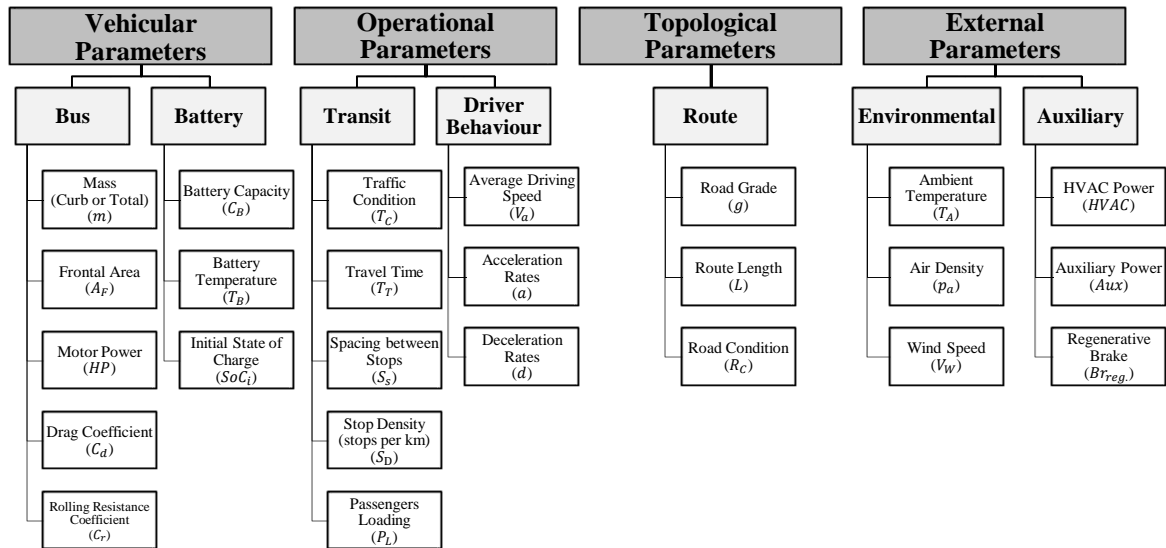
Figure 3-1. Parameters impacting the energy consumption of battery-electric buses

First, vehicular parameters include all the physical parameters related to the bus and the battery. These parameters are defined in the literature as vehicle mass (m), which is the total weight of the bus components, including the curb weight, motor, gearbox, wheels, battery, and passengers (Table 3.1) (Vepsäläinen et al. 2018a; Zhou et al. 2016; Kivekäs et al. 2018; Abdelaty and Mohamed 2020; Franca 2015; Gallet et al. 2018; Lajunen 2014; Lajunen 2018), frontal area (A_F) (Franca 2015; Lajunen 2014), drag coefficient (C_d) (Vepsäläinen et al. 2018a; Lajunen 2018), the initial status of charge (SoC_i), and battery temperature (T_B) (Vepsäläinen et al. 2018a; Vepsäläinen et al. 2019).

Second, operational parameters include the number of stops along the bus route (S_N), stop spacing (S_S) (Gallet et al. 2018; Vepsäläinen et al. 2018b), average speed (V_a), acceleration (a), and deceleration (d). Acceleration and deceleration rates are used to estimate the driving behaviour (i.e. slow, normal or aggressive) (Lajunen 2014;

Vepsäläinen et al. 2018b; Gao et al. 2017). Third, topological parameters include route features such as route length (L) and road grade (GR). Fourth, external parameters include environmental and auxiliary parameters such as air density (P_a), which depends mainly on the ambient temperature (T_A) and rolling resistance coefficient (Cr) which varies based on the road surface type and the weather (Vepsäläinen et al. 2019; Kunith et al. 2017). Furthermore, external parameters include the HVAC system, which depends on the ambient temperature such as the model developed by (Lajunen and Tammi 2016) and used by (Vepsäläinen et al. 2019), regenerative brake, and auxiliary systems such as bus doors and power steering and the hydraulic power for braking systems (Vepsäläinen et al. 2019; Gallet et al. 2018).

There are two common analyses used in the literature to study the association between the aforementioned parameters and BEB energy consumption: sensitivity analysis and regression analysis. Sensitivity analysis describes the importance of each input parameter in determining the variability of the model response. It defines how a variation in each parameter affects BEB energy consumption under a given set of assumptions and studying the uncertainty in these parameters (Diaz Alvarez et al. 2014). In contrast, regression models determine the significance, or the lack thereof, of the independent input parameters on the energy consumption (Liu et al. 2017).

Concerning sensitivity-based studies, Franca (2015) conducted a sensitivity analysis for the BEB energy consumption model using the bus mass, passenger loading, frontal area, drag coefficient, rolling resistance, drivetrain efficiency, and auxiliary power. Likewise, Basma et al. (2020) conducted a sensitivity analysis using the minimum state of

charge and the battery service life to study their impact on the BEB optimal battery design and the total energy cost, while Vepsäläinen et al. (2018b) used a surrogate modelling technique for BEB energy consumption based on the uncertainties in the weather and the operation parameters (i.e. driver aggressiveness and stops per km). Furthermore, a global sensitivity approach was used by (Kivekäs et al. 2018) to compare the impact of the number of stops and the average number of passengers on BEB energy consumption.

These studies argued that the variation in ambient temperature causes a considerable change in energy consumption due to its impact on auxiliary power and HVAC system (Franca 2015; Vepsäläinen et al. 2018b). Additionally, the variation in the rolling resistance showed the second highest impact (Franca 2015). Furthermore, passenger loading has a substantial effect on energy consumption rates (Franca 2015). Kivekäs et al. (2018) also found that the variation in the number of stops has more impact on energy consumption rates relative to passenger loading.

Although sensitivity-based studies provide a valuable contribution, sensitivity analysis falls short in accommodating the relationship between the independent parameters since it considers each parameter's impact independently (Bajtelsmit 1997; Christopher et al. 2002). As such, the model might be biased when there are many correlated parameters. Additionally, sensitivity models identify only the magnitude of the impact, not the significance. In contrast, regression models identify the relative impact of several parameters all at once, as well as their levels of significance (Yuan et al. 2017). Regression analysis allows predicting accurate models to estimate energy consumption rates (Liu et al. 2017; De Cauwer et al. 2017).

Relative to sensitivity analysis studies, a few studies employed regression models in analyzing the impact of the various parameters on the energy consumption of BEBs. Teoh et al. (2018) used a linear regression model to estimate the impact of route length and passenger loading on BEB energy consumption. They concluded that the route length and passenger loading are significant operational parameters; however, the authors argued that the operational parameters are greatly affected by various external parameters, including the headway and the charging type (e.g. normal or fast charging).

Vepsäläinen et al. (2018b) conducted a sensitivity analysis based on the results of multiple linear regression analysis to interpret the variation in the energy consumption rates due to the operation and environmental parameters. They used a linear regression analysis to determine the correlated and uncorrelated operation parameters and select the parameters that affect energy consumption. They also concluded the significance of the driver's aggressiveness and the stops per km on the consumed energy. After that, they performed the sensitivity analysis based on the selected parameters to interpret the energy consumption variation. That said, their regression model suffered from multicollinearity issues.

In addition, a deep learning network model (DLN) was developed to estimate the BEB's energy consumption (Pamula and Pamula 2020). They compared the DLN model to the results obtained from a multiple linear regression model (MLR). The comparison resulted in non-tangible differences between the two models. The results also indicated the high impact of the spacing between stops, travel time, elevation differences, and weather conditions on energy consumption.

Despite the scarcity of regression studies in the BEB domain, there is an abundance of studies that applied regression analysis in the electric vehicles (EVs) domain, as listed in Table 3.2. For example, Galvin (2017) developed a multivariate linear regression model to study the impact of driving behaviour parameters such as acceleration/deceleration rates and average speed on EVs' energy consumption, while ordinary least squares regression and multilevel mixed-effects regression models were used to assess the impact of the external parameters such as temperature and HVAC on the E_C (Liu et al. 2017; Wang et al. (2017).

Furthermore, several multiple linear regression models were developed to predict the energy consumption based on vehicular parameters such as rolling resistance and drag coefficient, topological parameters such as road grade and route length, operational parameters such as average speed and acceleration/deceleration rates, and external parameters such as temperature, auxiliary power, and HVAC (Liu et al. 2017; De Cauwer et al. 2015; Yuan et al. 2017; De Cauwer et al. 2017; Wang et al. 2017; Neaimeh et al. 2013). In this respect, the linear regression model is the dominant method utilized in the EV domain.

Table 3-1. Parameters affecting the energy consumption of BEB in the literature

	Vehicular Parameters						Operational Parameters					Topological Parameters		External Parameters			
	Mass	Frontal Area	Drag Coefficient	Rolling Resistance	Battery Temperature	Initial State of Charge	Number of Stops	Spacing Between Stops	Average Speed	Acceleration Rate	Deceleration Rate	Route Length	Road Grade	Ambient Temperature	Air Density	Auxiliary Power	HVAC
	m (ton)	A_F (m²)	C_d	C_r	T_B (C°)	SoC_i (%)	S_N	S_s (m)	V_a (Km/h)	a (m/s²)	d (m/s²)	L (Km)	g (%)	T_A (C°)	P_a (Kg/m³)	Aux (kW)	$HVAC$ (kW)
Gallet et al. (2018)	10.00 12.50 15.00 17.50 18.50	8.30 10.35	0.60 0.70 0.80	0.006 0.008 0.010	-	-	5	-	11.50 18.90 19.10 20.00 36.60	0.70 1.00 2.00	1.00 1.50 2.50	0.236 0.384 0.600 0.662 0.948	-	-	1.18	8.00 10.00 12.00 15.00	-
Vepsäläinen et al. (2019)	8.50 to 15.00	-	-	0.006 to 0.020	15 to 30	100 to 50	-	-	-	-	-	-	-	-30 to 35	-	2.00 to 7.00	2.00 to 25.00
Vepsäläinen et al. (2018a)	12.40 14.80	6.20	0.60	0.010	20	89	18	-	24.20	1.67 (max)	2.57 (max)	10.100	-	5	1.27	-	-
Kivekäs et al. (2017)	10.35 12.35 15.00	6.20	0.50	0.008	-	-	25	-	20.58	0.55 (avg.)	0.51 (avg.)	10.422	-	-	-	1.50	-
Lajunen (2014)	8.50 to 15.00	6.20	0.70	0.008	-	-	-	-	20.38	0.54 (avg.)	0.50 (avg.)	-	-	-	-	-	-
Lajunen et al. (2018)	12.70 14.65	-	-	0.008 0.012	-	-	-	-	10.90 to 41.20	0.152 to 0.336	-	3.300 to 42.400	0.00 to +9.69	-	-	6.00 9.00	-
Franca (2015)	16.92	8.90	0.66	0.008	-	-	-	-	20.23	1.50 1.30 (avg.)	2.50 2.70 (max)	3.220	-	-	-	2.50	-
Gao et al. (2017)	10.44 15.63	8.50	0.79	0.0098	-	-	-	-	5.75 11.00 13.42 19.85	3.00 4.06 4.60 6.10	4.30 4.50 5.13 5.60	0.980 6.650 6.860 10.530	-	-	-	3.75	-
Lajunen (2018)	10.00	6.20	0.60	0.010	-	-	-	99 175 344 384 714 1250	5.90 11.00 19.80 22.50 31.50 41.20	1.40 1.80 2.10 2.40 2.80 -	1.90 2.10 2.30 2.50 3.60 -	1.000 3.300 10.300 10.500 10.900 28.600	-	-	-	-	-
Kunith et al. (2017)	12.50 17.50 18.50	8.28 10.35	0.66	0.008	-	-	-	-	-	-	-	-	-	-15 to 30	1.29	2.57 to 17.06	-
Minimum	**8.50**	**6.20**	**0.50**	**0.006**	**15**	**50**	**5**	**99**	**5.75**	**0.15**	**0.50**	**0.236**	**0.000**	**-30**	**1.18**	**1.50**	**2.00**
Maximum	**20.00**	**10.35**	**0.80**	**0.020**	**30**	**100**	**25**	**1250**	**41.20**	**6.10**	**5.60**	**42.400**	**9.690**	**35**	**1.29**	**17.06**	**25.00**
Mean, (St. d)	**14.260 (3.349)**	**7.948 (1.674)**	**0.661 (0.0920)**	**0.0094 (0.0035)**	**21.667 (7.638)**	**79.667 (26.274)**	**16 (10.149)**	**494.333 (427.15)**	**20.262 (10.232)**	**2.000 (1.563)**	**-2.688 (1.504)**	**8.095 (10.486)**	**4.845 (6.852)**	**5 (28.062)**	**1.247 (0.059)**	**7.414 (5.072)**	**13.50 (16.26)**

Table 3-2. A concise list of regression models used to predict energy consumption for BEBs and EVs

	Model Parameters	**Model Type**	**Mode**
Pamula and Pamula (2020)	Spacing between Stops, Travel Time, Elevation Differences, Weather Condition	Linear Regression	Bus
Teoh et al. (2018)	Route Length, Number of Passengers	Linear Regression	Bus
Vepsäläinen et al. (2018b)	Idle Time, Driver Aggressiveness, Average speed, Stops per Km, Ambient Temperature, initial State of Charge	Linear Regression	Bus
Qi et al. (2018)	Initial State of Charge, Average Temperature, Route Length, Average Speed	Linear Regression	Vehicle
Liu et al. (2018)	HVAC, Average speed, Route Length, Ambient Temperature, Road Grade (-9% to 9%)	Ordinary Least Squares Regression Multilevel Mixed Effects Linear Regression.	Vehicle
Wang et al. (2017)		Linear Regression Multilevel Linear Regression	Vehicle
Liu et al. (2017)	Route Length, Average Speed, A/C and Heater Usage Ratio, Road Grade (-9% to 11%)	Linear Regression Three-Level Mixed Effects Models	Vehicle
De Cauwer et al. (2017)	Rolling Resistance, Aerodynamic Drag, Temperature, Auxiliary Power, Route Length, Travel Time, Acceleration/Deceleration Rates, Vehicle and Wind Speed, Elevation	Linear Regression	Vehicle
Yuan et al. (2017)	Road Grade, Acceleration/Deceleration Rates, Air Conditioning, Rolling Resistance, Aerodynamic Drag, Ambient Temperature	Linear Regression	Vehicle
Galvin (2017)	Acceleration/Deceleration Rates, Maximum Speed, Route Length, Average Speed, Travel Time	Linear Regression	Vehicle
De Cauwer et al. (2015)	Rolling resistance, Aerodynamic Drag, Auxiliary Power, Elevation, Acceleration/Deceleration Rates	Linear Regression	Vehicle
Zhang and Yao (2015)	Instantaneous Speed, Acceleration Rates, State of Charge	Linear Regression	Vehicle
Neaimeh et al. (2013)	Road Grade (-6% to 6%), Average Speed	Linear Regression	Vehicle
Badin et al. (2013)	Average Speed, Acceleration Rates, Stop Duration, Auxiliary Power, Regenerative Brake	Simulation Model resulted in Non-linear effect	Vehicle

* Note that: the parameters' terminology had been modified to maintain consistency. (For example, the stops per km is changed to the stop density).

Previous studies concluded that the average speed and acceleration rates significantly affect energy consumption (De Cauwer et al. 2017; Galvin 2017). Intuitively, air conditioning and heating systems increase the consumed energy significantly (De Cauwer et al. 2017; Liu et al. 2018). For topological parameters, the road grade shows the greatest change in energy consumption rates. Moreover, there is a significant range in the

consumed energy associated with positive and negative grades (Liu et al. 2017; De Cauwer et al. 2017; Wang et al. 2017; Liu et al. 2018). Furthermore, Qi et al. (2018) deduced that the consumed energy decreases when the trip distance or ambient temperature increases. Additionally, they argued that energy consumption does not change significantly when the initial state of charge changes.

A substantial variation was observed in the energy consumption rates for average speed less than 35 km/h, and a little variation occurred for average speeds more than 35 km/h, resulting in a non-linear relationship between the speed and the energy consumption (Neaimeh et al. 2013). The authors divided the dataset for each speed interval to overcome the nonlinearity in the regression model. Likewise, Badin et al. (2013) found a non-linear relationship between the average speed of less than 20 km/h and the energy consumption versus a linear relationship for more than 20 km/h. The authors found a high impact of the auxiliary power, driver aggressiveness, and the regenerative brake on the energy consumption rates at low speeds (less than 20 km/h), and this effect decreased with the speed increase.

Through this brief review, some research gaps are defined: (1) there is a lack of studies that quantify the impact of the transit networks' operational parameters on the energy consumption of BEBs. (2) The impact of topological parameters on energy consumption is an under-researched area. (3) The initial state of charge (SoC_i) also has a substantial impact on the operational features of the BEBs systems, which requires further assessment. (4) To the best of our knowledge, no prediction model has been developed to predict the energy consumption of BEBs based on the combined effect of vehicular,

operational, topological, and external parameters.

Therefore, the present study provides original contributions to the existing literature. (1) We developed a simulation model to estimate BEB energy consumption under different vehicular, operation, topological, and external conditions. The model has several parameters such as road gradient, road condition, and the initial state of charge that have not been previously considered together in the BEB's literature. The simulation model is calibrated to experimental BEB results. (2) We developed a prediction model using multiple linear regression to identify and predict the causal relationship between all four sets of parameters and BEB's energy consumption. (3) We utilized the prediction model to inform the optimal bus route design that enhances BEB energy utilization.

3.3. Methodology

This study adopts a four-step sequential methodology, as depicted in Figure 3.2. The following subsections explain each step in detail.
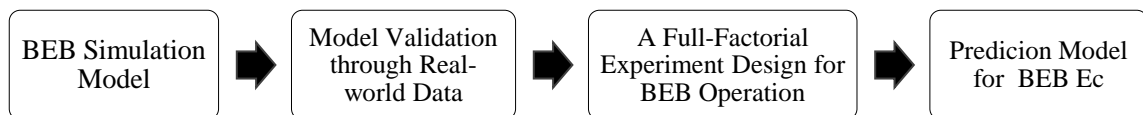
Figure 3-2. A simple flow chart of the four-step methodology

3.3.1. BEB Energy Estimation Simulation Models

Indeed, the best way to measure energy consumption for BEB is through field observation. However, collecting real-world data across all ranges of parameters is a challenging task. A large dataset is required to accommodate all possible combinations of parameters

affecting energy consumption. Therefore, the BEB literature is primarily based on the utilization of simulation models, which aim to mimic the real-world performance of BEBs under all the possible scenarios (Rupp et al. 2020; Vepsäläinen et al. 2018a; Hahn and Valentine 2019).

Toward that end, we developed a simulation model to predict BEB's energy consumption using a MATLAB Simulink platform following the approach advised by (Markel et al. 2002; Hahn and Valentine 2019), and taking into consideration the block designs used in the advanced vehicle simulator (ADVISOR). Figure 3.3 depicts the modelling blocks used for model development.
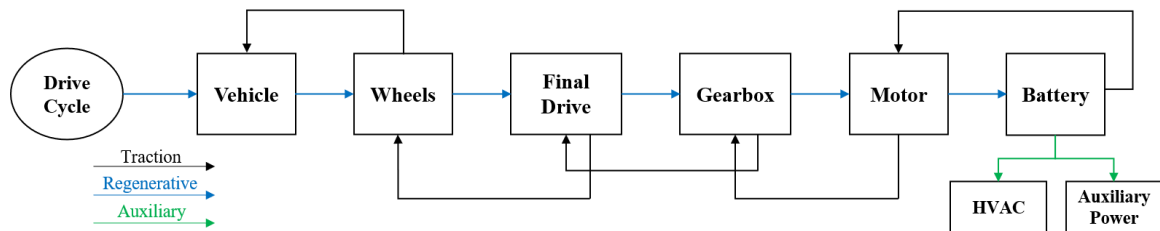
Figure 3-3. MATLAB Simulink model configuration for BEB powertrain

The required energy to propel the BEB is generated from the battery and passes through the vehicle powertrain components to the wheels. This energy is consumed during the bus's longitudinal dynamic movement and under numerous external parameters such as environmental parameters and operational parameters. Therefore, to determine the amount of energy consumed to overcome the corresponding resistances during the movement, the tractive force acting on the longitudinal dynamic movement of the bus up to wheels is calculated using Newton's second law of motion (Eq. 3.1) (Franca 2015; De Filippo et al. 2014; Rodríguez Pardo 2017; Beckers et al. 2019).

$$M.\vec{d} = \sum \vec{F} \qquad \text{Eq. 3.1}$$

The tractive force (F$_T$) should be equal or more than the summation of four resistance forces (Eqs. 3.2–3.5) that face the bus during the movement, as illustrated in Figure 3.4. These forces include: rolling resistance (F$_R$), which results from the friction between the tires and the road (Eq. 3.2); the magnitude of the rolling force, which is mainly based on the coefficient of rolling friction (C$_r$); vehicle mass (m); and gravitational acceleration (g) (Reimpell et al. 2001). Second, aerodynamic drag resistance (F$_{aero.}$) (Eq. 3.3) is the force required to overcome the air friction. The magnitude of the aerodynamic drag resistance is based on the frontal area of the bus, air density, aerodynamic drag coefficient, and the speed of the bus (Hogan and Latshaw 1973). Third, grade resistance (F$_g$), which depends mainly on the bus mass (m) and the grade of the roadway (g) (Eq. 3.4). The force required to accelerate the vehicle (F$_a$) depends on the bus mass and the acceleration/deceleration rates (Eq. 3.5). If the bus decelerates, the energy resulting from this force is either stored in the battery with regenerative braking and/or is partially lost in the braking system (Lajunen 2018; Vepsäläinen et al. 2018; Reimpell et al. 2001; Hogan and Latshaw 1973).

$$F_R = C_r.\text{V}.(m + m_{eq}).g.\sin(\alpha) \qquad \text{Eq. 3.2}$$

$$F_{aero} = 0.5.p.A_F.C_d.V^2 \qquad \text{Eq. 3.3}$$

$$F_g = (m + m_{eq}).g.\cos(\alpha) \qquad \text{Eq. 3.4}$$

$$F_a = m.(Acc.\,or\,Dec.) \qquad \text{Eq. 3.5}$$
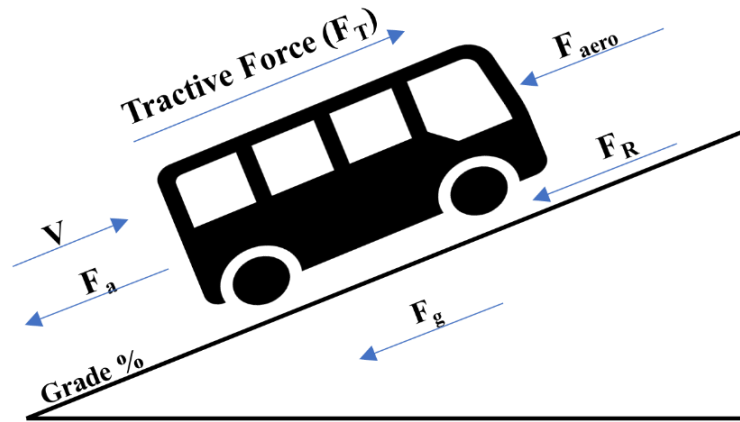
Figure 3-4. The longitudinal forces acting on the bus movements

In this respect, a backward-facing estimation approach is implemented to estimate the required energy to propel the vehicle, which considers energy losses in each powertrain component including, inverter, motor, gearbox, and the battery (De Filippo et al. 2014; Pelkmans et al. 2001).

Starting from the battery, we used the lithium-ion battery model used in the New Flyer XE40 electric bus. To perform the battery block model, we applied various equations to calculate different parameters such as the terminal voltage and the state of charge. The terminal voltage (V_t) is the difference between the open-circuit voltage and the internal resistance (R) multiplied by the current (I). The open-circuit voltage can be calculated based on the number of cells (n) and the depth of the discharge (DoD), where the DoD is the inverse of the SoC (DoD = 1 − SoC). The battery state of charge (SoC) can be estimated using the amount of electrical current in and out of the battery. The battery capacity (Q (I)) gives the amount of the current (electric charge) that the battery can deliver during the discharge state at the rated voltage (Eq. 3.8) (Gao et al. 2017); where $W_{discharge}$ is the battery

discharge power and W_{charge} is the battery charge power from regenerated kinetic energy. The battery capacity is considered as 200 kWh in the developed simulation model (NewFlyer-XE40 2017).

$$V_t = V - I.R \qquad\qquad \text{Eq. 3.6}$$

$$V = n.\ Sum(c.DoD^k) \qquad\qquad \text{Eq. 3.7}$$

$$SoC = 1 - \int_0^t (W_{discharge} - W_{charge})\,dt/Q(I) \qquad\qquad \text{Eq. 3.8}$$

Regarding the motor, we used Eq. 3.9 to calculate the motor rotational speed (W_m). The T_m is the motor torque, T_{load} is the load torque, and the J_r is the rotor inertia. The torque on the motor (T_m) is calculated using Eq. 3.10 by dividing the torque on the wheels (T_W) over the gear ratio of the gearbox (gr_{ratio}), differential gear ratio (fd_{ratio}), the efficiency of the gearbox (η_{gr}), and the efficiency of the final drive (η_{fd}) (Markel et al. 2002; Lajunen 2014; James et al. 2013). The torque on the wheels (T_W) is estimated in Eq. 3.11 by multiplying the tractive force (F_T) by the radius of the wheel (W_r) and taking into consideration the powertrain's inertial torque (T_{Wi}) (Franca 2015).

$$W_m(t) = \frac{1}{J_r}.\int (T_m - T_{load})dt \qquad\qquad \text{Eq. 3.9}$$

$$T_m(t) = \frac{T_w(t)}{gr_{ratio}.\eta_{gr}.fd_{ratio}.\eta_{fd}} + T_{mi} \qquad\qquad \text{Eq. 3.10}$$

$$T_W(t) = F_T.W_r + T_{Wi} \qquad\qquad \text{Eq. 3.11}$$

We calculate the mechanical power (P_m) by multiplying the motor torque (T_m) and the motor rotational speed (W_m), as shown in Eq. 3.12.

$$P_m(t) = T_m(t).W_m(t) \qquad\qquad \text{Eq. 3.12}$$

The electrical power is calculated by dividing the mechanical power (P_m) by the motor efficiency (η_m) (Eq. 3.13). The motor efficiency can be calculated by dividing the electric power coming from the battery (P_{in}) by the power required from the battery for the bus movement (P_{out}) (Eq. 3.14). The coefficient losses include copper losses (K_c), iron losses (K_i), and windage losses (K_w). The power required for bus movements (P_{out}) is estimated in Eq. 3.16 while taking into consideration the efficiency of the inverter and the motor (Markel et al. 2002; Franca 2015; Lajunen 2014; James et al. 2013).

$$P_e(t) = \frac{P_m(t)}{\eta_m} = V_t . I \qquad \text{Eq. 3.13}$$

$$\eta_m = \frac{P_{out}}{P_{in}} \qquad \text{Eq. 3.14}$$

$$P_{in}(t) = P_{out}(t) + P_{loss}(t) \qquad \text{Eq. 3.15}$$

$$P_{out}(t) = \frac{T_m(t).W_m(t)}{\eta_m.\eta_{inv}} + \frac{P_{aux.}}{\eta_{inv}} \qquad \text{Eq. 3.16}$$

$$P_{loss} = k_c.T_m{}^2 + k_i.W_m + k_w.W_m{}^3 + \text{Constant Power} \qquad \text{Eq. 3.17}$$

The energy required for both HVAC and auxiliary (AUX) systems is added as a function of various parameters. The HVAC system was estimated as a function of the ambient temperature and humidity following the model developed by (Lajunen and Tammi 2016) and modified by (Vepsäläinen et al. 2019). The auxiliary power (AUX), including the power for the door's air compressor, hydraulic control for the braking system, and other auxiliary devices, was assumed as a constant rate of 7 kW. This rate represents the worst-case scenario for AUX energy consumption, as reported by Vepsäläinen et al. (2019).

After developing the simulation model, a real-world BEB model is used as our base model. In this respect, we utilized the New Flyer XE40 electric bus, which is a standard 40

ft city transit bus. The input parameters for the base model, including the bus and operation parameters, are listed in Table 3.3. The charging is provided by a NewFlyer 100 kW portable depot charger. This charger charges four strings of seven lithium-ion batteries that supply power to the Siemens Model 1DB2016 drive motor and bus auxiliaries through the Siemens ELFA2 Electric Drive system.

Table 3-3. Base Model parameters for BEB energy consumption simulation

	Parameter	Value
	Battery Initial State of Charge (%)	100
	Max. Torque (N.m.)	2500
	Battery Capacity (kWh)	200
	Frontal area (m2)	8.32
	Dynamic Radius of the Tires (m)	0.5
Bus Specification	Gear Ratio	4.66
	Curb Weight (kg)	14860
	Recharge Efficiency	0.978
	Round Trip Efficiency	0.971
	Motor Efficiency	0.916
	Discharge Efficiency	0.992
	Drag Coefficient	0.6
	Rolling Resistance	0.01
Base-model Specification	Air Density (kg/m3)	1.27
	Ambient Temperature (o)	20
	HVAC (kW)	0
	Auxiliary Power (kW)*	0

* *As Auxiliary Power is not reported in Altoona Test, a value of Zero was used in the simulation model.*

3.3.2. Energy Consumption Model Validation

The validating process of BEB energy consumption models is applied, in the literature, using various methods. Essentially, the validation process is carried out using either real-world data or simulation models. In the present study, we validated our model using these two methods: real-world data and simulation mode.

3.3.2.1. Validation Using Altoona Real-World Test Results

To validate our simulation model, we utilized the New Flyer XE40 Altoona test results (Altoona 2015). The Altoona test is a real-world test for buses applied by Larson Transportation Institute's Bus Research and Testing Center in Altoona, Pennsylvania. It is being used to test buses based on their performance, reliability, and fuel economy to provide a holistic assessment of bus performance (Altoona 2015).

The Altoona test, for the New Flyer XE40 BEB, utilized three driving cycles in the test procedures, including (a) the arterial (ART) cycle, with a length of 3.073 miles that represents the high capacity urban road with speed limits between 30 mph and 50 mph; (b) the central business district (CBD) cycle with a distance of 3.073 miles, which includes 14 repetitions of a basic cycle composed of idle, acceleration, and deceleration modes with an average speed of 20 mph; and (c) the commuter (COM) cycle, which represents urban cycles without any stops during the trip with an average speed of 40 mph (Altoona 2015). We have extracted and used the three drive cycles (Appendix 3.1) to validate the simulation model, which resulted in estimated energy consumption within ±3.81% (0.0381 kWh/km) accuracy (Figure 3.5). This is consistent with the results of (De Filippo et al. 2014). More specifically, the differences in energy consumption between the Altoona test and our simulation model are 0.043 kWh/km (3%) in the ART cycle, 0.034 kWh/km (3.16%) in the CBD cycle, and 0.049 kWh/km (5.28%) in the COM cycle.
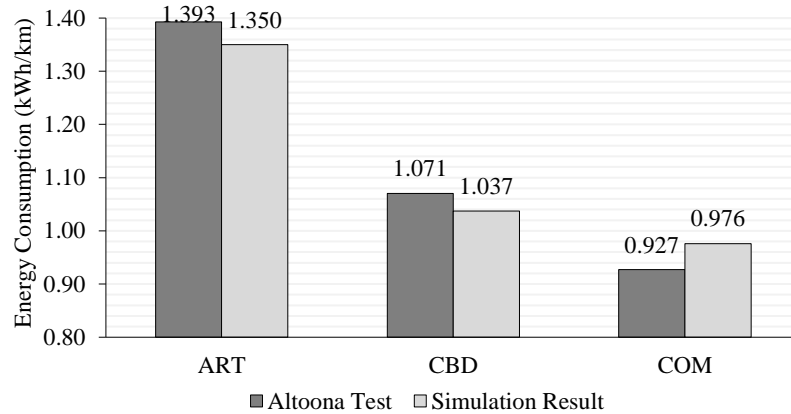
Figure 3-5. Energy consumption in the Altoona test compared to the developed simulation model

3.3.2.2. Validation Using Autonomie Software

Although the model validation through real-world Altoona test data provides a robust method to validate the model, some might argue that additional validation is required to cover the wide range of parameters used in the present study. Therefore, we followed the approach utilized by Gao et al. (2017) and validated our model using Autonomie software. The Nova Bus model in Autonomie was reconfigured to include the powertrain parameters of the New Flyer XE40 BEB (Table 3.3). This reconfigured model was tested across three bus cycles; Manhattan, New York, and RTE. The Autonomie predicted energy consumption is compared with our developed Simulink model across the same three cycles. The Autonomie validation process indicates only 6% error (simulated SoC minus Autonomie SoC), as depicted in Appendix 3.2.

3.3.3. Full-Factorial Experiment Design

We designed a full-factorial experiment (i.e. all possible combinations) to generate scenarios for BEB operation and study the impacts of all parameters on BEB's energy

consumption. The experimental design included vehicular, operational, topological, and external parameters on different drive cycles. Although some might argue that the developed scenarios might reflect extreme conditions, their inclusion is fundamental for model development.

The parameters, and their levels, are defined based on the literature, as detailed in Table 3.4. We also collected field data and speed profiles for eight different bus routes with different terrains (using a high-definition GPS device) to ensure the diversity of the parameters such as road grade, speed limit, and spacing between stops. Since bus routes are inherently a mix of positive and negative gradients, it would not be accurate to model the effect of average grade on energy consumption. Therefore, we divided the eight-speed profiles into 120 drive cycles (i.e. segments) to accommodate the impact of parameters such as road gradient, average speed, acceleration rates, and deceleration rates on the energy consumption. Therefore, we obtained a constant road gradient for each drive cycle, following the assumptions of (Liu et al. 2017; Sagaama et al. 2020) to overcome the uncertainties in using the average road gradient.

Table 3-4. List of input parameters used in full-factorial experiment design

Parameter – Unit	Parameter	Levels	Min.	Max.	Mean	Standard Deviation
Initial state of charge (SoC_i) – %		4	40	100	70.00	22.36
Mass (m) – Kg	Vehicular and	5	14932	20557	17482	1975.792
HVAC – kW	External	6	1.25	13.75	6.24	4.66
Rolling resistance (C_r) – Unitless		3	0.006	0.02	0.0120	0.0059
Road grade (g) – %		7	- 6	+ 6	0.00	4.00
Average speed (V_a) – Km/h		120	19.99	49.95	31.96	6.81
Maximum speed (V_m) – Km/h	Operational	120	31.60	72.91	51.67	13.66
Acceleration rates (a) – m/s^2	and	6	0.5	2.5	1.38	0.70
Deceleration rates (d) – m/s^2	Topological	6	1	4	2.33	0.99
Spacing between stops (S_S) – m		4	300	600	450.00	111.80
Cycle Length (L) – m		120	860.54	1817.000	1347.89	332.35
Total number of unique scenarios						**907,199**

Regarding vehicular and external parameters, we include four levels for the initial state of charge (40% till 100%, with 20% intervals). The total mass of the bus is calculated based on the curb weight and the number of passengers. The number of passengers is divided into five levels, where the minimum is zero passengers, while the maximum onboard capacity is 75 passengers. An average passenger weight (75 kg) is considered (Lajunen et al. 2018). The HVAC power is estimated based on the ambient temperature following the model developed by (Lajunen and Tammi 2016) and modified by (Vepsäläinen et al. 2019). The ambient temperature (six levels) used to estimate the HVAC power has been considered to be between -20 °C and 30 °C based on the historical Canadian weather dataset (Weather Canada 2019). The rolling resistance coefficient is considered based on the literature (three levels ranging from 0.006 to 0.02) (Vepsäläinen et al. 2019; Gallet et al. 2018; Lajunen 2018).

Operational and topological parameters include the minimum and maximum rates for acceleration and deceleration, selected based on the literature and field observation, and represent both average and aggressive driving behaviours (Gallet et al. 2018; Lajunen 2018; Kontou and Miles 2015). The rates varied between 0.50 m/s^2 and 2.5 m/s^2 for acceleration and 1 m/s^2 to 4 m/s^2 for deceleration, while the average speed is derived from the driving cycles. Furthermore, we used seven constant gradients levels ranging from -6% to 6%.

A total of 907,199 scenarios have been generated from all the possible combinations of the levels of the utilized parameters, using the full-factorial experimental design. We coded a loop in MATLAB to calculate the energy consumption rates for all generated scenarios, as shown in Figure 3.6. In the estimation process, only one parameter was

changed, while all other parameters were fixed.

3.3.4. Prediction Model

Based on the inputs of the model, new parameters have been defined. The new parameters include passenger loading, which represents the number of passengers during the trip. The road condition includes three levels I, II, and III based on the utilized rolling resistance coefficient and taking into consideration a constant bus tire pressure. Level I refers to a good dry road condition that includes the rolling resistance coefficients ≤ 0.006. Level II refers to a fair wet road condition that includes rolling resistance coefficients ≥ 0.01 and < 0.02, while level III refers to a poor icy road condition (slush) that includes rolling resistance coefficients ≥ 0.02.
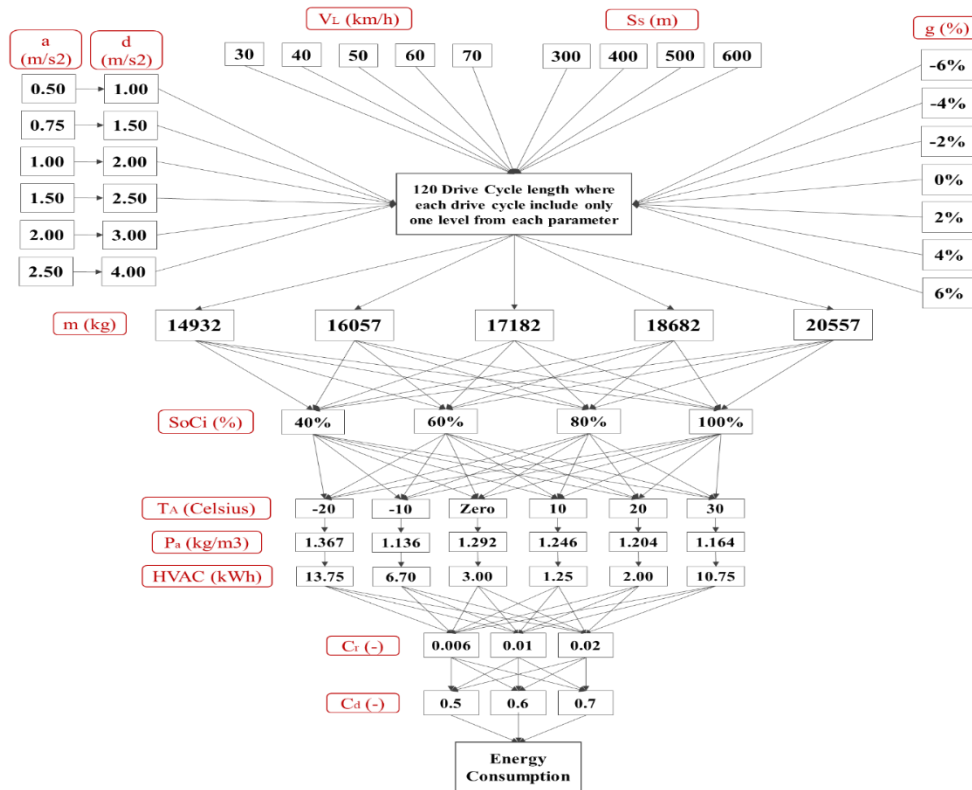
Figure 3-6. Full-factorial experimental design parameters (907,199 scenarios)

The driver's aggressiveness is defined based on the acceleration and deceleration rates of the bus drivers and divided into three levels. These levels are based on coupling six acceleration values with six deceleration values, resulting in a total of six pairs, which are grouped into three different levels, as depicted in Figure 3.6. Level I represents the slow driving behaviour with acceleration rates from 0.25 m/s^2 to 0.5 m/s^2 and deceleration rates from 1 m/s^2 and 1.5 m/s^2; level II represents the normal driving behaviour with acceleration rates from 1 m/s^2 to 1.5 m/s^2 and deceleration rates from 2 m/s^2 and 2.5 m/s^2; level III represents the aggressive driving behaviour with acceleration rates from 2 m/s^2 to 2.5 m/s^2 and deceleration rates from 3 m/s^2 and 4 m/s^2. In addition, a stop density parameter has been calculated by dividing the number of stops by the cycle length.

A multiple linear regression analysis (MLR) was applied to develop a prediction model for predicting the energy consumption of the BEBs. Besides, it quantifies the relationship between several independent parameters and a dependent parameter by fitting a linear equation to the observed dataset that we generated using the full-factorial experiment.

Given that the aim is to predict BEB energy consumption from vehicular, operational, topological, and external parameters, we used energy consumption (E_C) as the dependant parameter. The independent parameters included the road grade (GR), driver aggressiveness (D_{Agg}), road condition (R_C), HVAC, passenger loading (P_L), stop density (S_D), average speed (V_a), the initial state of charge (SoC_i), and route length (L), as depicted in Eq. 3.18. Our model is based on predicting the E_C in transit operation; thus, it can predict the E_C for micro trips.

$$E_C = \beta_0 + \beta_1 g + \beta_2 D_{Agg} + \beta_3 R_C + \beta_4 HVAC + \beta_5 P_L + \beta_6 S_D + \beta_7 V_a + \beta_8 SoC_i + \beta_9 L + \varepsilon$$

Eq. 3.18

where:

GR is the road grade (%),

D_{Agg} is the driver aggressiveness (three levels),

R_C is the road condition (three levels),

HVAC is the consumed energy due to heating, ventilation, and air conditioning (kW),

P_L is the passenger loading (passengers),

S_D is the stop density ratio along the route (stops/km),

V_a is the average speed during the trip (km/h),

SoC_i is the initial state of the battery charge (%),

L is the route length (m).

A set of analyses has been carried out to evaluate the appropriateness for the model specification, including linearity, normality, homoscedasticity, and the outliers as recommended by (Joseph et al. 2009). The results of these analyses confirm the model specification.

3.4. Results

3.4.1. Descriptive Statistics

The descriptive statistics listed in Table 3.5 show the statistical properties of the parameters utilized in the model. The resultant energy consumption rates ranged between -2.490 kWh/km and 6.119 kWh/km with a mean of 1.654 kWh/km. It should be noted that negative energy consumption values are attributed to scenarios that include a constant negative

grade. As such, the energy harnessed from the regenerative brake is higher than the energy

consumed by the bus. The average number of passengers is approximately 38 passengers,

while the stop density ranged between 1.651 stops/km and 3.486 stops/km. The driver

aggressiveness and the road condition have a mean of 2.0 and standard deviation of 0.816,

since both are divided into three levels.

Table 3-5. Descriptive statistics of the dataset (n=907,199)

Parameter	Minimum	Maximum	Mean	Standard Error	Standard deviation	Skewness	Skewness Std. Error	Kurtosis	Kurtosis Std. Error
E_C - (kWh/km)	-2.490	6.119	1.654	0.002	1.609	0.055	0.003	-0.890	0.005
g - (%)	-6.000	6.000	0.000	0.042	4.000	0.000	0.003	-1.250	0.005
SoC_i - (%)	40.000	100.000	70.000	0.000235	22.361	0.000	0.003	-1.360	0.005
R_C - (level)	1.000	3.000	2.000	0.001	0.816	0.000	0.003	-1.500	0.005
P_L - (passenger)	0.000	75.000	34.000	0.028	26.344	0.293	0.003	-1.190	0.005
$D_{Agg.}$ - (level)	1.000	3.000	2.000	0.001	0.816	0.000	0.003	-1.500	0.005
V_a - (km/h)	19.986	49.954	31.964	0.007	6.807	0.459	0.003	-0.559	0.005
S_D - (Stop/km)	1.651	3.486	2.377	0.0006	0.625	0.477	0.003	-1.194	0.005
HVAC - (kW)	1.250	13.750	6.242	0.005	4.661	0.461	0.003	-1.359	0.005
L - (m)	860.536	1817.000	1347.891	0.350	333.354	0.006	0.003	-1.353	0.005

The skewness and kurtosis are computed to test the normality of the data. The

skewness values between 0.5 and -0.5 show the normal distribution of the utilized

parameters (George and Mallery 2010). The skewness values closer to zero indicate that

the distribution is approximately symmetric, as shown in Table 3.5. The positive skewness

values for V_a, P_L, S_D, and HVAC indicate that the distribution is highly skewed to the left.

Kurtosis between 2.0 and -2.0 is considered acceptable (George and Mallery 2010).

The correlation coefficients (Table 3.6) show significant correlations between the

E_C and the independent parameters. The results also indicate the lack of significant

correlation between the independent parameters, except for route length and stop density (-

0.979). Therefore, we excluded the route length parameter from the analysis. This was

further assessed using statistical measures of multicollinearity (reported in the Results section) (Hair er al. 2010).

Table 3-6. The correlation matrix for the utilized parameters

	E_C	g	SoC_i	R_C	P_L	$D_{Agg.}$	V_a	S_D	HVAC	L
E_C	1.000									
g	0.945^{**}	1.000								
SoC_i	0.172^{**}	0.000	1.000							
R_C	0.132^{**}	0.000	0.000	1.000						
P_L	0.078^{**}	0.000	0.000	0.000	1.000					
$D_{Agg.}$	0.049^{**}	0.000	0.000	0.000	0.000	1.000				
V_a	0.027^{**}	0.000	0.000	0.000	0.000	0.491^{**}	1.000			
S_D	0.039^{**}	0.000	0.000	0.000	0.000	0.025^{**}	-0.388^{*}	1.000		
HVAC	0.105^{**}	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	
L	-0.037^{**}	0.000	0.000	0.000	0.000	-0.023^{**}	0.389^{**}	$\mathbf{-0.979^{**}}$	0.000	1.000

* and ** refer to significance at the 90% and 95% confidence levels, respectively.

3.4.2. BEB Energy Consumption Prediction Model

Linearity, normality, multicollinearity, and homoscedasticity were evaluated to verify the multivariate statistical assumptions. The linearity test confirmed the linear relationship between the dependent and independent parameters through a high coefficient of determination (R^2) for all the proposed regression models. Normality was confirmed through the errors between observed and predicted values (i.e. residuals), where a mean of 1.34×10^{-12} and a standard deviation of 1.00 were obtained. Moreover, the linearity and normality of the model are tested graphically, as shown in Figure 3.7.
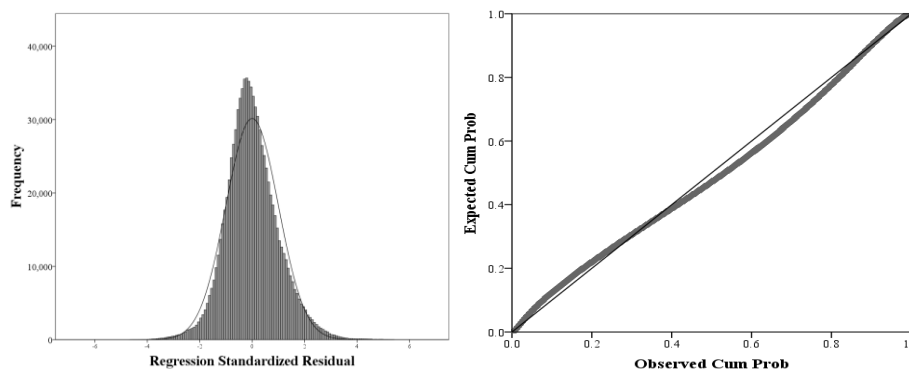
Figure 3-7. The distribution of the regression standardized residuals

For multicollinearity, the magnitude of the correlation coefficients was less than 0.50 (Table 3.6), the variance inflation factor (VIF) values varied between 1.0 and 1.67 ($<$ 5), and tolerance ranged between 0.6 and 1.0 (> 0.2), as shown in Table 3.7. These measures confirm the absence of multicollinearity between the parameters (James et al. 2013). Finally, the homoscedasticity results indicate that the variance of error terms is similar across the independent parameters, given the absence of a systematic pattern of errors in the distribution in Figure 3.8 (1% of the data) and in Appendix 3.3 (100% of the data).

The model goodness-of-fit is evaluated by inspecting the residuals, the coefficient of determination (R^2) (Table 3.7), outliers (Mahalanobis distance), and chi-square tests. The selected model shows a superior relationship between the dependent and independent parameters with an R^2 of 0.961. The proposed model explained 96.1% of the variance in energy consumption rates. The t-test values exceeded ±1.96, indicating a significant difference in the effect of the independent parameters on the E_C. Furthermore, the root means square error (RMSE) indicates the absolute fit of the model's predicted values to the observed dataset. The RMSE value of 0.3167 reflects an accurate model.

Table 3-7. Results of the energy consumption regression model

Parameters	Unstandardized Coefficients		Standardized Coefficients	t-test	Sig.	Collinearity Statistics		Adjusted R Square	RMSE
	B	Std. Error	Beta			Tolerance	VIF		
(Constant)	-0.782	0.003	-	-259.778	0.000	-	-		
g	0.380	0.008	0.945	4572.399	0.000	1.000	1.000		
SoC_i	0.0124	0.001	0.172	834.537	0.000	1.000	1.000	0.961	0.3167
R_C	0.260	0.000	0.132	638.564	0.000	1.000	1.000		
HVAC	0.036	0.000	0.105	506.847	0.000	1.000	1.000		
P_L	0.005	0.000	0.078	379.674	0.000	1.000	1.000		

$D_{Agg.}$	0.065	0.000	0.033	133.459	0.000	0.704	1.420
S_D	0.128	0.001	0.050	213.328	0.000	0.788	1.269
V_a	0.007	0.000	0.030	111.380	0.000	0.599	1.670
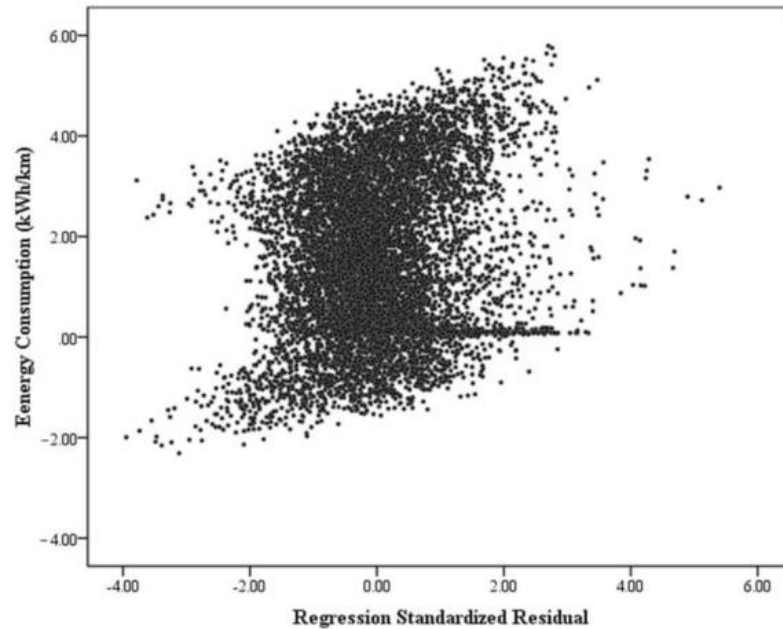
Figure 3-8. A scatter plot for the regression standardized residuals (n = 9071, 1% of the data)

Several relationships between the independent parameters and the dependent parameter (E_C) are depicted in Figure 3.9. For each 3D plot, we fixed some independent parameters (bold top legend) while the remaining parameters varied. That said, the figure depicts the significant change in the energy consumption rates (at the bivariate level) resulting from the change in vehicular, operational, topological, and external parameters.

The standardized coefficients (betas) are used to compare the impact of each independent parameter on the energy consumption. Accordingly, road grade has a substantial impact on the E_C (Std. β = 0.945), followed by the initial state of charge (Std. β = 0.172), road condition (Std. β = 0.132), HVAC (Std. β = 0.105), passenger loading (Std.

$\beta = 0.078$), driver aggressiveness (Std. $\beta = 0.033$), and stop density (Std. $\beta = 0.050$), while the average speed has the lowest weight in affecting the E_C (Std. $\beta = 0.030$). The prediction model is expressed mathematically in Eq. 3.19.

$$E_C = -0.782 + 0.380g + 0.0124SoC_i + 0.260R_C + 0.036HVAC + 0.005P_L +$$

$$0.065D_{Agg} + 0.128S_D + 0.007V_a + \varepsilon \qquad\qquad \text{Eq. 3.19}$$
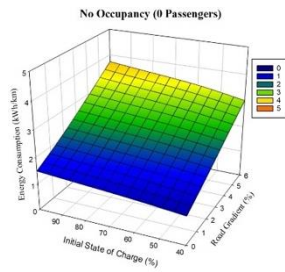
(a) Initial state of charge vs. road gradient at zero bus occupancy
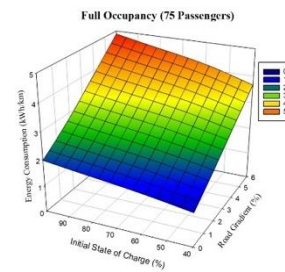
(b) Initial state of charge vs. road gradient at full bus occupancy
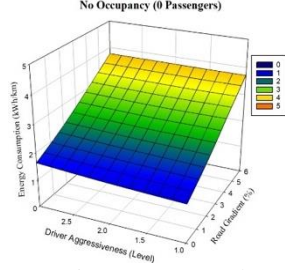
(c) Driver aggressiveness vs. road gradient at zero bus occupancy
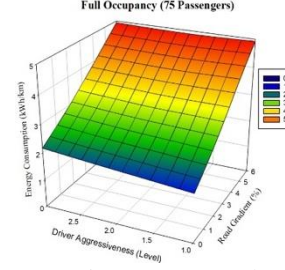
(d) Driver aggressiveness vs. road gradient at full bus occupancy
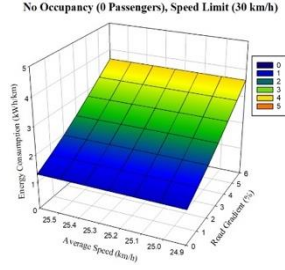
(e) Average speed vs. road gradient at zero bus occupancy and speed limit of 30 km/h
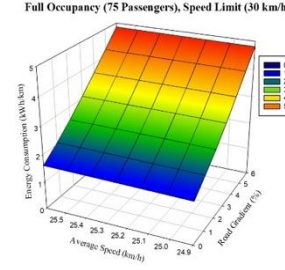
(f) Average speed vs. road gradient at full bus occupancy and speed limit of 30 km/h

(g) Average speed vs. road gradient at zero bus occupancy and speed limit of 70 km/h

(h) Average speed vs. road gradient at full bus occupancy and speed limit of 70 km/h

(i) Stop density vs. road gradient at zero bus occupancy and stop spacing of 300 m

(j) Stop density vs. road gradient at full bus occupancy and stop spacing of 300 m

(k) Stop density vs. road gradient at zero bus occupancy and stop spacing of 600 m

(l) Stop density vs. road gradient at full bus occupancy and stop spacing of 600 m

Figure 3-9. 3D plots for the relationships between E_C and the independent parameters

Put another way, the increase in the road grade by 1% increases the E_C by 0.380 kWh/km, while an increase in the initial state of charge by 10% increases the E_C by 0.124 kWh/km. The variation in the road condition and driver aggressiveness from a level to another level affects the E_C rates by 0.260 kWh/km and 0.065 kWh/km, respectively. The utilization of the HVAC system, which is attributed to temperature, has a significant bearing on the E_C, while each 10 km/h increase in the average speed increases the E_C rates by 0.07 kWh/km. Similarly, each increase in the stop density by one stop per km increases

the E_C rates by 0.128 kWh/km. An increase in the number of passengers by 10 passengers increases the consumed energy by 0.05 kWh/km, while an increase in HVAC by 1 kW increases the energy consumption by 0.036 kWh/km.

3.5. Discussion and Practical Relevance

3.5.1. Discussion of the Results

The estimated coefficients resulting from the multiple regression model show the significance of vehicular, operational, topological, and external parameters on the BEB's consumed energy. Comparing our findings to previous studies shows that our model produced more significant parameters, as shown in Figure 3.10. The figure shows the normalized relative weight (per study) associated with the significant parameters reported to impact energy consumption. It considers the absolute impact ratio of the parameters without considering the sign to compare the results from regression models and sensitivity analysis.

HVAC ■ **Average Speed** ■ **Initial State of Charge** ■ **Grade** ■ **Road Condition** ■ **Driver Agressiveness** ■ **Stop Density** ■ **Passenger Loading** ■ **Route Length**

* Note that: the parameters' terminology had been modified to maintain the terminologies consistency. (For example, the stops per km had been changed to the stop density).
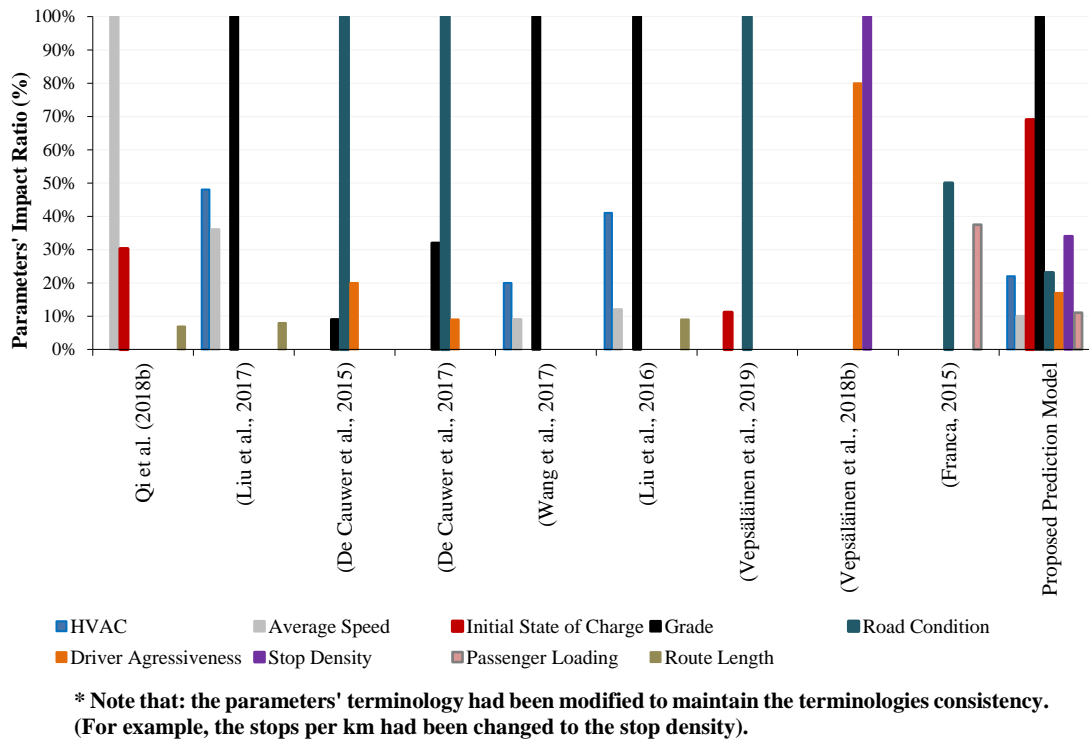
Figure 3-10. The relative weight (normalized per study) of significant parameters impacting electric powertrains energy consumption

The findings show that the main parameter driving the variation in the E_C rates is the road grade. Considering that no previous study has investigated the impact of the road grade on the E_C in the BEB literature, we compared our results with those in the EV literature. In this respect, Liu et al. (2017), Wang et al. (2017), and Liu et al. (2016) reported that road grade has the highest impact on E_C, and it increases almost linearly with increasing the absolute gradient.

The road condition and the initial state of the charge are the second and third most significant parameters, respectively. Both parameters have a positive relationship with the E_C rates, consistent with (Vepsäläinen et al. 2019). In contrast, Qi et al. (2018) found a negative relationship between the E_C and the initial state of charge for EVs. However, Qi

et al. (2018) argued that the state of charge is not significant in predicting the E_C, yet it has a marginal impact on E_C. Moreover, Vepsäläinen et al. (2019) indicated that the relation between energy consumption and the state of charge (less than 85%) is a fuzzy relationship. At the same time, the state of charge of more than 85% causes a lower energy recovery and therefore results in higher total energy consumption, which is in line with our results. We claim that the main reason for the high positive impact of the initial state of charge (SoC_i) on the E_C in the BEBs context is because our model is based on predicting the E_C in transit operation, which exhibits micro trips. This might be the reason for the level of significance associated with the initial state of charge, especially in situations where the micro trip features a negative grade coupled with a full batter (i.e. SoC_i around 100%). As such, the recovered energy from the regenerative break is marginal.

Regarding HVAC, the relationship between the ambient temperature and the consumed power is identified to have a non-linear relationship according to (Melaina et al. 2016) and the model developed by (Lajunen and Tammi 2016). To overcome this problem, the relationship between the ambient temperature and the HVAC power should be linearized around a specific value, as recommended by (De Cauwer et al. 2015; Vepsäläinen et al. 2018b). Our model was capable of capturing the significant impact of HVAC on E_C by using the HVAC power as an input in our prediction model instead of using the ambient temperature.

Driver aggressiveness has a slightly low bearing, yet significant, on E_C. De Cauwer et al. (2015), Vepsäläinen et al. (2018b), and De Cauwer et al. (2017) found a positive linear relationship between the driver aggressiveness and the E_C. They concluded that the impact

of driver aggressiveness on the E_C is lower than the impact of the road condition, which supports our findings. Similarly, the increase in passenger loading causes a slight increase in E_C rates. Franca (2015) reported similar findings for BEBs.

Stop density has a positive relationship with the E_C, indicating that increasing the number of stops contributes to higher energy consumption rates, which is in line with the results of (Vepsäläinen et al. 2018b), who stated a high positive linear relationship between the stop density and the energy consumption of BEB.

Comparing our model to the regression models available in the literature, our model has a higher goodness-of-fit than the previous models estimated by (Vepsäläinen et al. 2018b; Pamula and Pamula 2020) and includes more significant parameters in the same model. The impact of road grade as a topological parameter in estimating the energy consumption is more representative than using the elevation difference mentioned by (Pamula and Pamula 2020). This is because using the elevation difference between points results in ignoring the distance between these points.

Overall, our study spans to cover energy consumption rates for various possible operation scenarios. This, in turn, mimics the real-world operation of bus transit systems, where the values of the utilized parameters are continuously changing during the bus trips. However, it should be noted that some additional variables might impact the energy consumption of BEBs that are not included in the present study.

3.5.2. Practical Implications

We aimed to develop a prediction model for BEB energy consumption that incorporates

vehicular, operational, topological, and external parameters. Although this aim is achieved, the developed model holds significant practical implications to inform transit planners and decision-makers on the electrification of transit systems. These are articulated across two dimensions: BEB route selection and practical relevance.

First, we developed several hypothetical scenarios to inform transit planners on the best, moderate, and worst transit operational profile that enhances BEB energy utilization. The scenario development is based only on parameters associated with transit operation and design. Therefore, these scenarios incorporate average speed, stop density, passenger loading, driver aggressiveness, HVAC, and the initial state of charge, while they share a constant rolling grade (0%) and road condition (dry) (Table 3.8).

Table 3-8. Energy consumption scenarios for BEB route selection

Parameter	Unit	Coefficient	Scenario I	Scenario II	Scenario III	Scenario IV	Scenario V
Road Grade	%	0.380	0	0	0	0	0
Initial State of Charge	%	0.012	40	50	50	75	100
Road Condition	Level	0.260	3	3	3	3	3
HVAC (Temperature)	kWh (C°)	0.036	1.25 (20 °)	6.70 (-10 °)	6.70 (-10 °)	13.75 (-20 °)	13.75 (-20 °)
Passenger loading	Passenger	0.005	25	25	50	75	75
Driver Aggressiveness	Level	0.065	1	1	2	2	3
Stop Density	Stops/km	0.128	2	2.5	3	3	4
Average Speed	Km/h	0.007	20	20	30	30	40
Energy Consumption	kWh/km	Dependant	1.109	1.489	1.813	2.492	3.055
Difference (relative to Scenario I) %			**NA**	**134.28%**	**163.50%**	**224.71%**	**275.47%**

The scenario analysis indicates the significant bearing of transit network characteristics on the energy consumption of BEBs, which ranges from 1.109 to 3.055 kWh/km. Therefore, and considering operational parameters only, we encourage transit planners to implement BEBs on routes that feature lower stop density (one to two

stops/km), coupled with higher traffic level of service (i.e. LoS A and B). The aim is to reduce the frequency of buses coming to a complete stop during operation. This recommendation is to enhance the energy utilization of BEBs, but not to say that BEBs are not feasible in routes with higher stop densities.

On the other hand, electrifying routes with higher average speed and/or low passenger loads would not significantly increase energy savings. That said, we also recommend transit planners to pay attention to road grade as the most significant parameter impacting BEB energy consumption. Second, and concerning practical relevance, the proposed model could be implemented to quickly and efficiently model a transit network's energy consumption without the need for sophisticated and technically advanced simulation models. Each route could be divided into several segments with similar characteristics (e.g. speed, stop density, and passenger loading) used as inputs in the proposed prediction model to estimate BEBs energy consumption. This would be very beneficial while planning for the electrification of transit networks.

3.6. Conclusions

The accurate estimation of BEB energy consumption is a challenging task that requires laborious work of identifying the details of the BEB transit networks and how they affect the consumed energy. In this respect, the present study reveals the causal relationship between the transit network parameters, including vehicular, operational, topological, and external parameters and BEB's energy consumption, using multiple linear regression analysis. Furthermore, the study presents a prediction model for BEB energy consumption

to inform the optimal bus route design that intensifies the BEB energy efficiency.

A simulation model had been developed to predict the energy consumption rates using MATLAB Simulink. The developed model is validated using New Flyer XE40 Altoona test results. The validation process shows very promising results with ±5% accuracy, which confirms the validity of the developed simulation model with respect to the range of parameters tested in Altoona cycles.

Moreover, we generated BEB energy consumption data using a full-factorial experiment and based on real-world data collection (907,199 scenarios). A multiple linear regression model (MLR) was developed from the selected scenarios to predict the relationship between the independent vehicular, operational, topological, and external parameters and the dependent energy consumption parameter.

The results reveal a significant relationship between the BEB energy consumption and the independent parameters, including the road grade, the initial state of charge, road condition, HVAC, passenger loading, driver aggressiveness, average speed, and stop density. Besides, the estimated coefficients show that the main parameter driving the variation in the energy consumption rates is the road grade, while the stop density had a lower impact. The validity of the prediction model was verified using the goodness-of-fit, which shows that the prediction model explains about 96.1% of the variance in energy consumption. The prediction model was validated using a second dataset of 169,344, showing a very accurate E_C prediction.

Furthermore, we developed five hypothetical scenarios to inform the optimal transit

operation profile design that improves energy efficiency. We encourage transit planners to pay attention to the road grade while planning the bus routes electrification since it is the most significant parameter that affects the BEB's energy consumption. Additionally, routes with lower stop density should be considered for transit electrification. Simultaneously, the average speed and passenger loading do not have a considerable bearing, but a significant one, on the consumed energy compared to other parameters.

Although our results are in line with previous studies related to BEBs and EVs energy consumption, our study has some limitations. First, the regenerative braking energy recovery is considered in the energy consumption estimation. Yet, we did not study the impact of operational parameters on the change of the regenerative braking energy recovery. Additionally, in our model, auxiliary power was used at a constant rate, which is sensitive to the operation conditions (e.g. boarding/dwelling). Therefore, we encourage future studies to accommodate these limitations.

3.7. Acknowledgement

3.8. Appendices

Appendix 3-1. Altoona Speed Profile
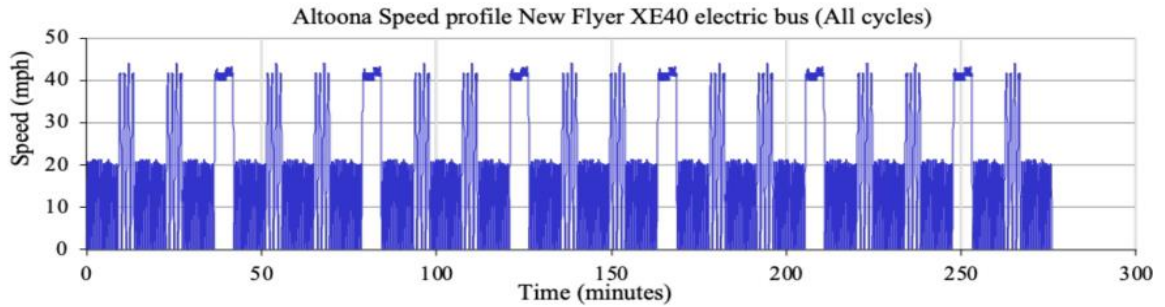
Appendix 3-2. Autonomie-Based Validation of the Developed Simulink Model

Drive Cycle	Drive Cycles Comparison	SoC Comparison	Δ SoC (%)
Manhattan			6.05%
New York Bus			6.29%
RTE			6.26%

Appendix 3-3. A Scatter Plot for the Regression Standardized Residuals (n=907,199)
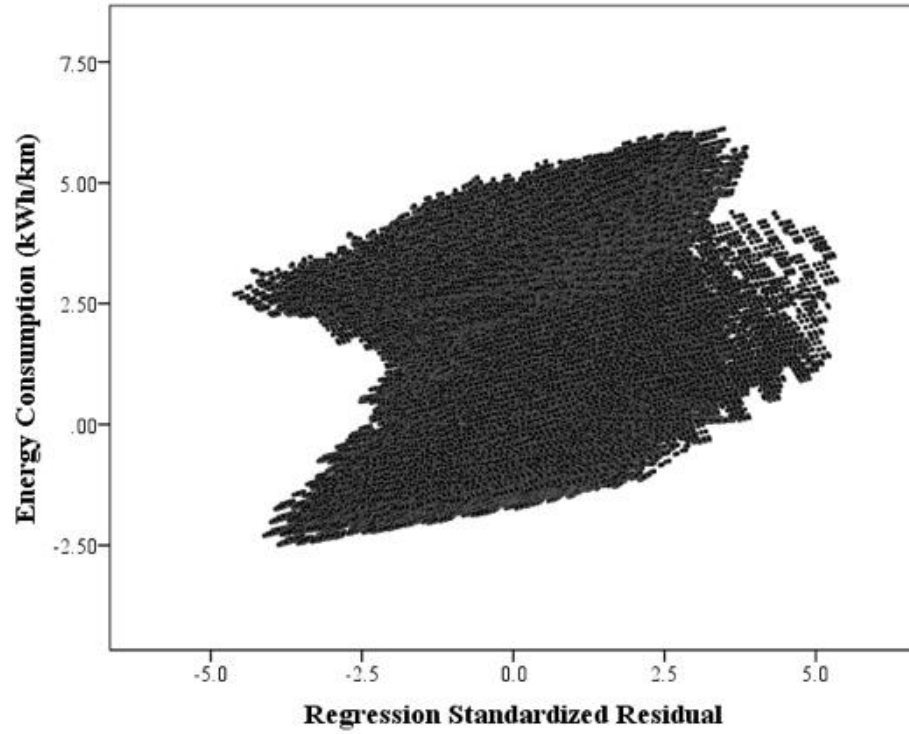
3.9. References

Perrotta, D.; Macedo, J.L.; Rossetti, R.J.; De Sousa, J.F.; Kokkinogenis, Z.; Ribeiro, B. Afonso, J.L. Route Planning for Electric Buses: A Case Study in Oporto. Procedia Soc. Behav. Sci. 2014, 111, 1004–1014.

Markel, T.; Brooker, A.; Hendricks, T.; Johnson, V.; Kelly, K.; Kramer, B.; O'Keefe, M.; Sprik, S.; Wipke, K. ADVISOR: A systems analysis tool for advanced vehicle modeling. J. Power Sources 2002, 110, 255–266.

Ferguson, M.; Mohamed, M.; Maoh, H. On the Electrification of Canada's Vehicular Fleets: National-scale analysis shows that mindsets matter. IEEE Electrif. Mag. 2019, 7, 55–65.

Mohamed, M.; Higgins, C.; Ferguson, M.; Kanaroglou, P. Identifying and characterizing potential electric vehicle adopters in Canada: A two-stage modelling approach. Transp. Policy 2016, 52, 100–112.

Kennedy, C. Key threshold for electricity emissions. Nat. Clim. Chang. 2015, 5, 179.181.

Mahmoud, M.; Garnett, R.; Ferguson, M.; Kanaroglou, P. Electric buses: A review of alternative powertrains. Renew. Sustain. Energy Rev. 2016, 62, 673–684.

Borén, S. Electric buses' sustainability effects, noise, energy use, and costs. Int. J. Sustain. Transp. 2020, 14, 1–16.

Kühne, R. Electric buses—An energy efficient urban transportation means. Energy 2010, 35, 4510–4513.

Mohamed, M.; Farag, H.; El-Taweel, N.; Ferguson, M. Simulation of electric buses on a full transit network: Operational feasibility and grid impact analysis. Electr. Power Syst. Res. 2017, 142, 163–175.

Quarles, N.; Kockelman, K.M.; Mohamed, M. Costs and Benefits of Electrifying and Automating Bus Transit Fleets. Sustainability 2020, 12, 3977.

Wellik, T.K.; Griffin, J.R.; Kockelman, K.M.; Mohamed, M. Utility-transit nexus: Leveraging intelligently charged electrified transit to support a renewable energy grid. Renew. Sustain. Energy Rev. 2021, 139.

Mohamed, M.; Ferguson, M.; Kanaroglou, P. What hinders adoption of the electric bus in Canadian transit? Perspectives of transit providers. Transp. Res. Part D Transp. Environ. 2018, 64, 134–149.

Chen, F.; Fernandes, T.; Roche, M.Y.; Carvalho, M.D.G. Investigation of challenges to the utilization of fuel cell buses in the EU vs transition economies. Renew. Sustain. Energy Rev. 2007, 11, 357–364.

El-Taweel, N.A.; Farag, H.E.Z.; Mohamed, M. Integrated Utility-Transit Model for Optimal Configuration of Battery Electric Bus Systems. IEEE Syst. J. 2020, 14, 738–748.

Liu, Z.; Song, Z.; He, Y. Economic Analysis of On-Route Fast Charging for Battery Electric Buses: Case Study in Utah. Transp. Res. Rec. J. Transp. Res. Board 2019, 2673, 119–130.

Abdelaty, H.; Elsayed, M.; Mohamed, M. Assessing the feasibility of energy storage

system for electric bus transit planning. In Proceedings of the 55th Annual Canadian Transportation Research Forum, Montreal, QC, Canada, 24–27 May 2020; pp. 1–21.

El-Taweel, N.A.; Mohamed, M.; Farag, H.E. Optimal design of charging stations for electrified transit networks. In Proceedings of the 2017 IEEE Transportation Electrification Conference and Expo (ITEC), Chicago, IL, USA, 22–24 June 2017; pp. 786–791.

He, Y.; Song, Z.; Liu, Z. Fast-charging station deployment for battery electric bus systems considering electricity demand charges. Sustain. Cities Soc. 2019, 48, 101530.

Rupp, M.; Rieke, C.; Handschuh, N.; Kuperjans, I. Economic and ecological optimization of electric bus charging considering variable electricity prices and CO2eq intensities. Transp. Res. Part D Transp. Environ. 2020, 81, 102293.

Liu, K.; Yamamoto, T.; Morikawa, T. Impact of road gradient on energy consumption of electric vehicles. Transp. Res. Part D Transp. Environ. 2017, 54, 74–81.

Qi, X.; Wu, G.; Boriboonsomsin, K.; Barth, M.J. Data-driven decomposition analysis and estimation of link-level electric vehicle energy consumption under real-world traffic conditions. Transp. Res. Part D Transp. Environ. 2018, 64, 36–52.

Teoh, L.E.; Khoo, H.L.; Goh, S.Y.; Chong, L.M. Scenario-based electric bus operation: A case study of Putrajaya, Malaysia. Int. J. Transp. Sci. Technol. 2018, 7, 10–25.

Vepsäläinen, J.; Kivekäs, K.; Otto, K.; Lajunen, A.; Tammi, K. Development and validation of energy demand uncertainty model for electric city buses. Transp. Res.

Part D Transp. Environ. 2018, 63, 347–361.

Wang, J.; Kang, L.; Liu, Y. Optimal scheduling for electric bus fleets based on dynamic programming approach by considering battery capacity fade. Renew. Sustain. Energy Rev. 2020, 130, 109978.

Zhou, B.; Wu, Y.; Zhou, B.; Wang, R.; Ke, W.; Zhang, S.; Hao, J. Real-world performance of battery electric buses and their life-cycle benefits with respect to energy consumption and carbon dioxide emissions. Energy 2016, 96, 603–613.

Basso, R.; Kulcsár, B.; Egardt, B.; Lindroth, P.; Sanchez-Diaz, I. Energy consumption estimation integrated into the Electric Vehicle Routing Problem. Transp. Res. Part D Transp. Environ. 2019, 69, 141–167.

De Cauwer, C.; Van Mierlo, J.; Coosemans, T. Energy Consumption Prediction for Electric Vehicles Based on Real-World Data. Energies 2015, 8, 8573–8593.

Kivekäs, K.; Lajunen, A.; Vepsäläinen, J.; Tammi, K. City Bus Powertrain Comparison: Driving Cycle Variation and Passenger Load Sensitivity Analysis. Energies 2018, 11, 1755.

Vepsäläinen, J.; Otto, K.; Lajunen, A.; Tammi, K. Computationally efficient model for energy demand prediction of electric city bus in varying operating conditions. Energy 2019, 169, 433–443.

Abdelaty, H.; Mohamed, M. Uncertainty in Electric Bus Energy Consumption: The Impacts of Grade and Driving Behaviour. In Proceedings of the 55th Annual Canadian Transportation Research Forum, Montreal, QC, Canada, 24–27 May 2020.

Abdelaty, H.; Al-Obaidi, A.; Mohamed, M.; Farag, H.E.Z. Machine Learning Prediction Models for Battery-Electric Bus Energy Consumption in Transit. Transp. Res. Part D Transp. Environ. 2021.

Kunith, A.; Mendelevitch, R.; Goehlich, D. Electrification of a city bus network—An optimization model for cost-effective placing of charging infrastructure and battery sizing of fast-charging electric bus systems. Int. J. Sustain. Transp. 2017, 11, 707–720.

Franca, A. Electricity Consumption and Battery Lifespan Estimation for Transit Electric Buses: Drivetrain Simulations and Electrochemical Modelling; University of Victoria: Victoria, BC, Canada, 2015.

Gallet, M.; Massier, T.; Hamacher, T. Estimation of the energy demand of electric buses based on real-world data for large-scale public transport networks. Appl. Energy 2018, 230, 344–356.

Lajunen, A. Energy consumption and cost-benefit analysis of hybrid and electric city buses. Transp. Res. Part C Emerg. Technol. 2014, 38, 1–15.

Lajunen, A. Lifecycle costs and charging requirements of electric buses with different charging methods. J. Cleaner Prod. 2018, 172, 56–67.

Vepsäläinen, J.; Ritari, A.; Lajunen, A.; Kivekäs, K.; Tammi, K. Energy Uncertainty Analysis of Electric Buses. Energies 2018, 11, 3267.

Gao, Z.; Lin, Z.; LaClair, T.J.; Liu, C.; Li, J.-M.; Birky, A.K.; Ward, J. Battery capacity and recharging needs for electric buses in city transit service. Energy 2017, 122,

588–600.

Lajunen, A.; Tammi, K. Energy consumption and carbon dioxide emission analysis for electric city buses. In Proceedings of the 29th World Electric Vehicle Symposium and Exhibition (EVS29), Montreal, QC, Canada, 19–22 June 2016.

Diaz Alvarez, A.; Serradilla Garcia, F.; Naranjo, J.E.; Anaya, J.J.; Jimenez, F. Modeling the Driving Behavior of Electric Vehicles Using Smartphones and Neural Networks. IEEE Intell. Transp. Syst. Mag. 2014, 6, 44–53.

Basma, H.; Mansour, C.; Nemer, M.; Stabat, P.; Haddad, M. Sensitivity analysis of bus line electrification at different operating conditions. In Proceedings of the 8th Transport Research Arena TRA, Helsinki, Finland, 27–30 April 2020; pp. 1–10.

Bajtelsmit, V. Risk Analysis and the Optimal Capital Budget; Lecture notes; University of Missouri: St. Louis, MO, USA, 1997.

Christopher Frey, H.; Patil, S.R. Identification and Review of Sensitivity Analysis Methods. Risk Anal. 2002, 22, 553–578.

Yuan, X.; Zhang, C.; Hong, G.; Huang, X.; Li, L. Method for evaluating the real-world driving energy consumptions of electric vehicles. Energy 2017, 141, 1955–1968. [

De Cauwer, C.; Verbeke, W.; Coosemans, T.; Faid, S.; Van Mierlo, J. A Data-Driven Method for Energy Consumption Prediction and Energy-Efficient Routing of Electric Vehicles in Real-World Conditions. Energies 2017, 10, 608.

Pamula, T.; Pamula, W. Estimation of the Energy Consumption of Battery Electric Buses

for Public Transport Networks Using Real-World Data and Deep Learning. Energies 2020, 13, 2340.

Galvin, R. Energy consumption effects of speed and acceleration in electric vehicles: Laboratory case studies and implications for drivers and policymakers. Transp. Res. Part D Transp. Environ. 2017, 53, 234–248.

Wang, J.-B.; Liu, K.; Yamamoto, T.; Morikawa, T. Improving Estimation Accuracy for Electric Vehicle Energy Consumption Considering the Effects of Ambient Temperature. Energy Procedia 2017, 105, 2904–2909.

Neaimeh, M.; Hill, G.A.; Hübner, Y.; Blythe, P.T. Routing systems to extend the driving range of electric vehicles. IET Intell. Transp. Syst. 2013, 7, 327–336.

Kivekäs, K.; Vepsäläinen, J.; Tammi, K.; Anttila, J. Influence of Driving Cycle Uncertainty on Electric City Bus Energy Consumption. In Proceedings of the 2017 IEEE Vehicle Power and Propulsion Conference (VPPC), Belfort, France, 11–14 December 2017; pp. 1–5.

Lajunen, A.; Kivekaes, K.; Baldi, F.; Vepsaelaeinen, J.; Tammi, K. Different Approaches to Improve Energy Consumption of Battery Electric Buses. In Proceedings of the 2018 IEEE Vehicle Power and Propulsion Conference (VPPC), Chicago, IL, USA, 27–30 August 2018; pp. 1–6.

Qi, Z.; Yang, J.; Jia, R.; Wang, F. Investigating Real-World Energy Consumption of Electric Vehicles: A Case Study of Shanghai. Procedia Comput. Sci. 2018, 131, 367–376.

Liu, K.; Wang, J.; Yamamoto, T.; Morikawa, T. Exploring the interactive effects of ambient temperature and vehicle auxiliary loads on electric vehicle energy consumption. Appl. Energy 2018, 227, 324–331.

Zhang, R.; Yao, E. Electric vehicles' energy consumption estimation with real driving condition data. Transp. Res. Part D Transp. Environ. 2015, 41, 177–187.

Badin, F.; Le Berr, F.; Briki, H.; Dabadie, J.-C.; Petit, M.; Magand, S.; Condemine, E. Evaluation of EVs energy consumption influencing factors, driving conditions, auxiliaries use, driver's aggressiveness. In Proceedings of the 2013 World Electric Vehicle Symposium and Exhibition (EVS27), Barcelona, Spain, 17–20 November 2013; pp. 1–12.

Hahn, B.; Valentine, D. SIMULINK®Toolbox. In Essential MATLAB for Engineers and Scientists; Academic Press: Cambridge, MA, USA, 2019.

De Filippo, G.; Marano, V.; Sioshansi, R. Simulation of an electric transportation system at The Ohio State University. Appl. Energy 2014, 113, 1686–1691.

Rodríguez Pardo, M. Uncertainty in Electric Bus Mass and Its Influence in Energy Consumption; Aalto University: Espoo, Finland, 2017.

Beckers, C.J.J.; Besselink, I.J.M.; Frints, J.J.M.; Nijmeijer, H. Energy consumption prediction for electric city buses. In Proceedings of the 13th ITS European Congres, Eindhoven, The Netherlands, 3–6 June 2019.

Reimpell, J.; Stoll, H.; Betzler, J.W. The Automotive Chassis: Engineering Principles; Butterworth-Heinemann: Woburn, MA, USA, 2001.

Hogan, C.M.; Latshaw, G.L. The relationship between highway planning and urban noise. In Proceedings of the ASCE Urban Transportation Division Environment Impact Specialty Conference, Chicago, IL, USA, 21–23 May 1973; pp. 109–126.

Pelkmans, L.; De Keukeleere, D.; Bruneel, H.; Lenaers, G. Influence of Vehicle Test Cycle Characteristics on Fuel Consumption and Emissions of City Buses. SAE Trans. 2001, 110, 1388–1398.

NewFlyer-XE40. Xcelsior Charge Technical Summary—400 Electric Bus; New Flyer: Anniston, AL, USA, 2017.

James, G.M.; Hastie, T.; Witten, D.; Tibshirani, R. An Introduction to Statistical Learning: With Applications in R; Springer: New York, NY, USA, 2013.

Altoona. Federal Transit Bus Test, Manufacturer: New Flyer, Model: XE40; Pennsylvania Transportation Institute: Pennsylvania, PA, USA, 2015.

Sagaama, I.; Kchiche, A.; Trojet, W.; Kamoun, F. Impact of Road Gradient on Electric Vehicle Energy Consumption in Real-World Driving. Springer International Publishing: Cham, Switzerland, 2020; pp. 393–404.

Weather Canada. Canada Daily Data Report, Environment and Natural Resources. Weather, Climate and Hazard; Government of Canada: Toronto, ON, Canada, 2019.

Kontou, A.; Miles, J. Electric Buses: Lessons to be Learnt from the Milton Keynes Demonstration Project. Procedia Eng. 2015, 118, 1137–1144.

Joseph, F.; Hair, J.; Black, W.C.; Babin, B.J.; Anderson, R.E. Multivariate Data Analysis,

7th ed.; Pearson Education Limited: Harlow, UK, 2009.

George, D.; Mallery, P. SPSS for Windows Step by Step: A Simple Guide and Reference; Prentice Hall Press: Hoboken, NJ, USA, 2010.

Hair, J.; Black, W.; Babin, B.; Anderson, R. Multivariate Data Analysis; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2010.

Liu, K.; Wang, J.; Yamamoto, T.; Morikawa, T. Modelling the multilevel structure and mixed effects of the factors influencing the energy consumption of electric vehicles. Appl. Energy 2016, 183, 1351–1360.

Melaina, M.; Bush, B.; Eichman, J.; Wood, E.; Stright, D.; Krishnan, V.; Keyser, D.; Mai, T.; McLaren, J. National Economic Value Assessment of Plug-In Electric Vehicles; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2016.

CHAPTER FOUR

4. MACHINE LEARNING PREDICTION MODELS FOR BATTERY-ELECTRIC BUS ENERGY CONSUMPTION IN TRANSIT

Abstract: The energy consumption (E_C) of battery-electric buses (BEB) varies significantly due to the intertwined relationships of vehicular, operational, topological, and external parameters. This variation is posing several challenges to predict BEB's energy consumption. Several studies are calling for the development of data-driven models to address this challenge. This study develops and compares seven data-driven modelling techniques that cover both machine learning and statistical models. The models are based on big data generated using full-factorial experimental design (n = 907,199) of a validated Simulink energy simulation model. The models are then used to predict E_C using a testing dataset (n = 169,344). The results show some minor discrepancies between the developed models. All models explained more than 90% of the energy consumption variance. Further, the results indicate that road gradient and the battery state of charge are the most influential factors on E_C, while driver behaviour and drag coefficient have the lowest impact.

Keywords: *Battery electric buses; Data-driven modelling techniques; Energy consumption; Factorial design; Sensitivity analysis; Operational/topological parameters*

4.1. Introduction

Battery electric bus (BEB) utilization in transit systems represents an innovative and promising technology that has emerged as environmentally friendly with the potential for zero greenhouse gas emissions (GHG) (Pihlatie et al., 2014; Quarles et al., 2020; Vepsäläinen et al., 2018b). Governments are promoting BEBs deployment by allocating substantial funding and propagating supportive policies (Qi et al., 2018a). As a result, about 47% of the global transit buses will be electric by 2025, increasing from 386,000 buses to 1.2 million buses (Chediak, 2018).

BEBs have several advantages over diesel buses; to name only a few it generates less noise, it has higher efficiency, higher torque at low speeds leading to better accelerations, and it enhances the energy efficiency by allowing the generation of electric energy from the kinetic energy through regenerative braking systems (Boren, 2019; El-Taweel et al., 2019; Kühne, 2010; Kumar, 2015; Mahmoud et al., 2016; Rodríguez Pardo, 2017).

Despite the advantages of using BEBs, there are still concerns regarding the required battery size to accommodate the varying operational demand (route length and operating hours) (El-Taweel et al., 2019; El-Taweel et al., 2017). Toward that, a convenient charging strategy is required to guarantee the efficient operation of BEBs (Wellik et al., 2021). The charging strategies include optimal allocation of charging stations (e.g. overnight and opportunity charging), wireless charging lanes, and battery swapping (An, 2020; Chen et al., 2018; Li, 2014).

There is an abundance of research studies investigating the optimal system sizing (battery size and charging stations) for BEB deployment in the transit system. However, all these studies hinge on having a reliable and accurate estimation of BEB's energy consumption (Mohamed et al., 2017; Offer et al., 2010).

BEB energy consumption (E_C) varies significantly due to the uncertainties in external factors such as route topology, driver behaviour, weather condition, and transit parameters (e.g. traffic condition and the number of bus stops). This variation is challenging and often requires very sophisticated modelling and simulation techniques or real-world data to identify the significant parameters that affect BEB's energy consumption (Rupp et al., 2020; Vepsäläinen et al., 2018b).

In this respect, previous studies have independently identified several parameters impacting BEB's energy consumption. These are broadly classified into four groups. 1) Vehicular parameters include bus mass (m) (Gallet et al., 2018; Kivekäs et al., 2017; Kunith et al., 2017), drag coefficient (C_d) (Gallet et al., 2018), rolling resistance (C_r) (Gallet et al., 2018; Lajunen, 2018), battery temperature (T_B), battery capacity (C_B), HVAC power, and the state of charge (SoC) (Vepsäläinen et al., 2019). 2) Operational parameters include driver behaviour parameters such as acceleration rate (a), deceleration rate (d), and average speed (V_a) (Franca, 2015; Gao et al., 2017; Lajunen, 2014). As well as transit parameters that include traffic condition (T_C), travel time (T_T), spacing between stops (SS) (Lajunen, 2014), passenger loading (P_L) (Kivekas et al., 2018), and regenerative brake (B_{reg}). 3) Topological parameters include route length (L) (Gallet et al., 2018; Gao et al., 2017; Lajunen, 2014), road gradient (g), and road condition (R_C). 4) External parameters include

air density (P_a) (Gallet et al., 2018; Vepsäläinen et al., 2018a), ambient temperature (T_A) (Vepsäläinen et al., 2019), wind speed (V_W), and auxiliary power (AUX) (Gallet et al., 2018; Vepsäläinen et al., 2019).

These parameters are identified, in the literature, based on two modelling approaches. The first approach uses real-world testing to measure BEB energy consumption and associate it with vehicular, operational, topological, and external parameters. This is similar to the works of (Gao et al., 2017; Qi et al., 2018a; Xu et al., 2015; Yuan et al., 2017). However, such an approach often explains the energy consumption behaviour under a limited number of real-world experiments, which limits the generalization of the results due to the small number of unique experiments (De Cauwer et al., 2017; Laurikko et al., 2015; Qi et al., 2018b).

The second approach relies on simulation models of BEB energy consumption based on kinematic and dynamic conditions such as the works of (Dib et al., 2014; Hahn and Valentine, 2019; Rupp et al., 2019). The simulation approach is often utilized to identify the influence of the external parameters and drivetrain parameters on the energy consumption under varying conditions (De Cauwer et al., 2015; Gao et al., 2017; Vepsäläinen et al., 2018b). However, it has some limitations since the simulation approach often depends on the vehicular parameters and the dynamic conditions, which leads to a shortage in studying the effect of the operational and the topological parameters on the consumed energy.

Both approaches are valid, yet they are limited in their practical relevance and ease of utilization for transit providers and municipality (Abdelaty and Mohamed, 2021). This

is due to the required technical expertise and/or investments to simulate/test BEB fleets (Mohamed et al., 2018). Therefore, there is a growing interest from service providers and municipalities in data-driven models that can provide a BEB energy consumption analysis 1) Without the need for sophisticated simulation models and 2) When there is a lack of access to BEB real-world data.

In this respect, this paper aims to develop, compare, and test the performance of seven dominant data-driven models for predicting the energy consumption of BEBs in transit operations. These seven models belong to two distinct approaches, statistical analysis and Machine Learning (ML), and were partially selected as the dominant models applied in previous studies. Such a comparison will inform transit providers and academia alike on the best data-driven model(s) for BEB energy consumption prediction, as well as the limitations of each modelling approach.

4.2. Literature Review

Several data-driven modelling approaches are frequently implemented in BEB E_C research (Table 4.1), including 1) Multiple Regression Analysis (MRA), which determines the significance of the input parameters in estimating the E_C. This is based on the contribution of each parameter in reducing the residual sum of squares (Liu et al., 2017). 2) Interpolation Method (IM), which interpolates the value of a dependent parameter as a function of a set of independent parameters (Vepsäläinen et al., 2019). 3) Machine Learning (ML), which provides the model with the ability to automatically analyze and interpret the underlying patterns and structures in the data (Mitchell, 1997). Machine learning models include

support vector machine and decision tree models. 4) Neural Network (NN), which recognizes the fundamental relationships in a set of data through a process that mimics the way the human brain operates (Dreyfus, 2005). Further, Sensitivity Analysis (SA) is often utilized as a follow-up analysis to describe the importance of each input parameter in predicting the E_C and explains how the variation in the parameters can affect the E_C (Diaz Alvarez et al., 2014).

Table 4-1. The employed modelling techniques and input parameters in the literature

Study	Context	Technique	Vehicular				Operational			Topological		External		
			m	SoC	C_r	C_d	acc/dec	V_a	S_N	L	g	T_A	HVAC	Aux.
(Ma et al., 2021)	e-Bus	GBDT	√	-	-	-	-	√	√	-	√	-	-	√
(Pamuła and Pamuła, 2020)	e-Bus	RA & NN	-	-	-	-	-	-	√	-	√	√	-	-
(Vepsäläinen et al., 2019)	e-Bus	IM	√	√	√	-	-	-	-	-	-	√	√	√
(Gallet et al., 2018)	e-Bus	SA	√	-	√	√	√	√	√	√	-	-	-	√
(Vepsäläinen et al., 2018)	e-Bus	IM & RA	√	√	√	√	√	√	√	√	-	√	-	-
(Teoh et al., 2018)	e-Bus	RA	√	-	-	-	-	-	-	√	-	-	-	-
(Lajunen et al., 2018)	e-Bus	SA	√	-	√	√	√	√	-	-	-	-	-	-
(Lajunen, 2018)	e-Bus	SA	√	-	√	-	√	√	-	√	-	-	-	√
(Gao et al., 2017)	e-Bus	SA	√	-	√	√	√	√	-	√	-	-	-	√
(Kunith et al., 2017)	e-Bus	SA	√	-	√	√	-	√	-	-	-	-	-	√
(Kivekäs et al., 2017)	e-Bus	SA	√	-	√	√	√	√	√	√	-	-	-	√
(Huang et al., 2017)	e-Bus	NN	√	-	-	-	√	√	√	-	-	√	-	-
(Franca, 2015)	e-Bus	SA	√	-	√	√	√	√	-	√	-	-	-	√
(Lajunen, 2014)	e-Bus	SA	√	-	√	√	√	√	√	√	-	-	-	-
(Kanarachos et al., 2019)	ICE Vehicle	NN	-	-	-	-	√	√	-	-	-	-	-	-
(Ping et al., 2019)	ICE Vehicle	NN	-	-	-	-	√	√	-	-	-	-	-	-
(Yamashita et al., 2018)	ICE Vehicle	NN	-	-	-	-	√	√	-	√	-	-	-	-
(Qi et al., 2018a)	ICE Vehicle	NN	-	-	-	-	√	√	-	-	√	-	-	-
(Du et al., 2017)	ICE Vehicle	NN	√	-	-	-	-	√	-	√	-	-	-	-
(Zeng et al., 2015)	ICE Vehicle	NN	-	-	-	-	-	√	√	√	-	-	-	-
(Diaz Alvarez et al., 2014)	ICE Vehicle	NN	-	-	-	-	√	√	-	-	-	-	-	-
(Masikos et al., 2014)	ICE Vehicle	NN	√	√	-	-	-	-	-	-	√	√	-	√
(Shankar and Marco, 2013)	ICE Vehicle	NN	-	-	-	-	√	√	√	√	-	-	-	-
(Liu et al., 2018)	EV	RA	-	-	-	-	-	√	-	√	√	√	√	-
(Qi et al., 2018b)	EV	RA	-	√	-	-	-	√	-	√	-	√	-	-
(Liu et al., 2017)	EV	RA	-	-	-	-	-	√	-	√	√	-	-	-
(Galvin, 2017)	EV	RA	-	-	-	-	√	√	-	√	-	-	-	-
(De Cauwer et al., 2017)	EV	RA	-	-	√	√	√	√	-	√	√	√	-	√
(Wang et al., 2017)	EV	RA	-	-	-	-	-	√	-	√	√	√	√	-

(Yuan et al., 2017)	EV	RA	-	-	√	√	√	-	-	-	√	√	-	-
(De Cauwer et al., 2015)	EV	RA	-	-	√	√	√	-	-	-	√	-	-	√
(Zhang and Yao, 2015)	EV	RA	-	√	-	-	√	√	-	-	-	-	-	-

GBR: Gradient Boosting Regression, MRA: Multiple Regression Analysis, NN: Neural Network, IM: Interpolation Method, SA: Sensitivity Analysis; m: Vehicle Mass, SoC: State of Charge, C_r: Rolling Resistance Coefficient, C_d: Drag Coefficient, Acc/dec: Acceleration and Deceleration Rates, V_a: Average Speed, T_A: Ambient Temperature, HVAC: Heating, Ventilation, and Air Conditioning, Aux: Auxiliary Power, S_N: Number of Stops, L: Route Length, g: Road Gradient.

Table 4.1 provides a summary of relevant studies, highlighting their modelling technique, modelling specification (e.g. parameters in the model), and the analysis context.

Regarding MRA, few studies have used regression models to predict the E_C of BEBs and study the impact of various parameters on the E_C. Teoh et al. (2018) used a linear regression model to estimate the effect of route length and passenger loading on the E_C of BEBs. Vepsäläinen et al. (2018b) used a linear regression model to identify the correlated parameters to select the parameters that affect energy consumption in a sensitivity analysis. Similarly, Pamula and Pamula (2020) used linear regression to develop an E_C prediction model based on the trip distance, trip time, the spacing between stops, and the difference between route elevations. They concluded that the route length and passenger loading are significant operational parameters; however, the E_C is greatly affected by various external parameters such as the headway and the charging type (Teoh et al., 2018). Also, Vepsäläinen et al. (2018b) concluded the significance of the driver's aggressiveness and the number of stops per km on the consumed energy of BEBs. Recently, Abdelaty and Mohamed (2021) developed a linear regression model that predicts the E_C of BEBs from several vehicular, operational, topological and external parameters.

Despite their relative scarcity in the BEB domain, MRA models have been extensively utilized in the electric vehicle (EV) context with promising results. (De Cauwer et al., 2015; De Cauwer et al., 2017; Liu et al., 2017; Qi et al., 2018b; Yuan et al., 2017) used road gradient, average speed, acceleration/deceleration rates, air conditioning and heating system, battery state of charge, ambient temperature, rolling resistance, and aerodynamic drag as independent variables to predict EVs' energy consumption. While

(Liu et al., 2018; Wang et al., 2017) used ordinary least squares regression and multilevel mixed-effects regression models to evaluate the impacts of ambient temperature, route length, road gradient, and HVAC on the E_C. Moreover, (Galvin, 2017; Zhang and Yao, 2015) developed a multivariate linear regression model to study the impact of driver behaviour parameters such as acceleration/deceleration rates and average speed on the E_C for EVs.

These studies, despite focusing on EVs, concluded that average speed and acceleration/deceleration rates have high impacts on E_C (De Cauwer et al., 2017; Galvin, 2017). Air conditioning and heating systems increase the consumed energy significantly (De Cauwer et al., 2017; Liu et al., 2018). The road gradient shows the most significant change in the E_C rates. Moreover, there are sharp divergences between the impact of upgrade and downgrade on the E_C (De Cauwer et al., 2017; Liu et al., 2018; Liu et al., 2017; Wang et al., 2017). Furthermore, Qi et al. (2018b) found an inverse relationship between the trip distance and ambient temperature and the E_C.

For the Interpolation Method (IM), Vepsäläinen et al. (2019) and Vepsäläinen et al. (2018b) used a surrogate modelling technique as an interpolation method to perform a sensitivity analysis for E_C rates based on the uncertainties in the weather (i.e. ambient temperature and rolling resistance), the driver behaviour (i.e. average speed and driver aggressiveness), and the state of charge. Vepsäläinen et al. (2019) acknowledged that the variation in ambient temperature causes the greatest change in the E_C rates due to its impact on auxiliary power and HVAC system, while the variation in rolling resistance showed the second-highest change in the E_C. At the same time, Vepsäläinen et al. (2018b) concluded

the high significance of stop density, ambient temperature, and the driver aggressiveness on the E_C rates.

Further, Franca (2015) conducted a sensitivity analysis for the BEB energy consumption model using input parameters such as passenger loading, temperature, frontal area, drag coefficient, rolling resistance, and auxiliary power. Likewise, Basma et al. (2020) conducted a sensitivity analysis using the battery state of charge and battery service life to studying their impact on the BEB optimal battery design and the total energy cost. Furthermore, Kivekas et al. (2018) used the global sensitivity approach to compare the impact of the number of stops and the average number of passengers on the E_C for BEB.

The SA results demonstrate that the variation in passenger loading impacts the E_C rates substantially (Franca, 2015). In contrast, Kivekas et al. (2018) found that variation in the number of stops has a higher impact on E_C rates compared to passenger loading. Basma et al. (2020) concluded the minimum state of charge's high incidence on the optimal battery sizing, where higher values indicate less effective battery energy capacity.

With respect to Machine Learning (ML) models, there is a lack of literature concerning the use of ML models in BEB or EV studies. That said, ML techniques are proved superior prediction models in several domains (Lawson et al., 2021). Therefore, we argue that the inclusion of ML in this study is essential to compare a different family of models in predicting the energy consumption of the BEBs.

Concerning Neural Network (NN), Pamula and Pamula (2020) combined Deep learning neural network (DLNN) with MRA to predict the energy consumption rates for

the BEBs considering route parameters such as trip distance, trip time, the spacing between stops, and the difference between route's elevation. They developed a model with an estimation error of less than 7% when comparing the estimated and actual E_C data for numerous real-world bus trips' data.

Outside the BEB domain, there is a diversity of studies that applied DLNN models for EVs and internal combustion engine (ICE) vehicles. For example, Diaz Alvarez et al. (2014) trained a neural network to predict EVs' energy consumption based on driver behaviour characteristics using data acquired from drivers' smartphones. The recurrent neural network (RNN) had been used to predict EVs' energy consumption rates considering the average speed, average acceleration/deceleration rates, stop density, trip distance, and trip time (Kanarachos et al., 2019; Shankar and Marco, 2013). While Masikos et al. (2014) applied the general regression neural network (GRNN) to study EVs' energy consumption based on the state of charge, mass, road gradient, and ambient temperature. Qi et al. (2018a) investigated the association between traffic congestion and EV energy consumption using regression analysis and neural network models. They represented traffic congestion as a factor of the average speed, acceleration/deceleration rates, road type, road gradient, PKE (a measure of the change in kinetic energy per unit distance due to acceleration), and NKE (a measure of the change in kinetic energy per unit distance due to deceleration, which may account for the regenerative braking impact).

While for Internal Combustion Engine (ICE) vehicles, the interested reader is referred to (Du et al., 2017; Ping et al., 2019; Wu and Liu, 2011, 2012; Yamashita et al., 2018).

Based on EV studies, the acceleration caused most of the energy consumption variation, and the average speed had the least effect (Diaz Alvarez et al., 2014). While Qi et al. (2018a) concluded that average speed is the highest significant parameter in estimating the energy consumption rates. The results also indicated an accuracy ranging from 70% to 80% when comparing the estimated and actual energy consumption for more than 800 vehicle trips (Shankar and Marco, 2013). While (Kanarachos et al., 2019; Masikos et al., 2014; Ping et al., 2019; Yamashita et al., 2018) achieved a model accuracy of more than 94% for both energy consumption (for EV) and fuel consumption (for ICE) models. Furthermore, when they compare neural network models and regression models, the former was deemed more accurate (Pamula and Pamula, 2020; Zeng et al., 2015). In contrast, (Qi et al., 2018a) concluded that the regression models are better than the neural network models due to the linear relationship between the utilized parameters and energy consumption.

Recently, Ma et al. (2021) used a gradient boosting decision tree to evaluate the influential parameters that affect diesel buses and electric buses' energy consumption. They found that BEBs are less affected by traffic congestion parameters than diesel buses. While the number of stops and distance between stops have high impacts on the E_C of BEBs compared to diesel buses.

Through this brief literature review, some research gaps are defined:

- There is no previous data-driven model that combines the operational, topological, vehicular, and external parameters as inputs in the BEBs' prediction model, which results in disregarding the impact of substantial parameters such as road gradient,

road condition, and stop density. Which, in turn, reduces the accuracy of the model estimation.

- The results of different data-driven models are not always consistent. This might be attributed to model specification, the modelling approach, or the sample characteristics.

- Although different data-driven models are utilized, there is no information about which technique is better or more accurate in predicting the E_C rates of BEBs in transit.

Toward that end, the present study aims to develop and compare different data-driven modelling techniques in predicting the energy consumption of BEBs using a wide range of vehicular, operational, topological, and external parameters. In particular, we have four overarching objectives: 1) Developing a full-factorial experiment to mimic BEB operation under all possible environments. 2) Validating a simulation-based BEB energy consumption model with results from real-world BEB experiments. 3) Predicting BEB's energy consumption from vehicular, operational, topological, and external parameters using seven different data-driven modelling techniques. 4) Comparing and evaluating the performance of these techniques for predicting the energy consumption of BEBs.

The findings of this study will provide tangible contributions to scholars and practitioners on the bearing of different modelling techniques on the prediction of BEB's energy consumption. Further, data-driven models provide transit operators and planners the means to plan, assess, and optimize the deployment of BEBs in transit.

4.3. Methodology

This study adopts a sequential methodological design, as illustrated in Figure 4.1. This includes three main subsections: experimental design, BEB energy estimation simulation model, and data-driven modelling techniques.
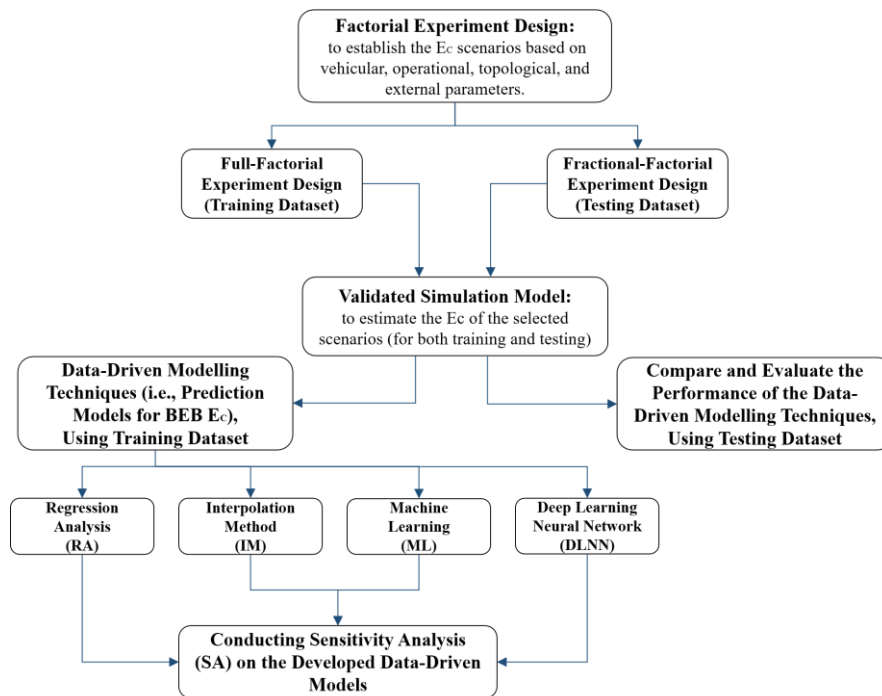
Figure 4-1. A flow chart explaining the utilized methodology

4.3.1. Factorial Experiment Design

A full-factorial experiment is used to define all the possible scenarios for BEB operation in transit based on vehicular, operational, topological, and external parameters. The experiment design includes 120 driving cycles with different terrain to ensure the diversity of driver behaviours and road topology. The drive cycles are collected from real-world data of fourteen different bus routes from the Hamilton Street Railway (HSR) transit service.

The dataset extracted from the full-factorial experiments is referred to hereafter as the training dataset (n = 907,199).

Also, we developed a fractional-factorial experiment design to generate a subset of scenarios (referred to as the testing dataset, n = 169,344) based on different levels for vehicular, operational, topological, and external parameters (Box et al., 2005). The generated datasets include 56 drive cycles collected from eight different bus routes from HSR.

For the training dataset, we include four levels for the initial state of charge (40% to 100%, with 20% intervals). The total mass of the bus is calculated based on the curb weight and the number of passengers. The number of passengers is divided into five levels, where the minimum is Zero passengers, and the maximum capacity is 75 passengers with an average passenger weight of 75 kg (Lajunen et al., 2018). The HVAC power is estimated based on the ambient temperature following the model developed by (Tammi and Lajunen, 2016) and modified by (Vepsäläinen et al., 2019). The ambient temperature (six levels) varies between -20 °C and 30 °C based on the historical Canadian weather dataset (Canada, 2019). The drag and the rolling resistance coefficients (three levels each) range from 0.5 to 0.7 and from 0.006 to 0.02, respectively (Gallet et al., 2018; Lajunen, 2018; Vepsäläinen et al., 2019).

Besides, the scenarios in the training dataset include seven levels of the road gradient ranging from - 6% to 6%. The minimum and maximum acceleration and deceleration rates varied between (0.50 to 2.5) m/s^2 and (1 to 4) m/s^2, respectively. These rates are selected based on the literature and field observation and represent both average

and aggressive driving behaviours (Gallet et al., 2018; Kontou and Miles, 2015; Lajunen, 2014). The rates varied between 0.50 m/s^2 to 2.5 m/s^2 for acceleration and 1 m/s^2 to 4 m/s^2 for deceleration, while the average speed, the spacing between stops, and cycle length are derived from the driving cycles.

For the testing dataset, we include nine levels of the initial state of charge (20% to 100%, with 10% intervals). The total bus mass is divided into sixteen levels based on passenger loading (0 passengers till 75 passengers, with 5 passengers' intervals). The HVAC includes seven levels based on ambient temperature between -30 °C and 30 °C. The drag coefficient is considered as a constant value of 0.6. The road gradient is divided into nine levels, including more extreme grades from - 8% to 8%.

After selecting the parameters and the levels, a total of 907,199 scenarios (training dataset) and 169,344 scenarios (testing dataset) have been generated based on the full- and fractional-factorial experiments, respectively. The parameters and their levels for both datasets are defined based on the literature as detailed in Table 4.2 and Appendix 4.1.

Table 4-2. List of input parameters used in factorial experiment design

Parameter - Unit	Parameter	Training Data			Testing Data			Source
		Levels	Min.	Max.	Levels	Min.	Max.	
Initial state of charge (SoC$_i$)– %		4	40	100	9	20	100	Vepsäläinen et al. (2019)
Mass (m) – Kg		5	14932	20557	16	14932	20557	Rupp et al. (2020)
HVAC – kW	Vehicular and External	6	1.25	13.75	7	1.25	22.30	Vepsäläinen et al. (2019)
Drag Coefficient (C$_d$)– Unitless		3	0.5	0.7	1	0.6	0.6	Gallet et al. (2018)
Rolling resistance (C$_r$)– Unitless		3	0.006	0.02	3	0.006	0.02	Gallet et al. (2018)
Battery Capacity – kWh		1	200		200	200		(NewFlyer-XE40, 2017)
Frontal area – m2		1	8.32		8.32	8.32		(NewFlyer-XE40, 2017)
Dynamic Tires Radius – m		1	0.5		0.5	0.5		(NewFlyer-XE40, 2017)
Road gradient (g) – %		7	- 6	+ 6	9	- 8	+ 8	Lajunen et al. (2018)
Average speed (V$_a$) – Km/h		120	19.99	49.95	56	19.99	49.95	Computed from speed profiles
Maximum speed (V$_m$) – Km/h	Operational	120	31.60	72.91	56	31.60	72.82	Computed from speed profiles
Acceleration rates (a) – m/s^2	and	6	0.5	2.5	6	0.5	2.5	Computed from speed profiles
Deceleration rates (d) – m/s^2	Topological	6	1	4	6	1	4	Computed from speed profiles
Spacing between stops (S$_S$) – m		4	300	600	4	300	600	Computed from speed profiles
Cycle Length (L) – m		120	860.54	1817	56	892.6	1817	Computed from speed profiles
Total number of unique scenarios		**907,199**			**169,344**			

To calculate the E_C for all scenarios, we coded a loop using MATLAB Simulink software to estimate the E_C rates interchangeably for all the selected parameters, as shown in Appendix 4.1. In each step, one parameter changes while all other parameters are fixed. This step is automated for all the scenarios, and the simulation model is detailed in the following section.

Based on the output of the E_C simulation, a new set of parameters is defined. These new parameters include the passenger loading (P_L), which represents the number of passengers during the trip; road condition (R_C), which is divided into three levels I (poor road condition), II (fair road condition), and III (good road condition), based on the utilized rolling resistance coefficient and taking into consideration a constant bus tire pressure; route length (L), is the cycle length; the driver aggressiveness (D_{Agg}), is defined based on the acceleration and deceleration rates of the drivers. D_{Agg} is divided into three levels, Level I for slow acceleration and deceleration rate, level II for normal rates, and level III represents the aggressive driver behaviour. Lastly, stop density (S_D) has been calculated by dividing the spacing between stops by the cycle length.

4.3.2. BEB Energy Consumption Simulation Model

4.3.2.1. Energy Simulation Model Development

We developed a simulation model to predict BEB's energy consumption for all the scenarios resulted from the experimental design using a MATLAB Simulink platform and following the approach advised by (Hahn and Valentine, 2019; Markel et al., 2002). The model is based on the block design used in the advanced vehicle simulator (ADVISOR) (Markel et al., 2002).

The required energy to move the BEB within the trip is generated from the battery. It passes through the vehicle powertrain components, including motor, gearbox, and inverter, until reaching the wheels. This amount of energy is consumed during the bus's longitudinal dynamic movement and under numerous external influences such as environmental parameters and operational parameters (De Filippo et al., 2014; Franca, 2015; Rodríguez Pardo, 2017). The amount of energy consumed to overcome the corresponding resistances during the movement is estimated using Newton's second law of motion (Eq. 4.1) (Beckers et al., 2019; De Filippo et al., 2014; Franca, 2015; Rodríguez Pardo, 2017).

$$M.\vec{d} = \sum \vec{F} \qquad \qquad \text{Eq. 4.1}$$

The tractive force (F_T) should be equal to or more than the summation of four resistance forces (Eqs. 4.2–4.5). These forces include: 1) rolling resistance (F_R) (Eq. 4.2), which is governed by the coefficient of rolling friction (C_r), vehicle mass (m), and gravitational acceleration (g) (Reimpell et al., 2001). 2) The aerodynamic drag resistance (F_{aero}) (Eq. 4.3), which is based on the bus frontal area, air density, aerodynamic drag coefficient, and the speed of the bus. 3) Grade resistance (F_g) (Eq. 4.4), which depends mainly on the bus mass (m) and the grade of the roadway (g). Further, the force required to accelerate the vehicle (F_a) (Eq. 4.5) depends on the bus mass and the acceleration/deceleration rates (De Filippo et al., 2014; Franca, 2015; Gallet et al., 2018; Rodríguez Pardo, 2017).

$$F_R = C_r.V.\left(m + m_{eq}\right).g.\sin(\alpha) \qquad \qquad \text{Eq. 4.2}$$

$$F_{aero} = 0.5 . p . A_F . C_d . V^2 \qquad \text{Eq. 4.3}$$

$$F_g = (m + m_{eq}) . g . \cos(\alpha) \qquad \text{Eq. 4.4}$$

$$F_a = m . (Acc. \, or \, Dec.) \qquad \text{Eq. 4.5}$$

The energy required for the HVAC system is estimated as a function of the ambient temperature and humidity following the model developed by (Tammi and Lajunen, 2016) and modified by (Vepsäläinen et al., 2019). While the auxiliary power (AUX), including the power for the door's air compressor, hydraulic control for the braking system, and other auxiliary devices, is assumed at a constant rate of 7 kW. This rate represents the worst-case scenario for AUX energy consumption, as reported by Vepsäläinen et al. (2019). A full description of the energy consumption model is available at (Abdelaty and Mohamed, 2021).

A real-world BEB model is used as our base model. We utilized the 40 ft New Flyer XE40 electric bus. The input parameters for the base model, including the bus and operation parameters, are listed in Table 4.3. Please note that for the Altoona test, the bus was equipped with a Lithium-ion battery XSYST 7 (7kWh, 4 strings of 7 batteries), which results in ≈ 200 kWh. Further, New Flyer XE40 utilizes an electrical-based water heater with a capacity of 98,000 Btu/hr, while the AC is provided through a Copeland Scroll air compressor.

It should be noted that during the Altoona test, AUX power was not reported, and a value of Zero is therefore used in the simulation model, while the HVAC systems were turned off during the test, hence a value of Zero is also used in the base model.

Table 4-3. Base Model parameters for BEB energy consumption simulation

Parameter		Value
Bus Specification	Battery Initial State of Charge (%)	100
	Max. Torque (N.m.)	2500
	Battery Capacity (kWh)	200
	Frontal area (m^2)	8.32
	Dynamic Radius of the Tires (m)	0.5
	Gear Ratio	4.66
	Curb Weight (kg)	14,860
	Recharge Efficiency	0.978
	Round Trip Efficiency	0.971
	Motor Efficiency	0.916
	Discharge Efficiency	0.992
Base-model Specification	Drag Coefficient	0.6
	Rolling Resistance	0.01
	Air Density (kg/m^3)	1.27
	Ambient Temperature (º)	20
	HVAC (kW)	0
	Auxiliary Power (kW)*	0

4.4. Validation of the Energy Simulation Model

We utilized two approaches to validate the developed energy consumption simulation model. These approaches are based on previous literature (illustrated in Table 4.4) and include real-world-based and simulation-based validation.

First, for real-world-based validation, we utilized the Altoona test for the Ne Flyer XE40 BEB. The test includes three driving cycles: Arterial (ART = 3.073 miles), Central Business District (CBD = 3.073 miles), and Commuter (COM = 6.147 miles). We run the three cycles in our simulation model using all the specifications of the New Flyer XE40 BEB (Altoona, 2015). The validation resulted in estimated energy consumption within ± 3.81% (3.81 kWh/100 km) accuracy (Figure 4.2), which is consistent with the results of (De Filippo et al., 2014), who validated their model using the Altoona test for Proterra BEB and with estimated E_C of 5% accuracy.

Second, and for the simulation-based validation, we followed the approach utilized by Gao et al. (2017), which utilized Autonomie software. A BEB in Autonomie was reconfigured to match the powertrain parameters of the New Flyer XE40 BEB. Then the model was tested across three bus cycles (Manhattan, New York, and RTE). The Autonomie-validation process indicates only 6% error only as depicted in Appendix 4.2.

Table 4-4. BEB energy consumption validation methods in previous studies.

| Study	E$_C$ Estimation Method	Validation Method
Rupp et al. (2020)	Equation-based	NA
Vepsäläinen et al. (2019)	Simulation (Simulink)	Validated through a developed model
Gallet et al. (2018)	Equation-based	NA
Vepsäläinen et al. (2018a)	Simulation (Simulink)	Experimental test (one Bus Route)
Lajunen (2018)	Mathematical Model (MATLAB)	NA
Lajunen et al. (2018)	Simulation (Autonomie)	NA
Kivekäs as et al. (2017)	Simulation (Simulink)	Experimental test (one Bus Route)
Gao et al. (2017)	Simulation	Autonomie software (one cycle)
Kunith et al. (2017)	Simulation	NA
Franca (2015)	Simulation (Simulink)	Experimental Altoona
Lajunen (2014)	Simulation (ADVISOR)	NA
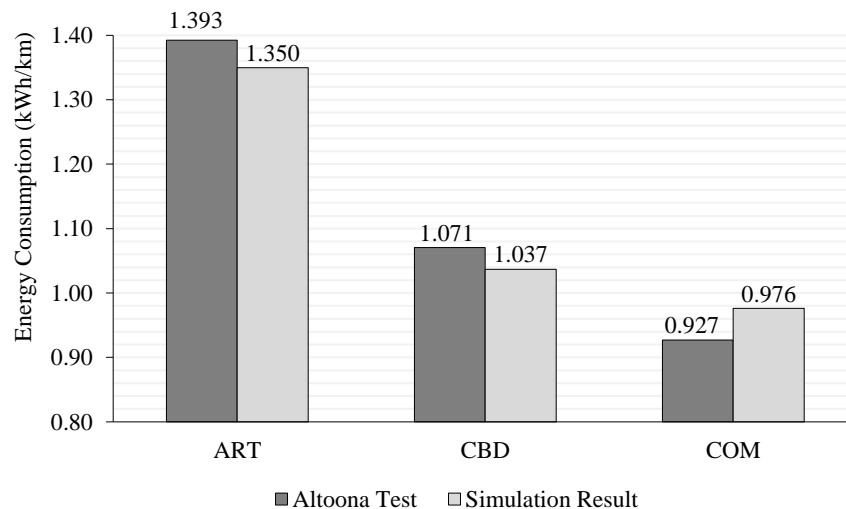
Figure 4-2. Energy consumption in Altoona test compared to the developed simulation model

4.4.1. Data-Driven Modelling Techniques

Seven data-driven modelling techniques were developed to predict the E_C, including Multiple Linear Regression Analysis (MLR), Interpolation Method (IM), Decision Tree (DT), Gradient Boosting Decision Tree (GBDT), Support Vector Machine (SVM), Radial basis Neural Network (RB-NN), and Multilayer Perception Neural Network (MLP-NN). These models span to cover different families of modelling, including statistical analysis and machine learning techniques. We used the training dataset (n = 907,199 scenarios) to develop each prediction model. While the testing data (n = 169,344 scenarios) is used to evaluate the prediction accuracy of each model.

In all models, the energy consumption (E_C) is the output (dependant) parameter. While the input parameters include the road gradient (g - %), the initial state of charge (SoC_i - %), drag coefficient (C_d - levels), road condition (R_C- Levels), HVAC (kWh), driver aggressiveness (D_{Agg} - levels), passenger loading (P_L – number of passengers), stop density (S_D -stops/km), average speed (V_a - Km/h), and route length (L - meter).

4.4.1.1. Multiple Regression Analysis

A Multiple Linear Regression model (MLR) was used to quantify the relationship between several independent parameters and a dependent parameter by fitting a linear equation to the observed data (Eq. 4.6).

$$E_C = \beta_0 + \beta_1 g + \beta_2 D_{Agg} + \beta_3 R_C + \beta_4 HVAC + \beta_5 P_L + \beta_6 S_D + \beta_7 V_a + \beta_8 SoC_i +$$

$$\beta_9 L + \beta_{10} C_d + \varepsilon \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{Eq. 4.6}$$

A set of analyses has been carried out to evaluate the appropriateness for the model

specification, including linearity, normality, homoscedasticity, and outliers as recommended by (Hair et al., 2009). The Least-Square Estimation (LSE) is used to determine the best fit for the regression model by minimizing the residual sum of squares. The goodness-of-fit of the model is evaluated using the coefficient of determination (R^2). The root means square error (RMSE) is used to assess the spread of the residuals. A forward stepwise entry method was implemented, where the model starts by the intercept only, and parameters are added to the model (one at each step) based on their contribution to overall goodness-of-fit and the amount of variance explained.

4.5. Interpolation Method

We used the Radial Basis Function (RBF) surrogate model, which approximates a target model using a series of weighted radial functions (Bajer and Holena, 2012). The radial function $\phi: R^n \rightarrow R$ is a function that is defined considering the distance between the input and a fixed reference point p as follows (Nedelkova et al., 2015):

$$\phi(x) = \phi(\parallel x - p \parallel),$$ Eq. 4.7

$$x \subset \mathbb{R}^n \rightarrow \mathbb{R}$$ Eq. 4.8

where $\parallel . \parallel$ is a norm on $\mathbb{R}^n$.

RBF surrogate model can produce accurate approximations of high-order models with lower computational power in comparison to other interpolation models. This increases the number of training points that are utilized in the construction of the surrogate model (Durantin et al., 2017). RBF interpolation method can be mathematically expressed as follows:

$$f(x) = \sum_{i=1}^{n} \pi_i \phi(\|x - p_i\|) \qquad \text{Eq. 4.9}$$

Where x is the input vector, n is the number of training points, π_i is the RBF interpolating weight, and p_i is the target response of training point i. The RBF (Eq. 4.9) can be rewritten in a matrix form as follows:

$$\begin{bmatrix} \phi(x_1,x_1) & \phi(x_1,x_2) & \cdots & \phi(x_1,x_n) \\ \phi(x_2,x_1) & \phi(x_2,x_2) & \cdots & \phi(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(x_n,x_1) & \phi(x_n,x_2) & \cdots & \phi(x_n,x_n) \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix} = \begin{bmatrix} f(\mathbf{x_1}) \\ f(\mathbf{x_2}) \\ \vdots \\ f(\mathbf{x_N}) \end{bmatrix} \qquad \text{Eq. 4.10}$$

The system of linear (Eq. 4.11) can be solved using the least square method to calculate the RBF weights:

$$\pi = \phi^{-1} f \qquad \text{Eq. 4.11}$$

Since the relation between energy consumption and the input parameters is mostly linear, the RBF chosen in this model is a linear function $\phi(x) = \|x - p\|$ (Wang and Qin, 2019). Therefore, the model predicts the energy consumption based on the following equation:

$$\hat{f}(x) = \sum_{i=1}^{n} \pi_i \|x - p_i\| \qquad \text{Eq. 4.12}$$

4.6. Decision Trees

A Decision Tree (DT) is a supervised machine learning technique that can be used to build regression models in the shape of a hierarchical data structure (Nedelkova et al., 2015). DT predicts the target response variable through decision rules generated from training data and their properties. DT models start with a root node that is split into subsequent decision nodes, which end with terminal leaves. The model applies a test on the input value at each

decision node and chooses a branch based on the outcome. The process continues until a terminal leaf is reached, which represents the predicted value of the target response variable (Bajer and Holena, 2012).

We applied the Classification and Regression Trees (CART) algorithm (Durantin et al., 2017) to construct the DT model as follows: for all predictor variables $x_{j,j} = 1,2\cdots m$, the algorithm splits a decision node d by calculating the weighted mean squared error (MSE) of the target response in node d as given by:

$$MSE_d = \frac{1}{N_d}\sum_{i \subset T} (y_i - \bar{y}_d)^2 \qquad\qquad Eq.\ 4.13$$

Where N_d is the number of samples in node d, T is the subset of samples in node d, y_i is the response variable value, and yd is the mean of response values of samples in node d. An iterative greedy algorithm is applied to find the split that results in the biggest reduction of MSE value. The splits continue until the error value is less than the required threshold, in which the algorithm stops and assigns the value of y_d to the terminal leaf (Wang and Qin, 2019). The output of the decision tree model for input x is represented as follows:

$$\hat{f}(x) = \sum_{i=1}^{M} c_i I \qquad\qquad Eq.\ 4.14$$

Where M is the number of leaf nodes in the model, c_i is a value stored in leaf i that represents input subset R_i, and I is an identity unction that gives 1 only when $x \in R_i$.

4.6.1. Gradient-Boosting Decision Tree

A boosting method works by combining multiple decision trees with weak prediction

accuracy to create a strong predictive model (Ma et al., 2021). The gradient boost algorithm is performed iteratively to reduce the prediction error of the model until a target value is reached. For a function $F(x)$, an approximation $F(x)$ is obtained by minimizing the loss function $L_{y,x}(y, F(x))$ over the training dataset $\{(x_i, y_i)\}$, $i = 1,2, \ldots, N$, as given by the following:

$$\tilde{F}(x) = \arg\min_{F(x)} L_{y,x}(y, F(x)) \qquad \text{Eq. 4.15}$$

In this work, we apply the least error function as the loss function as follows:

$$L(y, F(x)) = (y - F(x))^2 \qquad \text{Eq. 4.16}$$

The algorithm starts with an initial value $F_0(x)$:

$$F_0(x) = \arg\min_{c} \sum_{i=1}^{N} L(y_i, c) \qquad \text{Eq. 4.17}$$

Then in each iteration m, the following steps are performed:

The negative gradient $\tilde{y}_i$ of the loss function is calculated:

$$\tilde{y}_{m,i} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}, i = 1, \ldots, N \qquad \text{Eq. 4.18}$$

A new DT model is trained using the dataset $\{(x_i, \tilde{y}_{m,i})\}$:

$$\hat{f}_m(x) = \sum_{i=1}^{L_m} c_{m,i} I_{L_{m,i}} \qquad \text{Eq. 4.19}$$

Where L_m is the number of leaves for the mth iteration model. $c_{m,i}$ is value stored in leaf i that represents input subset $L_{m,i}$, and $I_{L_{m,i}}$ is an identity function that gives 1 only when $x_i \in L_{m,i}$.

In GBDT, the $c_{m,i}$ of the mth iteration for input subset $L_{m,i}$ is calculated:

$$c_{m,i} = \arg\min_{c} \sum_{x_i \in L_{m,i}} L\left(y_i, F_{m-1}(x) + c\hat{f}_m(x)\right)$$

Eq. 4.20

Finally, the model is updated as follows:

$$F_m(x) = F_{m-1}(x) + \sum_{i=1}^{L_m} c_{m,i} I\left(x \in L_{m,i}\right)$$

Eq. 4.21

4.6.2. Support Vector Machine Learning

Support vector machine (SVM) is a supervised machine learning technique that can be used to create classification and regression models (Dhiman et al., 2020). In this study, we use SVM to train a regression model based on a sample training dataset that consists of a predictor variable set $X \in R_n$ with n features and N observations, as well as a response set $Y \in R$ (Dhiman et al., 2020; Yang, 2019). The analysis of the training dataset shows that the relation between the target response variable and input parameters are mostly linear. Therefore, we use a linear function f(x) in the SVM model as follows:

$$f(x) = x.\beta^T + b,$$

Eq. 4.22

Where β is a weight coefficient, and b is a bias term. The aim of SVM is to find the optimal f(x) that have the minimum deviation from y_i for each x_i, and also be as flat as possible. To achieve this aim, we formulate an optimization problem that minimizes the norm value of the weight coefficients:

$$\min_{\beta} \frac{1}{2} \parallel \beta \parallel^2$$

Eq. 4.23

For each x_i, the deviation from y_i must be less than tolerance factor ε:

$$|y_i - (x_i. \beta^T + b)| \leq \varepsilon \ \forall i \in N \qquad\qquad \text{Eq. 4.24}$$

To deal with outlier data points that might prevent the optimization problem in (Eqs. 4.23–4.24) from converging, we introduce slack variables as soft constraints. These variables can manage the degree to which prediction errors and outlier data points are tolerated. Therefore, the optimization problem in (Eqs. 4.23–4.24) is reformulated as follows:

$$\min_{\beta, \xi_i, \xi_i^*} \ \frac{1}{2}\beta'\beta + C\sum_{i=1}^{N} (\xi_i + \xi_i^*) \qquad\qquad \text{Eq. 4.25}$$

$$y_i - (x_i. \beta^T + b) \leq \varepsilon + \xi_i \ \forall i \in N \qquad\qquad \text{Eq. 4.26}$$

$$(x_i. \beta^T + b) - y_i \leq \varepsilon + \xi_i^* \ \forall i \in N \qquad\qquad \text{Eq. 4.27}$$

$$\xi_i \geq 0 \qquad\qquad \forall i \in N \qquad\qquad \text{Eq. 4.28}$$

$$\xi_i^* \geq 0 \qquad\qquad \forall i \in N \qquad\qquad \text{Eq. 4.29}$$

Where ξ_i and ξ_i^* are slack variables, while C is a penalty factor that can be varied to control the degree of prediction error tolerance. The optimization problem in (Eqs. 4.25–4.29) can be converted to Lagrange dual formulation to make it easier to solve. Since that this problem is convex, the solution of the dual problem is equal to the primal solution. The dual formula is given by the following:

$$\min_\alpha \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i'x_j + \varepsilon\sum_{i=1}^{N} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i^* - \alpha_i)$$

$$\text{Eq. 4.30}$$

$$\sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0 \qquad\qquad \text{Eq. 4.31}$$

$$0 \leq \alpha_i \leq C \qquad\qquad \text{Eq. 4.32}$$

$$0 \leq \alpha_i^* \leq C \qquad \text{Eq. 4.33}$$

Where α_i and α_i^* are non-negative multipliers. The weight coefficients can be calculated from the multipliers obtained by solving the problem in (Eqs. 4.30–4.33) as follows:

$$\beta = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) x_i \; \forall i \in N \qquad \text{Eq. 4.34}$$

Eq. 4.34 will give a vector that contains a β value for each predictor parameter in the training dataset. The output function of the SVM regression model is therefore given by the following:

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)(x_i^T x) + b \qquad \text{Eq. 4.35}$$

4.6.3. Multilayer Perception Neural Network (MLP-NN)

We used a Multilayer Perceptron (MLP) algorithm, a back-propagation neural network, which is considered the most typical deep neural network model (Kain, 2018). Back-propagation consists of two main steps. First, a training feed-forward network is estimated by calculating the gradient from the input parameters and moving forward toward the hidden layer and then from the hidden layer to the output layer. Second, the error is calculated, and it is propagated back from the last layer to the previous layers.

MLP-NN entails three or more layers (input layer, hidden layer(s), and output layer), where each layer is represented as $f(x) = f (w_i x_i + b)$, where f is the activation function, w is the weights in the layer, x is the input parameter, and b is the bias vector. The input layer represents a set of nodes [$x_{j,j} = 1, 2, 3, \cdots\cdots, m$] for all the parameters. While each node in the hidden layer transforms the values from the input layer with a weighted

linear summation $[w_1x_1, w_2x_2, w_3x_3, \cdots\cdots, w_mx_m]$, followed by a non-linear activation function. The output layer receives values from the hidden layer and transforms them into an output value (Malik et al., 2020).

The MLP outputs include coefficients and intercepts; the coefficient is a list of weight matrices where the weight matrix at index I represents the weights between layers i, j, and k. The intercepts are a list of bias vectors at index i that represents the bias values added to layer j.

4.6.3.1. Radial-Basis Neural Network (RB-NN)

The training process in RB-NN has a similar structure to the MLP-NN, but it is faster since it requires relatively fewer interactions. Besides, in MLP, we only have linear separability because they are composed of input and output layers (some hidden layers in MLP), and it needs at least one hidden layer to derive a non-linearity separation. But in the RBNN, the model transforms the input signal into another form, which can then be fed into the network to get linear separability (Oblitas et al., 2021; Samui et al., 2017). RB-NN consists of input, hidden, and output layers. However, it includes only one hidden layer called the feature vector.

4.6.4. Descriptive Statistics

The correlation coefficients between the input parameters are listed in Table 4.5. It indicates either the lack of significant correlation or the absence of correlation in some cases. However, there is a strong correlation between route length L and stop density S_D (-0.979). Therefore, we excluded the route length parameter from the analysis.

Table 4-5. The correlation matrix for the input parameters

	g	D_{Agg}	R_C	HVAC	P_L	S_D	V_a	SoC_i	C_d	L
g	1.000									
D_{Agg}	0.000	1.000								
R_C	0.000	0.000	1.000							
HVAC	0.000	0.000	0.000	1.000						
P_L	0.000	0.000	0.000	0.000	1.000					
S_D	0.000	0.025**	0.000	0.000	0.000	1.000				
V_a	0.000	0.491**	0.000	0.000	0.000	-0.388**	1.000			
SoC_i	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000		
C_d	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	
L	0.000	-0.023**	0.000	0.000	0.000	**-0.979****	0.389**	0.000	0.000	1.000

** Correlation is significant at the 0.01 level (2-tailed)

The descriptive statistics of all parameters are listed in Table 4.6 for both the training and testing datasets. The resultant energy consumption rates range between -2.392 kWh/km and 5.359 kWh/km for the training dataset and from -3.468 kWh/km to 7.237 kWh/km for the testing dataset. It should be noted that negative energy consumption values are attributed to scenarios with constant negative grades, whereas the energy harnessed from the regenerative brake is higher than the energy consumed by the bus. The average number of passengers is approximately 38 passengers, while the stop density ranged between 1.651 stops/km and 3.486 stops/km. The driver aggressiveness and the road condition have the mean (2.0) and standard deviation of (0.816) since both are divided into three levels.

The skewness and kurtosis are computed to test the normality of the dataset. The skewness between 0.5 and -0.5 shows that the distribution is approximately symmetric (George and Mallery, 2010). Thus, all the utilized parameters are normally distributed except; S_D and HVAC testing data, which are skewed to the right. While the kurtosis between 2.0 and -2.0 is considered acceptable to prove the normal distribution (George and Mallery, 2010). The kurtosis values for the entire dataset are within this range (Table 4.6).

Table 4-6. Descriptive statistics of the training and testing dataset

Dataset	Parameter	Minimum	Maximum	Mean	Std. Error	Std. deviation	Skewness	Kurtosis
Training Dataset (907,199)	E_C	-2.490	6.119	1.654	0.002000	1.609	0.055	-0.890
	g	-6.000	6.000	0.000	0.042000	4.000	0.000	-1.250
	D_{Agg}	1.000	3.000	2.000	0.001000	0.816	0.000	-1.500
	R_C	1.000	3.000	2.000	0.001000	0.816	0.000	-1.500
	HVAC	1.250	13.750	6.242	0.004893	4.661	0.461	-1.359
	P_L	0.000	75.000	34.000	0.028000	26.344	0.293	-1.190
	S_D	1.651	3.486	2.377	0.000656	0.625	0.477	-1.194
	V_a	19.986	49.954	31.964	0.007147	6.807	0.459	-0.559
	SoC_i	40.000	100.000	70.00	0.000235	22.361	0.000	-1.360
	C_d	0.500	0.700	0.600	0.000086	0.082	0.000	-1.500
Testing Dataset (169,344)	E_C	-3.468	7.237	1.491	0.004917	2.023	0.059	-0.810
	g	-8.000	8.000	0.000	0.012000	5.078	-0.020	-1.174
	D_{Agg}	1.000	3.000	2.000	0.002000	0.816	0.000	-1.500
	R_C	1.000	3.000	2.000	0.002000	0.816	0.000	-1.500
	HVAC	1.250	22.300	7.614	0.016737	6.888	1.155	0.162
	P_L	0.000	75.000	37.500	0.056000	23.049	0.000	-1.209
	S_D	1.651	3.361	2.283	0.001462	0.602	0.662	-0.897
	V_a	19.990	49.950	33.327	0.018335	7.545	0.207	-0.909
	SoC_i	20.000	100.000	60.000	0.063000	25.820	0.000	-1.230
	C_d	0.6000	0.6000	0.600	0.000000	0.000	-	-

4.7. Results of the Prediction Models

The training dataset is used in seven models to predict BEB energy consumption. The developed models are based on the aforementioned data-driven modelling techniques that include multiple linear regression model (MLR), radial basis function interpolation model (RBF), decision tree model (DT), gradient boosting decision tree (GBDT), Support Vector Machine (SVM), and multilayer perception neural network model (MLP-NN), and radial-basis neural network (RB-NN). It should be noted that the influence of each parameter on the E_C of BEB is assessed through a sensitivity analysis of the results, which is presented in Section 4.8.3.

4.7.1. Multiple Linear Regression Model (MLR)

Linearity, normality, multicollinearity, and homoscedasticity were evaluated to verify the

multivariate statistical assumptions for the MLR model. First, the linearity test confirmed the linear relationship between the dependent and independent parameters through a high coefficient of determination (R^2). Second, normality was confirmed through the errors between observed and predicted values (i.e. residuals), as shown in Figure 4.3.a. Third, multicollinearity was checked using correlations, variance inflating factor (VIF), and tolerance. The results showed no multicollinearity between the parameters, where the VIF values varied between 1.0 and 2.3 (less than 5), and the tolerance ranged between 0.4 and 1.0 (greater than 0.2). These findings are taken together to confirm the absence of multicollinearity between the parameters (James GM, 2013). Finally, homoscedasticity stated that the variance of error terms was similar across the independent parameters' values.

The model goodness-of-fit is evaluated using the coefficient of determination (R^2), t-test, and root mean squared error (RMSE) (Table 4.7). The selected model shows a robust relationship between the dependent and independent parameters with an R^2 of 0.961, indicating that the proposed model explained 96.1% of the variance in energy consumption. The t-test values exceeded $\pm$ 1.96 indicating a significant difference in the effect of the independent parameters on the E_C. While the root mean squared error shows the absolute fit of the model's predicted values to the observed training dataset with an error of 0.3163.
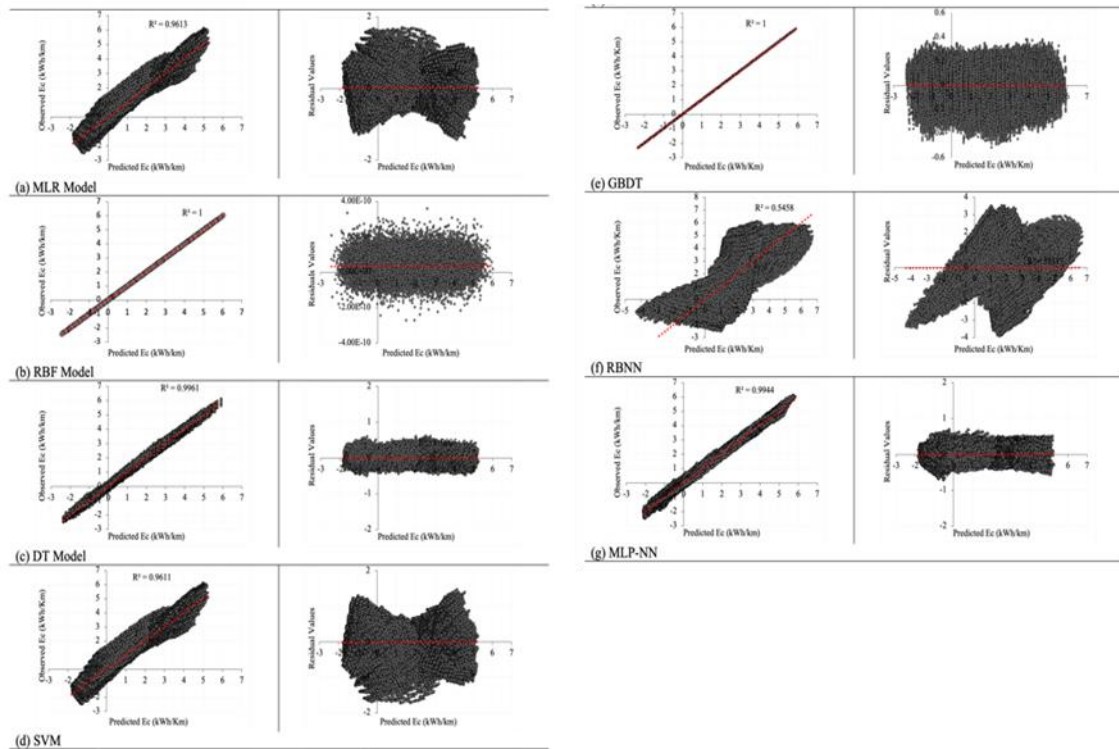
Figure 4-3. The normal probability and the residuals plots for all predicted data-driven models

Table 4-7. Results of the MLR model

Parameters	Unstandardized Coefficients		Standardized Coefficients	t-test	Sig.	Collinearity Statistics		Adjusted R Square	RMSE
	B	Std. Error	Beta			Tolerance	VIF		
(Constant)	-0.885	0.004	-	-228.619	0.000	-	-		
g	0.380	0.000	0.945	4576.930	0.000	1.000	1.000		
SoC$_i$	0.012	0.000	0.172	835.366	0.000	1.000	1.000		
R$_C$	0.260	0.000	0.132	639.197	0.000	1.000	1.000		
HVAC	0.036	0.000	0.105	507.350	0.000	1.000	1.000	0.961	0.3163
P$_L$	0.005	0.000	0.078	380.050	0.000	1.000	1.000		
D$_{Agg.}$	0.065	0.000	0.033	133.592	0.000	0.704	1.420		
S$_D$	0.128	0.001	0.050	213.539	0.000	0.788	1.269		
V$_a$	0.007	0.000	0.030	111.491	0.000	0.599	1.670		
Cd	0.173	0.004	0.009	42.416	0.000	1.000	1.000		

The model indicates the significant impact of all independent parameters on energy consumption, with varying magnitude. The significant parameters include road gradient, the initial state of charge, road condition, HVAC, passenger loading, driver aggressiveness,

stop density, average speed, and drag coefficient. The MLR model is expressed mathematically as:

$$E_C = -0.885 + 0.380 \, g + 0.012 \, SoC_i + 0.260 \, R_C + 0.036 \, HVAC + 0.005 \, P_L + 0.065 \, D_{Agg} + 0.128 \, S_D + 0.007 \, V_a + 0.173 \, C_D$$

Eq. 4.36

The standardized coefficient (betas) is used to compare the impact of each independent variable on the energy consumption. Accordingly, road gradient has the most substantial impact on the E_C (Std. $\beta = 0.945$), followed by the initial state of charge (Std. $\beta = 0.172$), road condition (Std. $\beta = 0.132$), HVAC (Std. $\beta = 0.105$), passenger loading (Std. $\beta = 0.078$), driver aggressiveness (Std. $\beta = 0.033$), stop density (Std. $\beta = 0.050$), average speed (Std. $\beta = 0.030$), while the drag coefficient has the lowest weight in affecting the E_C (Std. $\beta = 0.009$).

4.7.2. Radial Basis Function Interpolation Model (RBF)

The RBF model's validation is assessed using goodness-of-ft, which is evaluated using the coefficient of determination (R^2). Besides, error measurements, including root mean squared error (RMSE), mean square error (MSE), and mean absolute error (MAE), are used to compute the different error values between the predicted E_C from the RBF model and the actual observed E_C values.

The RBF results demonstrate that the $R^2 = 1$ and the error measurements values = zero. The reason is that interpolation models pass through their training data points, which means that when the same training data are used to input to the model, the output will be

the same. This was tested later using the testing dataset.

To validate the results, graphical analysis is performed between the predicted E_C from the RBF model versus the observed E_C values, as shown in Figure 4.3.b. It can be noticed from the figure that the model results in a perfect correlation between the predicted and observed E_C.

Table 4-8. The relationships between the variables in the RBF model

Input Parameter	Relationship
Road Gradient	0.965
SoC initial	0.164
Drag Coefficient	-0.013
Average Speed	0.043
Stop Density Ratio	0.063
Passenger load	0.047
Driver Aggressiveness	0.010
Road Condition	0.083
HVAC	0.044

Further, the nonparametric measure of the relationships between the variables is detailed in Table 4.8. The results indicate that, road gradient poses the highest impact on E_C, followed by the initial status of charge, road condition, and stop density.

4.7.3. Decision Tree Model (DT)

The decision tree model is constructed from the training dataset by optimally partitioning the utilized input parameters into 16,575 nodes. Tree nodes in this model consist of 8,387 decision nodes and 8,188 leaf nodes, where the number of decision nodes for each input parameter is allocated as per Table 4.9. Figure 4.4 illustrates the top four layers of the constructed regression model. As depicted in the figure, the road gradient (g) is assigned by the model as the root node, followed by the initial state of charge (SoC_i) and road

condition (R_C), which implies that they are the most important input parameters for the prediction model.

Table 4-9. The number of decision nodes for each input parameters

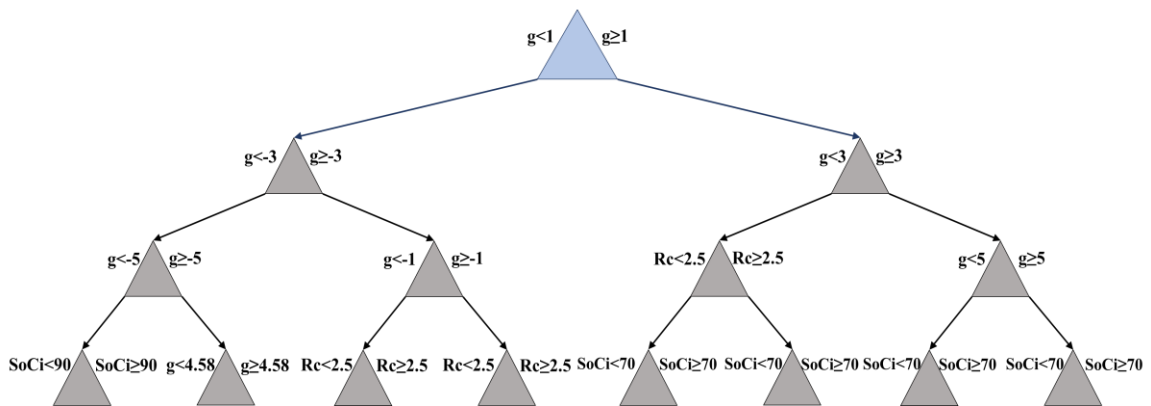| Input Parameter	Number of Decision Nodes
Road Gradient	6
SoC initial	461
Drag Coefficient	0
Average Speed	2634
Stop Density Ratio	2125
Passenger load	1227
Driver Aggressiveness	380
Road Condition	547
HVAC	1007

Figure 4-4. Upper section in the DT Model

The model validity is tested using the goodness-of-ft, which is evaluated using the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). The error indicators (MSE = 0.0102, RMSE = 0.1010, MAE = 0.0776) are relatively low, and the R^2 is 0.996, which shows an exceptional relationship between the input parameters and the energy consumption, as shown in Figure 4.3.c. The residuals distribution is symmetrically and scattered around the zero line, indicating that the model has a strong fit for the dataset.

4.7.4. Gradient-Boosting Decision Tree (GBDT)

In the GBDT model, the number of trained weak decision tree models is 35. The weak models are assigned an equal weight of 1, with the GBDT model response representing the aggregated response of these weak models. In order to estimate the importance of all model parameters, we considered all decision trees in the GBDT. For each decision tree, we summed the changes in MSE values in each parameter split and divided the sum by the number of branch nodes, and then averaged the estimates over all the decision trees in the model. The error indicators (MSE = 0.0078, RMSE = 0.0884, MAE = 0.0683) are relatively low, and the R^2 is 0.997, as shown in Figure 4.3.e. The normalized importance of model parameters is listed in Table 4.10.

Table 4-10. The importance of the input parameters in the GBDT model

Parameter	Importance
Road Gradient	1.0000
SoC initial	0.0380
Drag Coefficient	0.0000
Average Speed	0.0142
Stop Density Ratio	0.0103
Passenger load	0.0159
Driver Aggressiveness	0.0033
Road Condition	0.0213
HVAC	0.0124

4.7.5. Support Vector Machine Learning

The goodness-of-fit measures of the SVM model are utilized to assess the model. The error indicators (MSE = 0.1008, RMSE = 0.3175, MAE = 0.2405) are relatively low, and the R^2 is 0.961. This shows that the input parameters and the energy consumption have a very

strong association, as shown in Figure 4.3.d. The beta factors for the input parameters are

shown in Table 4.11.

Table 4-11. Beta factors β showing the importance of each parameter in the model.

Parameter	Beta factor
g	6.52
SoC_i	1.087
C_d	0.063
V_a	0.221
S_D	0.32
P_L	0.533
$D_{Agg.}$	0.213
R_C	0.919
HVAC	0.727

Similar to all other models, the SVM results show that the road gradient is the

dominant factors for BEBs energy consumption, followed by the initial status of charge and

road conditions.

4.7.6. Multilayer Perception Neural Network Model (MLP-NN)

The trained neural network consists of nine input parameters and a single hidden layer that

has eight neurons, as shown in Figure 4.5. Each link in the network has a weight that

determines the influence of each input parameter on each neuron (Table 4.12) and the

influence of each neuron on the BEBs energy consumption (Figure 4.5).

The selected model shows a superior relationship between the input and output

parameters with an R^2 of 0.994, indicating that the proposed model explained 99.4% of the

energy consumption variance. The mean squared error and the root mean squared error

indicate the model's predicted values' utter fit to the observed training dataset with errors
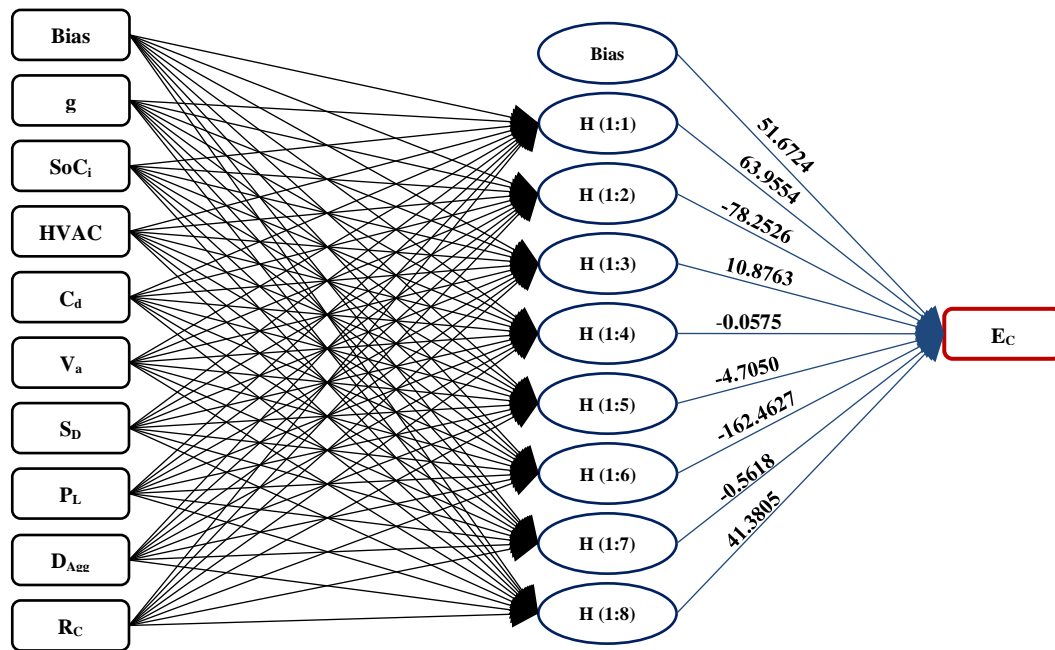
of 0.0146 and 0.1207, respectively.

Figure 4-5. The developed DLNN model including the input, hidden, and output layers

Table 4-12. The weights from the input parameters to the hidden layer's neurons.

	H (1:1)	H (1:2)	H (1:3)	H (1:4)	H (1:5)	H (1:6)	H (1:7)	H (1:8)
Bias	-0.0080	-0.9384	-3.1602	30.4437	0.6571	0.3725	1.0179	2.3411
g	-0.1471	0.7077	-1.7818	-17.1978	1.7828	-0.4078	-1.3180	0.1819
SoC$_i$	0.0960	-0.3043	0.7293	9.8315	-0.6795	0.1880	-0.1328	0.0246
HVAC	-0.0076	0.0573	-0.1548	-1.2134	0.1442	-0.0323	0.1055	0.0133
C$_d$	-0.0038	0.0016	0.0010	-0.0452	0.0010	-0.0023	-0.0086	-0.0005
V$_a$	-0.4956	0.1023	0.5927	-2.2426	-0.2958	-0.2464	-1.1487	-0.2084
S$_D$	-0.2520	0.1130	0.1384	-0.3356	0.0627	-0.1570	-0.5310	-0.1001
P$_L$	-0.0340	0.0104	0.0640	0.6776	-0.0040	-0.0252	-0.7445	-0.1030
D$_{Agg}$	0.0732	-0.0396	-0.0317	-1.5123	-0.0541	0.0506	0.1458	0.0467
R$_C$	-0.0243	0.0939	- 0.2366	47.9737	0.2249	-0.0567	-0.1470	0.0209

The strong linear relationship between the observed and predicted energy consumption shows the DLNN model's normality, as shown in Figure 4.3.g. While the residuals, which represent the distance between the observed and the predicted data points, are very close to the zero-line, which confirms the immense accuracy of the predicted

DLNN model. The mathematical description of the developed DLNN model is represented as follows:

$$H(1:1) \text{ to } H(1:8) = \text{Bias} + \left(\sum_{i=1}^{9} x_i * w_i\right) = \text{Bias} + g * w_g + SoC_i * w_{SoCi} +$$

$$HVAC * w_{HVAC} + C_d * w_{C_d} + V_a * w_{V_a} + S_D * w_{S_D} + P_L * w_{P_L} + D_{Agg} * w_{D_{Agg}} + R_C *$$

$$w_{R_C}$$

Eq. 4.37

$$E_C = 51.6724 + 63.9554H(1:1) - 78.2526H(1:2) + 10.8763H(1:3) -$$

$$0.0575H(1:4) - 4.7050H(1:5) - 162.4627H(1:6) - 0.5618H(1:7) +$$

$$41.3805H(1:8)$$

Eq. 4.38

4.7.7. Radial-Basis Neural Network (RB-NN)

The trained neural network consists of nine input parameters and a hidden layer that has ten neurons. Each link in the network has a weight that determines the influence of each input parameter on each neuron and the influence of each neuron on the BEBs energy consumption (Table 4.13).

Table 4-13. The weights between the input parameters, the hidden layer, and the output parameter

Parameter	H(1)	H(2)	H(3)	H(4)	H(5)	H(6)	H(7)	H(8)	H(9)	H(10)
g	-0.204	0.004	0.232	0.381	-0.082	0.101	0.582	0.174	-0.908	-0.303
SoC_i	0.372	-0.362	0.114	0.226	-0.117	0.133	0.152	-0.030	-0.224	-0.163
HVAC	-0.677	0.605	0.115	-0.471	-0.673	-0.654	1.117	1.169	1.146	-0.715
C_d	-0.077	0.072	0.090	0.381	-0.177	-0.009	0.033	0.098	-0.322	-0.086
V_a	-0.795	-0.829	0.054	0.619	-0.148	1.070	-0.189	1.008	0.215	0.025
S_D	1.055	1.072	1.370	-0.475	-0.344	-0.539	-0.520	-0.574	-0.709	-0.669
P_L	-0.211	-0.051	0.108	0.172	1.090	-0.002	0.016	0.002	-0.091	-0.651
$D_{Agg.}$	-0.400	-0.504	1.192	0.393	-0.408	1.083	-0.520	0.938	-0.133	-0.384
R_C	-0.251	0.100	0.130	1.084	-0.201	-0.593	0.227	-0.358	-0.069	-0.071
Hidden Unit Width	1.162	1.236	1.189	1.194	1.161	1.107	1.165	1.151	1.133	1.115
E_C	-0.639	-0.165	0.746	1.235	-0.382	-0.238	4.151	0.969	-5.293	-1.650

The selected model shows a relationship between the input parameters and the BEB energy consumption with an R^2 of 0.896. The MSE and the RMSE indicate the model's predicted values fit the observed values in the training dataset with errors of 0.452 and 0.672, respectively. The linear relationship between the observed and predicted energy consumption questions the RBNN model's accuracy, as shown in Figure 4.3.f.

4.8. Discussion

4.8.1. Discussion of the Result

The results show the significance of the vehicular, operational, topological, and external parameters on the BEB's energy consumption, with varying magnitude. The significant parameters include the road gradient, the initial state of charge, road condition, HVAC, passenger loading, driver aggressiveness, average speed, stop density and drag coefficient.

The developed data-driven models have a powerful goodness-of-fit with R^2 (based on the training dataset – Figure 4.3) ranging from 0.961, 1.000, 0.996, 0.961, 1.000, 0.896, and 0.994 for MLR, RBF, DT, SVM, GBDT, RBNN, and MLPNN models, respectively. Therefore, all developed models, except RBNN, can explain more than 96% of the variance in the energy consumption rates for the scenarios similar to those used in developing the models. Also, the residual errors are symmetrically distributed around the predicted E_C axis (x_{axis}), as shown in Figure 4.5, with a root mean square error below 0.32 for all the predicted models, which indicates the strong fit of the models when applying on the training dataset.

4.8.2. Models' Validation Using Testing Dataset

Towards a more robust validation for the developed data-driven models, we used the testing

dataset (n = 169,344) to compare the predicted E_C from the validated simulation model to the E_C predicted from the data-driven models. In this respect, we plugged in the input parameters in the four developed data-driven prediction models to calculate the energy consumption rates. Then, we compared the calculated E_C from the models with the original observed E_C rates of the testing dataset (Figure 4.6).
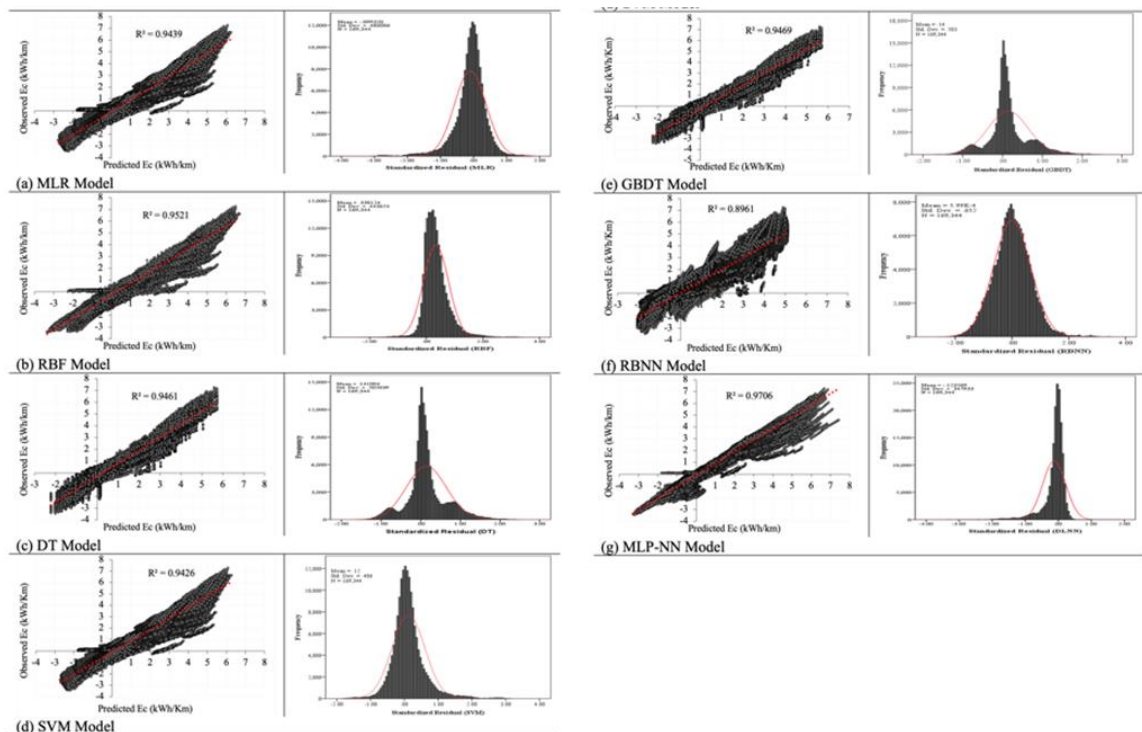

Figure 4-6. The normal probability and standardized residual distribution plots

The linear relationship between the observed and predicted E_C clearly exists in all the developed models. The R^2 values are 0.943, 0.952, 0.946, 0.942, 0.946, 0.546, and 0.971 for MLR, RBF, DT, SVM, GBDT, RBNN, and MLPNN models, respectively. The values indicate a high fitting of the results derived from the developed model and the simulated E_C. However, that does not confirm the accuracy of the utilized data-driven techniques.

First, the E_C rates predicted from the RBF and MPL-NN models are not symmetric around the inclined line (see Figures 4.6.b and 4.6.g), which means that the predicted and observed E_C values will not be close to each other. Further, the residual term distribution for the RBF model (Histograms in Figure 4.6.b) is gradually skewed towards the negative value, resulting in less predicted energy consumption than expected. While the MLP-NN model (Figure 4.6.g), the residual distribution is sharply skewed towards the positive value, which results in dramatically more predicted energy consumption than expected.

Second, both DT models (Figures 4.6.c & 4.6.e) show slight improvements to the RBF or the DLNN models. The predicted E_C values are scattered around the inclined line but not firmly symmetric around it, especially at the end of the line (with higher E_C rates). That indicates that DT models are poor in predicting E_C rates of more than 4.5 kWh/km. The residual distribution of both models (DT & GBDT) is slightly skewed toward the left, which results in slightly less predicted E_C than expected. Further, the residuals are not normally distributed for both models, raising some questions about their prediction accuracy.

Third, for the RBNN, and despite having normally distributed residuals (Figure 4.6.f), there is a large spread of the predicted E_C resulting in lower R^2 compared to all other models.

Fourth, both MLR and SVM models (Figures 4.6.a & 4.6.d, respectively) show promising results. The predicted E_C data is almost at a slope of 1, which results in more adjacent observed and predicted E_C rates. In addition, the residuals are normally distributed with a marginally skewed towards the positive side (in MLR model), which results in

marginally more predicted energy consumption than expected, as shown in Figure 4.8.a. In contrast, the opposite is true for the SVM model, with a marginal negative skewness.

Another way to assess the accuracy of the models is by examining how the variation in the input parameters affects the energy consumption rates; therefore, a sensitivity analysis is conducted for all the developed data-driven models.

4.8.3. Sensitivity Analysis

Sensitivity analysis studies how the output target varies in response to each input variation (Plischke et al., 2013). In this work, a Sobol analysis, which is a global variance-based sensitivity analysis method, is performed on the four data-driven prediction models (Sobol, 1993). This method is chosen because it considers the whole input domain and can measure the parameter's interactions. Sobol analysis calculates the variance of the model response and decomposes it in terms of input parameter contributions. Two Sobol indices are studied, which are first-order effect index S_i that represents the effect of X_i alone, and total-effect index S_{Ti} that takes into consideration the interactions of X_i with other parameters as follows:

$$S_i = \frac{V\left[E\left(\frac{Y}{X_i}\right)\right]}{V(Y)} \qquad \text{Eq. 4.39}$$

$$S_{Ti} = 1 - \frac{V[E(Y/X_{\sim i})]}{V(Y)} \qquad \text{Eq. 4.40}$$

where V[.] and E(.) are the variance and expected values, respectively. For all four models in this study, Sobol sampling (Sobol, 1967) and Saltelli estimator (Saltelli et al., 2007) are utilized to estimate these two indices.

The sensitivity results show that road gradient is the most significant parameter impacting E_C in all the developed models for both first-order and total-order effects, as shown in Table 4.14 and Table 4.15. The SoC_i has the second-highest impact on the E_C for all models except the MLP-NN model. While driver aggressiveness and drag coefficient have the lowest impact on the E_C for all seven models. The order effect of the other parameters is varying among the models, so for a better description, Figure 4.7 is presented to show the variation using Sobol indices values.
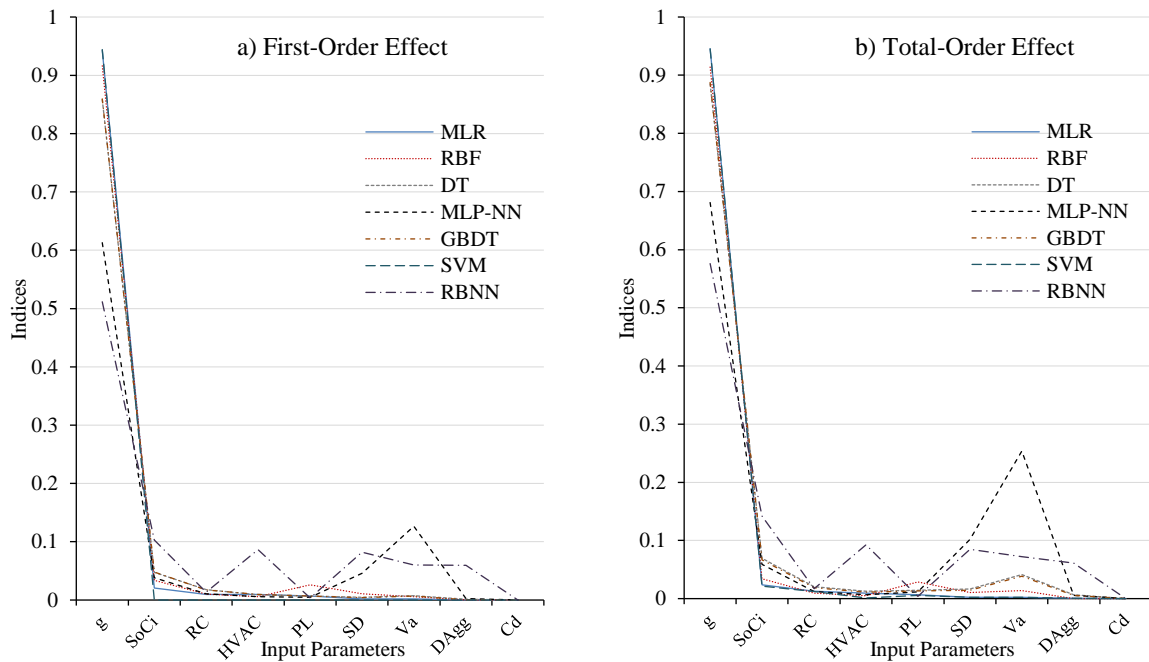

Figure 4-7. The Sobol indices of (a) first-order and (b) total-order effects for the four developed data-driven models

Table 4-14. The order effects of the independent parameters for the developed models

Order-effect	MLR		RBF		DT		MLP-NN		GBDT		SVM		RBNN	
	First-order effect	Total-order effect	First-order effect	Total-order effect	First-order effect	Total-order effect	First-order effect	Total-order effect	First-order effect	Total-order effect	First-order effect	Total-order effect	First-order effect	Total-order effect
1	g	g	g	g	g	g	g	g	g	g	g	g	g	g
2	SoC_i	SoC_i	SoC_i	SoC_i	SoC_i	SoC_i	V_a	V_a	SoC_i	SoC_i	SoC_i	SoC_i	SoC_i	SoC_i
3	HVAC	R_C	P_L	P_L	R_C	V_a	S_D	S_D	R_C	V_a	R_C	R_C	HVAC	HVAC
4	R_C	HVAC	S_D	V_a	HVAC	R_C	SoC_i	SoC_i	HVAC	R_C	HVAC	HVAC	S_D	S_D
5	P_L	P_L	R_C	S_D	V_a	S_D	R_C	R_C	V_a	S_D	P_L	P_L	V_a	V_a
6	S_D	S_D	HVAC	RC	P_L	P_L	HVAC	P_L	P_L	P_L	S_D	S_D	D_{Agg}	D_{Agg}
7	V_a	V_a	V_a	HVAC	S_D	HVAC	P_L	HVAC	S_D	HVAC	V_a	V_a	R_C	R_C
8	D_{Agg}	D_{Agg}	D_{Agg}	D_{Agg}	D_{Agg}	D_{Agg}	C_d	D_{Agg}	D_{Agg}	D_{Agg}	D_{Agg}	D_{Agg}	P_L	P_L
9	C_d	C_d	C_d	C_d	C_d	C_d	D_{Agg}	C_d	C_d	C_d	C_d	C_d	C_d	C_d

Table 4-15. Sobol indices of the independent parameters for the developed models

Parameters	First-order effect							Total-order effect						
	MLR	RBF	DT	MLP-NN	GBDT	SVM	RBNN	MLR	RBF	DT	MLP-NN	GBDT	SVM	RBNN
g	0.9428	0.9162	0.8584	0.6129	0.85961	0.9442	0.51155	0.9455	0.9139	0.8850	0.6810	0.88752	0.9452	0.5760
SoCi	0.0204	0.0336	0.0477	0.0384	0.04791	0.0225	0.10289	0.0237	0.0340	0.0695	0.0588	0.06674	0.0214	0.1422
RC	0.0094	0.0100	0.0176	0.0106	0.01737	0.0123	0.01373	0.0125	0.0093	0.0202	0.0124	0.01902	0.0119	0.0169
HVAC	0.0096	0.0056	0.0090	0.0059	0.00945	0.0084	0.08653	0.0092	0.0052	0.0129	0.0064	0.01105	0.0095	0.0924
PL	0.0068	0.0263	0.0066	0.0045	0.00677	0.0060	0.00360	0.0064	0.0286	0.0145	0.0115	0.01322	0.0058	0.0040
SD	0.0018	0.0107	0.0049	0.0456	0.00396	0.0023	0.08182	0.0025	0.0105	0.0168	0.1018	0.01446	0.0023	0.0850
Va	0.0018	0.0055	0.0072	0.1265	0.00706	0.0013	0.06016	0.0020	0.0140	0.0411	0.2535	0.03828	0.0020	0.0723
DAgg	0.0005	0.0005	0.0012	0.0027	0.00019	0.0005	0.05972	0.0008	0.0001	0.0064	0.0054	0.00575	0.0006	0.0613
Cd	0.0002	0.0002	0.0000	0.0001	0.00000	0.0003	0.00000	5×10^{-5}	5×10^{-5}	0.0000	0.0001	0.00000	0.0001	0.0000

In short, the values of the Sobol indices for the utilized parameters using MLR, RBF, DT, GBDT, SVM, MLP-NN, and RBNN show that the road gradient has the highest impact on the predicted E_C rates, while driver aggressiveness and the drag coefficient have the lower impact on the predicted E_C rates. For the remaining parameters, there are fluctuations among their order-effect without a clear pattern. Therefore, a sample-based sensitivity analysis is carried out to study how the E_C varies in response to each input parameter variation for both training and testing datasets. Then, we compare the result of this sensitivity analysis with the result of the sensitivity analysis for the four data-driven models to define which model behaviour is closer to the dataset's behaviour.

The sample-based sensitivity for the training and testing datasets is shown in Figure 4.8. Comparing the result with the seven developed models, we found that the MLR model has the closest behaviour to both datasets where the order effect of the parameters are accurately the same (compared to the Std. β) except for the driver aggressiveness parameter.
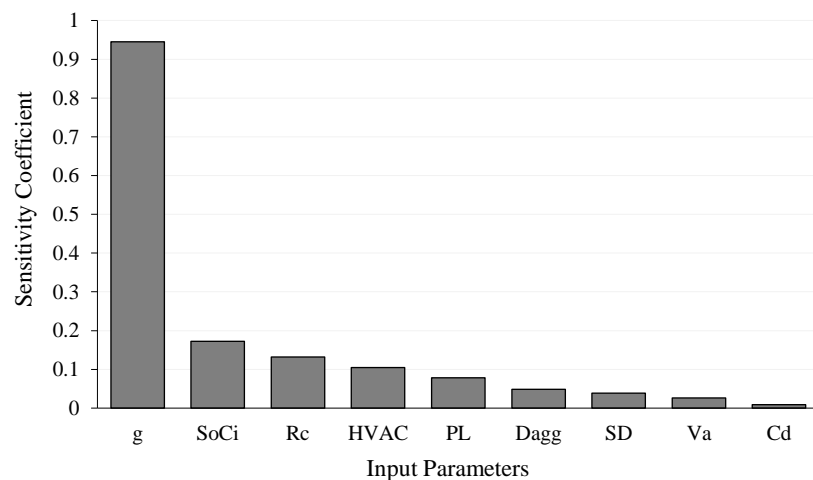

Figure 4-8. A sample-based sensitivity for the training and testing datasets

Overall, we recommend the utilization of MLR and SVM models in the prediction of BEB's E_C in transit applications. This recommendation is based on multiple factors (Table 4.16), including the model results, the goodness-of-ft, and the ease of utilization.

From a statistical modelling perspective, MLR is available in almost all software packages and does not require deep knowledge of machine learning and data-driven models. In addition, the models could be estimated very quickly. From the machine learning family of modelling, we recommend the SVM model. That said, implementing this model by transit providers might be challenging.

Table 4-16. A holistic comparison of the developed models

Model	R^2 (Testing data)	Residual Distribution	Distribution of Predicted E_C	Total-order effect
MLR	0.943	Normal	Slight deviation from the zero-line	Consistent with the literature
RBF	0.952	Skewed - Left	Slight deviation from the zero-line	Consistent with the literature
DT	0.946	Non-normal	Symmetric around the zero-line	Consistent with the literature
GBDT	0.942	Non-normal	Symmetric around the zero-line	Consistent with the literature
SVM	0.946	Normal	Slight deviation from the zero-line	Consistent with the literature
MLP-NN	0.971	Skewed - Right	Slight deviation from the zero-line	Not consistent with the literature
RBNN	0.546	Normal	Significant deviation from the zero-line	Not consistent with the literature

4.9. Conclusion

Estimating the energy consumption rates of battery-electric buses vary significantly due to factors related to vehicular, operational, topological, and external parameters. This variation calls for the development of data-driven modelling techniques to accurately predict E_C rates in varying conditions. The present study started with two overarching aims to 1) Develop BEB energy consumption data-driven models that accommodate vehicular, operational, topological and external parameters. 2) Evaluate and compare the results of seven different data-driven modelling techniques: regression analysis (MLR), interpolation

method (RBF), decision tree (DT), gradient boosting decision tree (GBDT), support vector machine (SVM), multilayer perception neural network (MLP-NN), and radial basis neural network (RBNN). Besides, we performed a sensitivity analysis of these seven models to evaluate the influence of each parameter in the energy consumption variation.

Toward that end, we developed a simulation model to predict the energy consumption rates using MATLAB Simulink to overcome the challenges of collecting real world E_C data that represents all possible operational conditions. The simulation model is validated using the real-world results of the Altoona testing facility. A full-factorial experiment is then used to define all the possible scenarios for BEB operation in transit based on the selected parameters. After that, we used the simulated energy consumption scenarios in developing four different data-driven modelling techniques.

A two-stage validation/assessment technique is developed. First, we utilized a testing dataset based on a fractional-factorial experiment to assess the accuracy of the developed models. Second, a Sobol sensitivity analysis was carried out for the results of these models to evaluate and compare the impact of the utilized parameters on predicting the energy consumption of BEBs.

The results show the significance of the road gradient, the initial state of charge, road condition, HVAC, passenger loading, driver aggressiveness, average speed, stop density and drag coefficient on the BEB's energy consumption, with varying magnitude. The developed data-driven modelling techniques show a powerful goodness-of-fit that can explain more than 96% and 89% of the energy consumption variance using the training and testing datasets, respectively. The validation result for the seven modelling techniques

using the testing dataset shows that the MLR and SVM models are deemed the most accurate models.

The sensitivity analysis confirms that the road gradient is the most significant parameter impacting the E_C. Also, the results show that driver aggressiveness and the drag coefficient have the lowest impact on the E_C in the four data-driven models for both the first-order effect and the total-order effect. Comparing the Sobol sensitivity for the developed models with the sample-based sensitivity for the training and testing datasets, we found that the MLR model has the closest behaviour to the training and testing datasets. Likewise, the order effect of the parameters using MLR coincides with the original dataset, which confirms the high accuracy of the MLR model in estimating the energy consumption rates of BEBs.

In conclusion, both MLR and SVM techniques provide various opportunities to practitioners and scholars to develop accurate, rapid, and reliable data-driven models in BEB energy consumption studies. In addition, the results of the sensitivity analysis provide transit planners with a means to configure the optimal transit operation profile that enhances BEB energy utilization.

Future research studies are encouraged to investigate the proposed models' sensitivity to different BEBs, with different powertrain configurations. This will indeed yield more clarification on the sensitivity of the bus itself to vehicular, operational, topological, and external parameters.
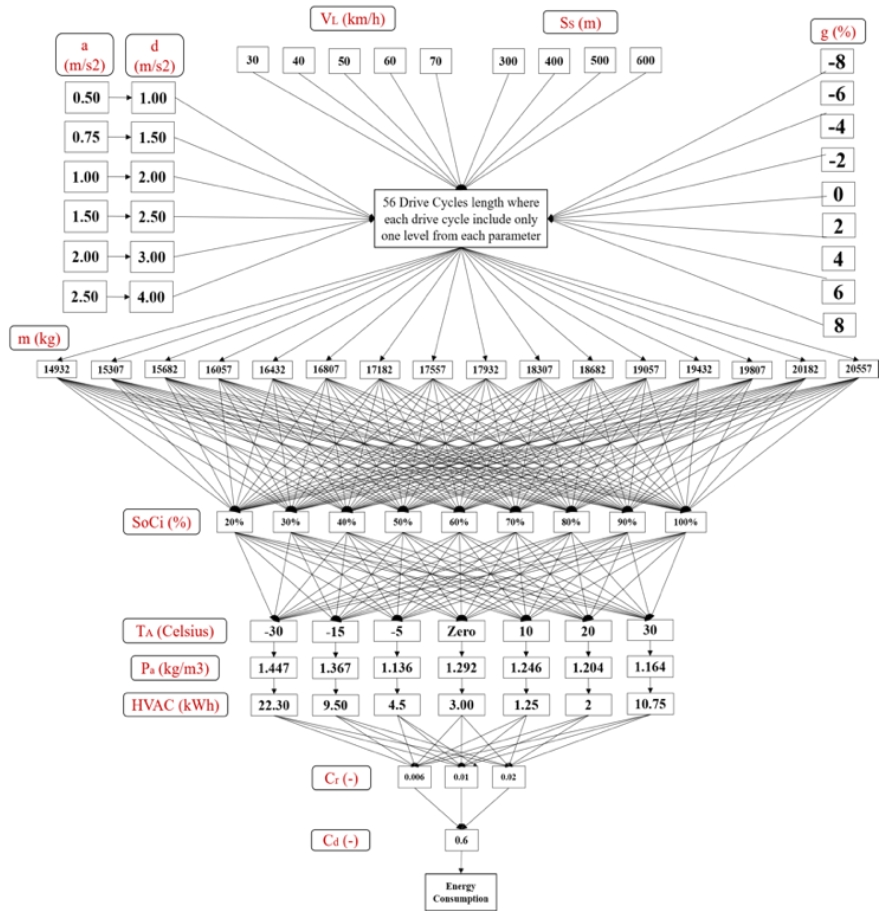
4.10. Acknowledgement

4.11. Appendices

Appendix 4-1. Full and fractional-factorial experimental design
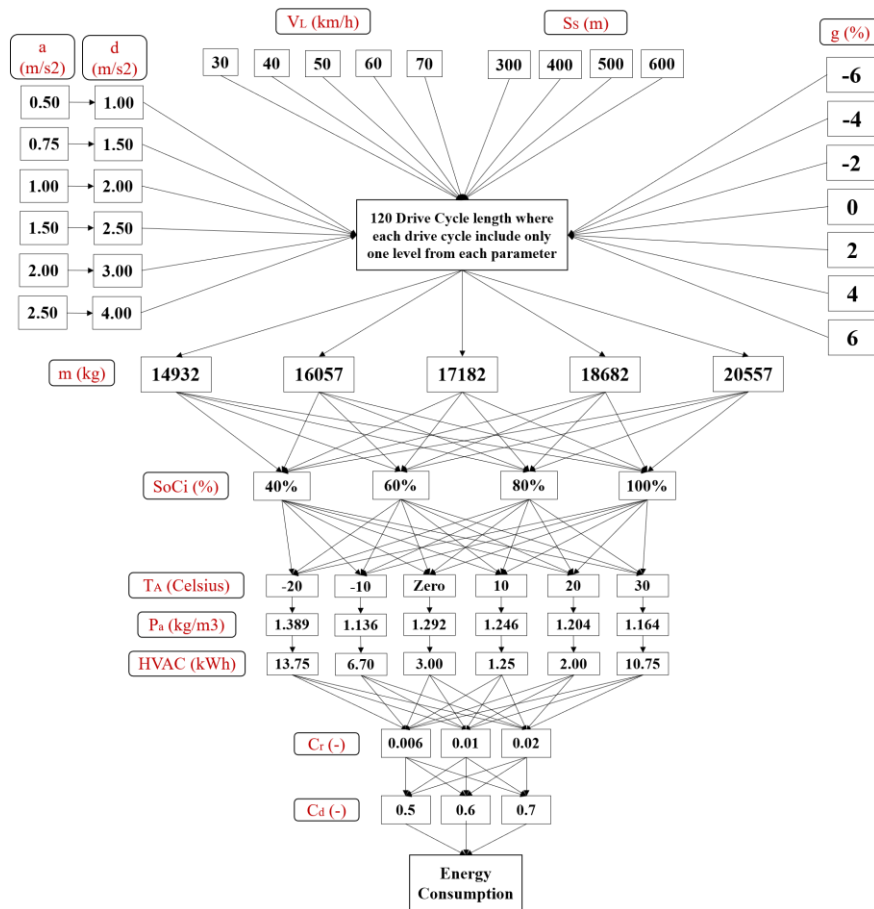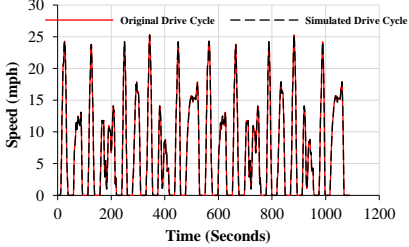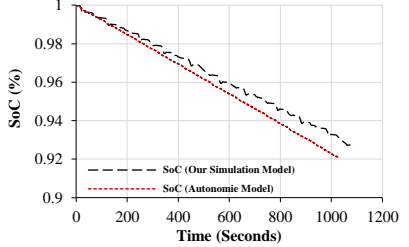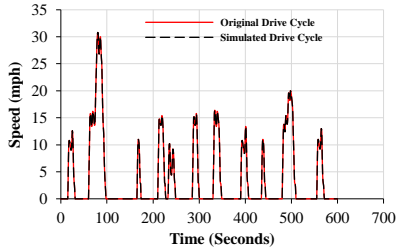
a) Training dataset

b) Testing dataset

Full and fractional Factorial experimental design parameters: (a) Training dataset of 907,199 scenarios and (b) testing dataset of 169,344 scenarios

Appendix 4-2. Simulation-based validation of the energy consumption model

Drive Cycle	Drive Cycles Comparison	SoC Comparison	Δ SoC (%)
Manhattan			6.05%
New York Bus			6.29%
RTE			6.26%

4.12. References

Altoona, 2015. New Flyer - Model: XE40, FEDERAL TRANSIT BUS TEST. Pennsylvania Transportation Institute. http://apps.altoonabustest.psu.edu/buses/reports/ 458.pdf?1441118410.

Abdelaty, H., Mohamed, M., 2021. A prediction model for battery electric bus energy consumption in transit. Energies 14, 2824. https://doi.org/10.3390/ en14102824.
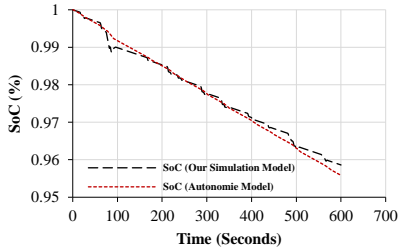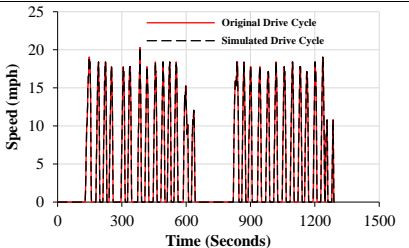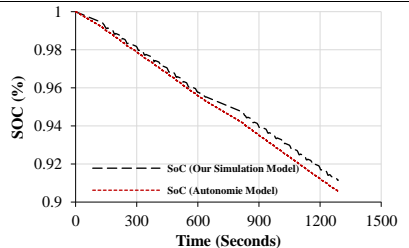
An, K., 2020. Battery electric bus infrastructure planning under demand uncertainty. Transportation Research Part C: Emerging Technologies 111, 572–587.

Bajer, L., Holena, M., 2012. RBF-based surrogate model for evolutionary optimization, ITAT.

Basma, H., Mansour, C., Nemer, M., Stabat, P., Haddad, M., 2020. Sensitivity analysis of bus line electrification at different operating conditions, Proceedings of 8th Transport Research Arena TRA, Helsinki, Finland, pp. 1-10. https://hal.archives ouvertes.fr/hal-02611573.

Beckers, C.J.J., Besselink, I.J.M., Frints, J.J.M., Nijmeijer, H., 2019. Energy consumption prediction for electric city buses, 13th ITS European Congress Eindhoven, Netherlands.

Bor´en, S., 2019. Electric buses' sustainability effects, noise, energy use, and costs. International Journal of Sustainable Transportation 14 (12), 956–971.

Box, G.E.P., Hunter, J.S., Hunter, W.G., 2005. Statistics for Experimenters: Design,

Innovation, and Discovery, 2nd Edition ed.

Canada, 2019. Daily Data Report, environment and natural resources, Weather, Climate and Hazard. Retrieved from Government of Canada.

Chediak, M., 2018. Electric Buses Will Take Over Half the World Fleet by 2025 pp. https://www.bloomberg.com/news/articles/2018-2002-2001/electric-buses willtake-over-half-the-world-by-2025.

Chen, Z., Yin, Y., Song, Z., 2018. A cost-competitiveness analysis of charging infrastructure for electric bus operations. Transportation Research Part C: Emerging Technologies 93, 351–366.

De Cauwer, C., Van Mierlo, J., Coosemans, T., 2015. Energy Consumption Prediction for Electric Vehicles Based on Real-World Data. Energies 8 (8), 8573–8593.

De Cauwer, C., Verbeke, W., Coosemans, T., Faid, S., Van Mierlo, J., 2017. A Data-Driven Method for Energy Consumption Prediction and Energy-Effcient Routing of Electric Vehicles in Real-World Conditions. Energies 10 (5).

De Filippo, G., Marano, V., Sioshansi, R., 2014. Simulation of an electric transportation system at The Ohio State University. Applied Energy 113, 1686–1691.

Dhiman, H.S., Deb, D., Balas, V.E., 2020. Supervised machine learning in wind forecasting and ramp event prediction. Academic Press.

Diaz Alvarez, A., Serradilla Garcia, F., Naranjo, J.E., Anaya, J.J., Jimenez, F., 2014. Modeling the Driving Behavior of Electric Vehicles Using Smartphones and Neural

Networks. IEEE Intelligent Transportation Systems Magazine 6 (3), 44–53.

Dib, W., Chasse, A., Moulin, P., Sciarretta, A., Corde, G., 2014. Optimal energy management for an electric vehicle in eco-driving applications. Control Engineering Practice 29, 299–307.

Dreyfus, G., 2005. Neural Networks: Methodology and Applications. Springer.

Du, Y., Wu, J., Yang, S., Zhou, L., 2017. Predicting vehicle fuel consumption patterns using floating vehicle data. J Environ Sci (China) 59, 24–29.

Durantin, C., Rouxel, J., D´ esid´eri, J.-A., Gli`ere, A., 2017. Multifdelity surrogate modeling based on radial basis functions. Structural and Multidisciplinary Optimization 56 (5), 1061–1075.

El-Taweel, N.A., Farag, H.E., Mohamed, M., 2019. Integrated Utility-Transit Model for Optimal Configuration of Battery Electric Bus Systems. IEEE Systems Journal 14(1), 738–748.

El-Taweel, N.A., Mohamed, M., Farag, H.E., 2017. Optimal design of charging stations for electrified transit networks, 2017 IEEE Transportation Electrification Conference and Expo (ITEC). IEEE 786–791.

Franca, A., 2015. Electricity consumption and battery lifespan estimation for transit electric buses: drivetrain simulations and electrochemical modelling. University of Victoria.

Gallet, M., Massier, T., Hamacher, T., 2018. Estimation of the energy demand of electric

buses based on real-world data for large-scale public transport networks. Applied Energy 230, 344–356.

Galvin, R., 2017. Energy consumption effects of speed and acceleration in electric vehicles: Laboratory case studies and implications for drivers and policymakers. Transportation Research Part D: Transport and Environment 53, 234–248.

Gao, Z., Lin, Z., LaClair, T.J., Liu, C., Li, J.-M., Birky, A.K., Ward, J., 2017. Battery capacity and recharging needs for electric buses in city transit service. Energy 122, 588–600.

George, D., Mallery, P., 2010. SPSS for Windows Step by Step: A Simple Guide and Reference 18.0 Update. Prentice Hall Press, One Lake Street Upper Saddle River, NJUnited States.

Hahn, B., Valentine, D., 2019. SIMULINK® Toolbox. Essential MATLAB. for Engineers and Scientists.

Hair, J.F., Black, W.C., Badin, B.J., Anderson, R.E., 2009. Multivariate data analysis, Seventh, edition ed. Pearson Education Limited, Harlow.

Huang, D., Xie, H., Ma, H., Sun, Q., 2017. Driving cycle prediction model based on bus route features. Transportation Research Part D: Transport and Environment 54, 99–113.

James GM, H.T., Witten D, Tibshirani R. , 2013. An Introduction to Statistical Learning: With Applications in R. Kain, N.K., 2018. Understanding of Multilayer perceptron.

Kanarachos, S., Mathew, J., Fitzpatrick, M.E., 2019. Instantaneous vehicle fuel consumption estimation using smartphones and recurrent neural networks. Expert Systems with Applications 120, 436–447.

Kivekas, K., Vepsäläinen, J., Tammi, K., 2018. Stochastic Driving Cycle Synthesis for Analyzing the Energy Consumption of a Battery Electric Bus. IEEE Access 6, 55586–55598.

Kivekas, ¨ K., Vepsal ¨ ¨ ainen, J., Tammi, K., Anttila, J., 2017. Influence of Driving Cycle Uncertainty on Electric City Bus Energy Consumption, 2017 IEEE Vehicle Power and Propulsion Conference (VPPC), pp. 1-5.

Kontou, A., Miles, J., 2015. Electric Buses: Lessons to be Learnt from the Milton Keynes Demonstration Project. Procedia Engineering 118, 1137–1144.

Kühne, R., 2010. Electric buses – An energy effcient urban transportation means. Energy 35 (12), 4510–4513.

Kumar, R., 2015. Regenerative Brake: To Harness the Kinetic Energy of Braking. International Journal of Emerging Technologies and Innovative Research 2 (1), 124–129.

Kunith, A., Mendelevitch, R., Goehlich, D., 2017. Electrification of a city bus network— An optimization model for cost-effective placing of charging infrastructure and battery sizing of fast-charging electric bus systems. International Journal of Sustainable Transportation 11 (10), 707–720.

Lajunen, A., 2014. Energy consumption and cost-benefit analysis of hybrid and electric

city buses. Transportation Research Part C: Emerging Technologies 38, 1–15.

Lajunen, A., 2018. Lifecycle costs and charging requirements of electric buses with different charging methods. Journal of Cleaner Production 172, 56–67.

Lajunen, A., Kivekäs, K., Baldi, F., Vepsäläinen, J., Tammi, K., 2018. Different Approaches to Improve Energy Consumption of Battery Electric Buses, 2018 IEEE Vehicle Power and Propulsion Conference (VPPC), pp. 1-6.

Laurikko, J., Pihlatie, M., Nylund, N., Halmeaho, T., Kukkonen, S., Lehtinen, A., Karvonen, V., M¨ akinen, R., Ahtiainen, S., 2015. Electric city bus and infrastructure demonstration environment in Espoo, Finland.

Lawson, C.E., Marti, J.M., Radivojevic, T., Jonnalagadda, S.V.R., Gentz, R., Hillson, N.J., Peisert, S., Kim, J., Simmons, B.A., Petzold, C.J., Singer, S.W.,

Mukhopadhyay, A., Tanjore, D., Dunn, J.G., Garcia Martin, H., 2021. Machine learning for metabolic engineering: A review. Metab Eng 63, 34–60.

Li, J.-Q., 2014. Battery-electric transit bus developments and operations: A review. International Journal of Sustainable Transportation 10 (3), 157–169.

Liu, K., Wang, J., Yamamoto, T., Morikawa, T., 2018. Exploring the interactive effects of ambient temperature and vehicle auxiliary loads on electric vehicle energy consumption. Applied Energy 227, 324–331.

Liu, K., Yamamoto, T., Morikawa, T., 2017. Impact of road gradient on energy consumption of electric vehicles. Transportation Research Part D: Transport and

Environment 54, 74–81.

Ma, X., Miao, R., Wu, X., Liu, X., 2021. Examining influential factors on the energy consumption of electric and diesel buses: A data-driven analysis of large-scale public transit network in Beijing. Energy 216.

Mahmoud, M., Garnett, R., Ferguson, M., Kanaroglou, P., 2016. Electric buses: A review of alternative powertrains. Renewable and Sustainable Energy Reviews 62, 673–684.

Malik, A., Rai, P., Heddam, S., Kisi, O., Sharafati, A., Salih, S.Q., Al-Ansari, N., Yaseen, Z.M., 2020. Pan Evaporation Estimation in Uttarakhand and Uttar Pradesh States, India: Validity of an Integrative Data Intelligence Model. Atmosphere 11 (6).

Markel, T., Brooker, A., Hendricks, T., Johnson, V., Kelly, K., Kramer, B., O'Keefe, M., Sprik, S., Wipke, K., 2002. ADVISOR: a systems analysis tool for advanced vehicle modelling. Journal of Power Sources 110 (2), 255–266.

Masikos, M., Demestichas, K., Adamopoulou, E., Theologou, M., 2014. Mesoscopic forecasting of vehicular consumption using neural networks. Soft Computing 19 (1), 145–156.

Mitchell, T., 1997. Textbook: Machine Learning.

Mohamed, M., Farag, H., El-Taweel, N., Ferguson, M., 2017. Simulation of electric buses on a full transit network: Operational feasibility and grid impact analysis. Electric Power Systems Research 142, 163–175.

Mohamed, M., Ferguson, M., Kanaroglou, P., 2018. What hinders adoption of the electric bus in Canadian transit? Perspectives of transit providers. Transportation Research Part D: Transport and Environment 64, 134–149.

Nedelkova, Z., Lindroth, P., Stromberg, ¨A.-B., Patriksson, M., 2015. Integration of expert knowledge into radial basis function surrogate models. Optimization and Engineering 17 (3), 577–603.

NewFlyer-XE40, 2017. Xcelsior Charge Technical Summary - 40' Electric Bus.

Oblitas, J., Mejia, J., De-la-Torre, M., Avila-George, H., Seguí Gil, L., Mayor Lopez, ´L., Ibarz, A., Castro, W., 2021. Classification of the Microstructural Elements of the Vegetal Tissue of the Pumpkin (Cucurbita pepo L.) Using Convolutional. Neural Networks. 11 (4), 1581.

Offer, G.J., Howey, D., Contestabile, M., Clague, R., Brandon, N.P., 2010. Comparative analysis of battery electric, hydrogen fuel cell and hybrid vehicles in a future sustainable road transport system. Energy Policy 38 (1), 24–29.

Pamula, T., Pamula, W., 2020. Estimation of the Energy Consumption of Battery Electric Buses for Public Transport Networks Using Real-World Data and Deep Learning. Energies 13 (9).

Pihlatie, M., Kukkonen, S., Halmeaho, T., Karvonen, V., Nylund, N., 2014. Fully electric city buses - The viable option. In: 2014 IEEE International Electric Vehicle Conference (IEVC), pp. 1–8.

Ping, P., Qin, W., Xu, Y., Miyajima, C., Takeda, K., 2019. Impact of Driver Behavior on

Fuel Consumption: Classifcation, Evaluation and Prediction Using Machine Learning. IEEE Access 7, 78515–78532.

Plischke, E., Borgonovo, E., Smith, C.L., 2013. Global sensitivity measures from given data. European Journal of Operational Research 226 (3), 536–550.

Qi, X., Wu, G., Boriboonsomsin, K., Barth, M.J., 2018a. Data-driven decomposition analysis and estimation of link-level electric vehicle energy consumption under real-world traffic conditions. Transportation Research Part D: Transport and Environment 64, 36–52.

Qi, Z., Yang, J., Jia, R., Wang, F., 2018b. Investigating Real-World Energy Consumption of Electric Vehicles: A Case Study of Shanghai. Procedia Computer Science 131, 367–376.

Quarles, N., Kockelman, K.M., Mohamed, M., 2020. Costs and Benefts of Electrifying and Automating Bus Transit Fleets. Sustainability 12 (10), 3977.

Reimpell, J., Stoll, H., Betzler, J.W., 2001. Tyres and wheels, pp. 86-148.

Rodríguez Pardo, M., 2017. Uncertainty in electric bus mass and its influence in energy consumption, Mechanical engineering. Aalto University, p. 59.

Rupp, M., Handschuh, N., Rieke, C., Kuperjans, I., 2019. Contribution of country-specifc electricity mix and charging time to environmental impact of battery electric vehicles: A case study of electric buses in Germany. Applied Energy 237, 618–634.

Rupp, M., Rieke, C., Handschuh, N., Kuperjans, I., 2020. Economic and ecological

optimization of electric bus charging considering variable electricity prices and CO2eq intensities. Transportation Research Part D: Transport and Environment 81.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2007. Sensitivity Analysis: From Theory to Practice. Global Sensitivity Analysis. The Primer 237–275.

Samui, P., Roy, S.S., Balas, V.E., 2017. Handbook of neural computation. Academic Press.

Shankar, R., Marco, J., 2013. Method for estimating the energy consumption of electric vehicles and plug-in hybrid electric vehicles under real-world driving conditions. IET Intelligent Transport Systems 7 (1), 138–150.

Sobol, I., 1993. Sensitivity Estimates for Nonlinear Mathematical Models.

Sobol, I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Computational Mathematics and Mathematical Physics 7 (4), 86–112.

Tammi, K., Lajunen, A., 2016. Energy consumption and carbon dioxide emission analysis for electric city buses, 29th.

Teoh, L.E., Khoo, H.L., Goh, S.Y., Chong, L.M., 2018. Scenario-based electric bus operation: A case study of Putrajaya, Malaysia. International Journal of Transportation Science and Technology 7 (1), 10–25.

Vepsäläinen, J., Kivekäs, K., Otto, K., Lajunen, A., Tammi, K., 2018a. Development and validation of energy demand uncertainty model for electric city buses.

Transportation Research Part D: Transport and Environment 63, 347–361.

Vepsäläinen, J., Otto, K., Lajunen, A., Tammi, K., 2019. Computationally efficient model for energy demand prediction of electric city bus in varying operating conditions. Energy 169, 433–443.

Vepsäläinen, J., Ritari, A., Lajunen, A., Kivekäs, K., Tammi, K., 2018b. Energy Uncertainty Analysis of Electric Buses. Energies 11 (12).

Wang, H., Qin, Q.-H., 2019. Basics of fundamental solutions and radial basis functions. Methods of Fundamental Solutions in Solid Mechanics 91–105.

Wang, J.-B., Liu, K., Yamamoto, T., Morikawa, T., 2017. Improving Estimation Accuracy for Electric Vehicle Energy Consumption Considering the Effects of Ambient Temperature. Energy Procedia 105, 2904–2909.

Wellik, T.K., Griffn, J.R., Kockelman, K.M., Mohamed, M., 2021. Utility-transit nexus: Leveraging intelligently charged electrified transit to support a renewable energy grid. Renewable and Sustainable Energy Reviews 139 (2021). https://doi.org/10.1016/j.rser.2020.110657.

Wu, J.-D., Liu, J.-C., 2011. Development of a predictive system for car fuel consumption using an artificial neural network. Expert Systems with Applications 38 (5), 4967 4971.

Wu, J.-D., Liu, J.-C., 2012. A forecasting system for car fuel consumption using a radial basis function neural network. Expert Systems with Applications 39 (2), 1883– 1888.

Xu, Y., Gbologah, F.E., Lee, D.-Y., Liu, H., Rodgers, M.O., Guensler, R.L., 2015. Assessment of alternative fuel and powertrain transit bus options using real-world operations data: Life-cycle fuel and emissions modeling. Applied Energy 154, 143–159.

Yamashita, R.-J., Yao, H.-H., Hung, S.-W., Hackman, A., 2018. Accessing and constructing driving data to develop fuel consumption forecast model. IOP Conference Series: Earth and Environmental Science 113.

Yang, X.-S., 2019. Introduction to algorithms for data mining and machine learning. Academic press.

Yuan, X., Zhang, C., Hong, G., Huang, X., Li, L., 2017. Method for evaluating the real world driving energy consumptions of electric vehicles. Energy 141, 1955–1968.

Zeng, W., Miwa, T., 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, 906–921.

Zhang, R., Yao, E., 2015. Electric vehicles' energy consumption estimation with real driving condition data. Transportation Research Part D: Transport and Environment 41, 177–187.

CHAPTER FIVE

5. BATTERY-ELECTRIC BUS ENERGY PREDICTION FRAMEWORK USING A LOW-RESOLUTION OPEN-SOURCE DATA-DRIVEN MODEL

Abstract: The accurate estimation of the energy consumption (E_C) rates of Battery Electric Buses (BEBs) might be ambiguous due to uncertainties in amassing real-world operation data such as speed profiles, route topology, passenger loading, and weather conditions. Therefore, attaining an extensive validated E_C prediction model is essential to surmount the challenges in collecting real-world data. In this respect, this study develops and assesses an open-source low-resolution data-driven framework to estimate BEB's E_C in transit operation, using vehicular, operational, topological, and external parameters. Moreover, a three-step validation process is used to assess the proposed framework's performance. The results show that the prediction model provides a reasonable error margin (20%). The validation analyses show a powerful goodness-of-fit where the prediction model can explain more than 90% of the E_C variation. The framework can superbly provide transit planners and agencies with an optimal transit operating profile that reinforce the BEB's energy savings.

5.1. Introduction and Background

The accurate estimation of the energy consumption (E_C) rates of the Battery Electrical Buses (BEBs) might be equivocal since it relies on the circumstances surrounding the bus trips and the route where they operate. The uncertainties associated with BEB energy consumption are well documented in the literature and include speed profiles, driver behaviour, route topology, passenger-hour load profile, and weather conditions, among others (Abdelaty et al., 2021; El-Taweel et al., 2021; Hjelkrem et al., 2021; Rupp et al., 2020).

Furthermore, the inconsistencies in the energy consumption estimation process challenges research activities in the areas of planning the electrification of the transit network, as well as configuring the optimum BEB's transit operation profile that boosts the energy efficiency (Abdelaty and Mohamed, 2021a; El-Taweel et al., 2017). Claims for a proper E_C identification are frequently made to develop cost optimization, battery sizing, battery performance, charging station's spatial distribution, and transit-related greenhouse gases emissions (He et al., 2018; Kivekas et al., 2017; Qi et al., 2018a; Quarles et al., 2020; Teoh et al., 2018), all of which hinges on the energy consumption rates of BEBs. Therefore, there is no doubt that the implementation of BEBs in public transit systems requires quantifying and modelling the consumed energy for the bus operation (Abdelaty and Mohamed, 2021b; El-Taweel et al., 2020).

Two approaches have been proposed for modelling the E_C of BEBs (Mahmoud et al., 2016). The first approach employs real-world testing/demonstration to collect E_C data

from BEBs operation. Although it seems an attractive solution, it compromises external validity since the E_C behaviour is estimated within a restricted number of real-world tests, weakening the generality of the results owing to the limited number of unique trials (De Cauwer et al., 2017; Qi et al., 2018b). In contrast, the second approach depends on simulating the energy consumption performance of the BEBs using simulation models and according to the vehicular kinematic and dynamic terms that represent the bus motion (Hahn and Valentine, 2019; Rupp et al., 2019). In this case, the main limitation is the dependency on the vehicular parameters and the dynamic terms. That is, in turn, leading to a deficiency in studying the impact of the operational, topological, and external parameters on the consumed energy.

Even though both strategies are viable, their utility for transit operators and municipalities is restricted (Ferguson et al., 2019; Mohamed et al., 2018). Both approaches require vast technical expertise to implement, which is often not available for transit agencies. Consequently, transit services and agencies largely demand the development of data-driven models to predict BEB energy consumption while avoiding complex simulation models or restricted access to BEB real-world data. This arena is an ideal territory for data-driven prediction models.

Toward that end, our study aims at developing a data-driven framework to predict BEB energy consumption from open-source low-resolution data that are typically collected by all transit providers and municipalities. Although, there are several recent studies that have developed data-driven models to predict the energy consumption of BEBs, all of which are built on high-resolution microscopic data extracted from simulation models or

real-world data (Gao et al., 2017; Pamuła and Pamuła, 2020; Vepsäläinen et al., 2018b).

In essence, data-driven models produce E_C estimates based on substantial amounts of observed data (high or low resolution) considering representative parameters of the BEB transit operation such as operational, vehicular, topological, and external parameters.

First, operational parameters include parameters of the transit system, passenger-hour load profile, and driver behaviour. The transit system includes the number of stops along the bus route, the spacing between bus stops, and the BEBs' state of charge. The passenger-hour load profile represents the daily passenger loading along the bus trip, which can be collected through automatic passenger count (APC) gathered by the transit providers (Gallet et al., 2018; Kivekas et al., 2018). While, the driver behaviour is estimated from the average speed and acceleration/deceleration rates, which can be calculated from the speed profile data during the bus trips (Abdelaty and Mohamed, 2021b; Gao et al., 2017; Lajunen, 2014).

The key to operational data is collecting high-resolution speed profiles. Hardly may it be done since data depend on the circumstances of each bus trip, which involve variations due to traffic conditions and the drivers' behaviours. Besides, data collection requires experimental work using high-accuracy devices attached to the BEBs during the trips or synthetic speed profiles to create instantaneous speed cycle data that mimic the real-world BEBs' speed profiles.

Examples of high-resolution speed profile compilation for data-driven models can be found in the following cases. Firstly, Ma et al. (2021) generated stochastic synthetic

speed profiles using predefined data and information extracted from Google Maps. Observations include the length of the segments between the stops, the maximum and minimum speed for each segment, maximum acceleration/deceleration rates, maximum and minimum stop time at the traffic slights, and probability of making a stop during road intersections. A second example is the work of El-Taweel et al. (2021), who created probabilistic synthetic profiles using the typical characteristics of the speed profile, including the speed limit, trip distance, traffic condition, the distance between trip stops, dwelling time, upper/lower bounds of the speed, and acceleration/deceleration rates. Finally, Abdelaty and Mohamed (2021a) used a handheld GPS device connected to an android cell phone to collect realistic speed profiles data for a Canadian transit network. They collected the bus trip data during the daily peak hours and recorded more than 420,000 data points, which were analyzed to extract longitude/latitude, distance, and elevation for each data point. The acceleration/deceleration rates, average speed, and grade were also calculated between each pair of points.

Second, vehicular parameters include all the physical parameters related to the bus and the battery, such as (a) the bus mass, which represents the bus total weight , including the curb weight, motor, gearbox, wheels, battery, and passengers; (b) the bus's frontal area, which varies according to the BEBs type; (c) the drag coefficient and the rolling resistance coefficient, which differs following the road surface type and the weather; and (d) the battery-related parameters (Abdelaty and Mohamed, 2021b; Kivekäs et al., 2018).

Third, topological parameters include the route features such as the road gradient, which can be collected through the road topography platforms (e.g. google maps,

CanElevation, and GPS data) (Abdelaty et al., 2021; De Cauwer et al., 2017).

Lastly, external parameters involve the weather conditions and the auxiliary power. The weather condition represents the ambient temperature of the BEBs that is considered the essence of the air density parameter and the HVAC power. The HVAC power is a combination of heating, ventilation, and air conditioning powers that is predicted based on the ambient temperature, such as the model developed by (Tammi and Lajunen, 2016) and used by (Vepsäläinen et al., 2019). On the other hand, the auxiliary power implies the regenerative brake, which is used to determine the amount of recovered energy; and the auxiliary power that includes all other auxiliary components in the BEB, such as bus doors and power steering, and the hydraulic power for braking systems (Kunith et al., 2017; Vepsäläinen et al., 2019).

From a data-driven modelling perspective, these four sets of parameters are often used, separately or jointly, in three modelling techniques: Multiple Regression Analysis (MRA), Deep Learning Neural Network (DLN), and Machine Learning (ML) algorithms (Abdelaty et al., 2021; Gallet et al., 2018; Pamuła and Pamuła, 2020; Vepsäläinen et al., 2018b).

Regarding MRA, Teoh et al. (2018) conducted a multiple linear regression analysis to predict an E_C model from operational parameters, including passenger loading and the route length. The proposed model achieved a 0.85 coefficient of determination (R^2), which suggests high accuracy. However, the model was significantly influenced by external parameters, such as the charging type. Similarly, Vepsäläinen et al. (2018b) used a multiple linear regression analysis to define the association between the BEBs' E_C and the converter

power, power steering power, air compressor power, average speed, idle time, stop density, driver aggressiveness, battery temperature, and the battery state of charge. The predicted models explain 76% to 81% of the variation in the E_C with a high importance of the stop density and the driver aggressiveness on the consumed energy. Finally, (Abdelaty et al., 2021; Abdelaty and Mohamed, 2021b) developed an E_C prediction model using a multiple linear regression analysis using various operational, topological, and external parameters. The model shows the high impact of the road gradient, road condition, driver aggressiveness, stop density, HVAC power, the initial state of charge, passenger loading, average speed, and drag coefficient. The model shows high goodness-of-fit between the predicted and the simulated E_C with an R^2 of 0.9613.

A deep learning network (DLN) was used by Pamuła and Pamuła (2020) to estimate the E_C for BEBs based on route characteristics such as travel time, the distance between bus stops, weather, and elevation differences. Besides, they have developed an MLR technique to generate an E_C prediction model using the same parameters; then, they compared the results with the DLN's results. The estimated error for the DLN and MLR reached a maximum of 18% and 22%, respectively. In another research, (Abdelaty et al., 2021) developed an E_C prediction model using distinct neural network techniques (radial basis and multilayer perception) based on road gradient, road condition, driver aggressiveness, stop density, HVAC power, the initial state of charge, passenger loading, average speed, and drag coefficient. The model results explain more than 99.4% of the variation in the predicted E_C.

Finally, Machine Learning (ML) models in BEB's E_C application are scarce. Ma et

al. (2021) applied a gradient boosting decision tree to assess the impact of various parameters on the E_C of the diesel and electric buses. On their side, Abdelaty et al. (2021) used support vector machine learning and gradient boosting decision tree techniques to predict the BEBs' E_C, with a superior fit for each model ($R^2 = 0.970$ and 0.961, respectively).

5.1.1. Research Gap and Contributions

The aforementioned studies suggest some important factors affecting the E_C rates of BEBs. However, two fundamental gaps remain. Most notable is the lack of data-driven prediction models that combine the parameters concerning the speed profiles, route topology, passenger-hour load profile, and weather conditions through open-source low-resolution data to surmount the difficulties in collecting real-world data. Secondly, the lack of an extensive validation process to stabilize and further generalize the data-driven prediction models in the BEBs' energy consumption research. The latter is essential to overcoming the uncertainty in the operational and weather parameters.

Very few attempts have been made in the literature to address the first gap. Hjelkrem et al. (2021) used event-based low-resolution data for developing an energy consumption model using limited vehicular parameters, including bus mass, bus frontal area, rolling resistance, average speed and drag coefficient. Besides, they integrated the energy model with another model to calculating the auxiliary power from the HVAC and doors opening/closing power. The speed profiles data are generated using algorithms assuming that the driver reaches the average speed with constant acceleration/deceleration rates. The passenger loading is assumed as 75% of the bus capacity along the bus route.

Similarly, Fiori et al. (2021) estimated the energy consumption rates based on the traction power model that depends on bus speed, total mass, rolling resistance, drag coefficient, and frontal area. The auxiliary system model estimates the auxiliary power as a role of the ambient temperature and the battery state of charge. The CAN bus and GPS device were used to collect speed profiles, total bus mass, ambient temperature, and the battery state of charge.

Additionally, Li et al. (2021) predicted the energy consumption rates using kinematic and external parameters such as average speed, ambient temperature, departure time, and passenger loading. They generated stochastic synthetic speed profiles from various bus speeds, waiting time at the bus stops, the distance between stops, and acceleration/deceleration rates. They assumed the bus mass, bus frontal area, rolling resistance, and drag coefficient as constant values.

However, these studies focus more on using Newton's law of motion in estimating the E_C from the tractive force acting on the longitudinal dynamic movement of the bus up to wheels. That does not take into consideration the impact of the interaction that vehicular, operational, topological, and external parameters have on the E_C. Moreover, given the interlinked relationship between the topological parameters (e.g. road gradient) and the operational parameters of the BEB, it is necessary to account for this relationship in estimating an accurate E_C.

In this respect, the present study overcomes these gaps and develops and assesses an open-source low-resolution data-based framework to estimate BEB's E_C in transit operation. This framework integrates a data-driven prediction model using a MLR model

that accommodates vehicular, operational, topological, and external parameters. The framework relies on low-resolution open-source data collected by all transit providers. The model is applied in the Canadian context. The open-source include weather and topological data collected using environment Canada and CanElevation, as well as operational/topological data collected using the automatic vehicle location (AVL), automatic passenger count (APC), and topological parameters. Moreover, the model is validated using a three-step process to assess the proposed framework's performance in order to provide transit planners and agencies with the optimum operating profile for the BEB's transit systems that enriches the energy usage.

After this brief introduction, the study is organized as follows: the energy prediction framework is described in section 5.2, including the low-resolution data preparation and the data-driven modelling technique. Then, section 5.3 applies the developed model to the cases study and assesses the performance of the prediction model through a three-step validation process. That is followed by a discussion highlighting the practical implications of the proposed framework in section 5.4. Section 5.5 outline the conclusion of the study and discusses the key remarks.

5.2. Energy Prediction Framework

The energy consumption prediction framework consists of four main stages, as shown in Figure 5.1. The first stage involves preparing the low-resolution data from different sources that include AVL data, APC data, elevation data, and weather data. The second stage includes the experimental design that generates all possible BEB operation scenarios on the

selected routes, according to the selected vehicular, operational, topological, and external parameters. The third stage (i.e. modelling stage) includes calculating the E_C through a validated simulation model to mimic BEBs operation real-world operation (referred hereafter as the observed E_C dataset); and predicting the E_C using a validated data-driven model (referred hereafter as the predicted E_C dataset). The fourth stage evaluates the proposed framework through error estimation between the observed and predicted datasets, then using sensitivity analysis to identify the error variations in response to each E_C parameter.
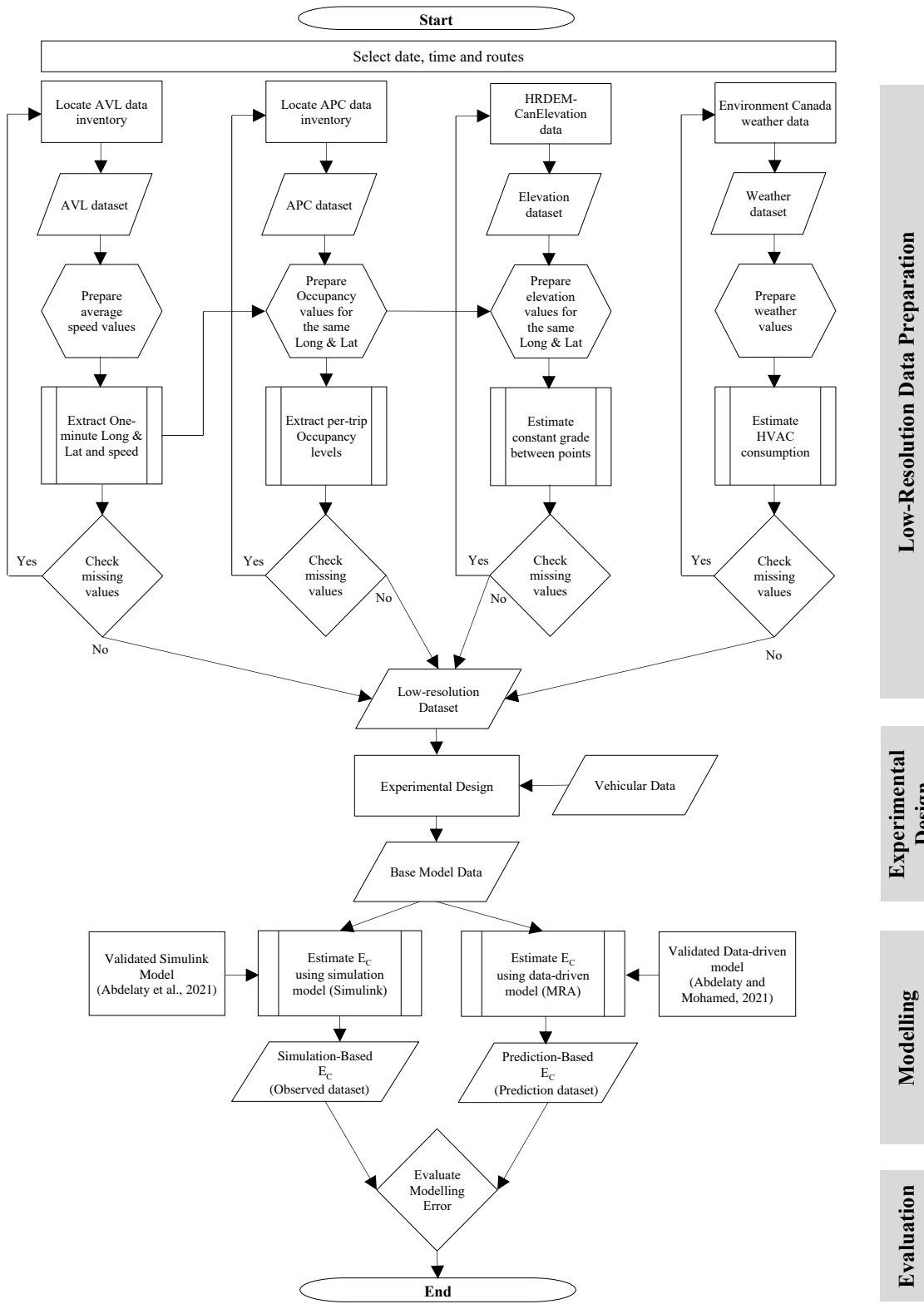
Figure 5-1. The proposed energy consumption prediction framework

5.2.1. Low-Resolution Data

The energy consumption rates of the BEBs varies significantly based on several parameters that can be classified into vehicular, operational, topological, and external parameters. Towards that end, a total of six bus routes located in the City of Hamilton (Ontario, Canada) have been selected (Table 5.1). The routes have different characteristics, topologies, and represents different transit operation including local routes, collector routes, and arterial routes, to accommodate all the bus transit operations in Canada (Nikel et al., 2020).

5.2.1.1. Speed Profiles

The speed profile data is collected twice (Figure 5.2). The first low-resolution speed dataset is collected through the Automatic Vehicle Location (AVL) data. AVL data is recorded using a GPS device mounted on the vehicle. The data is registered online every one minute with the longitude and latitude of the vehicle. Although the AVL could be treated as separate one-minute entries, we have applied various ad-hoc analyses to extract additional parameters such as the average acceleration, average deceleration, average speed, and the maximum speed, as shown in Table 5.1.

Table 5-1. AVL-based operational features for the observed speed profiles

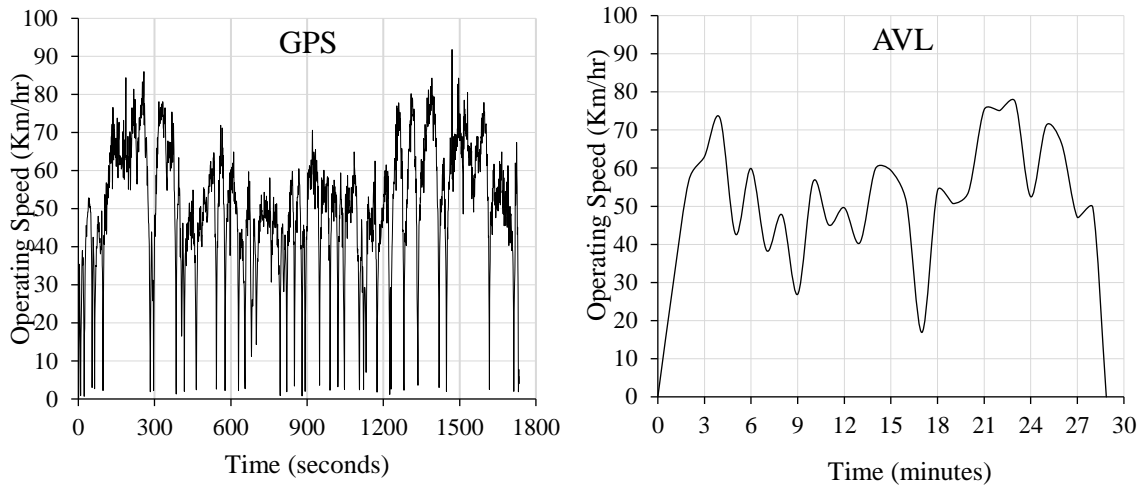| Bus Route	Route Type	# Of one-minute sections	Av. Acc. (m/s^2)	Av. Dec. (m/s^2)	Av. Speed (km/h)	Max Speed (km/h)
1	Local	8	2.55	-4.38	32.30	55.50
16	Local	17	1.46	-2.63	33.57	60.80
5	Collector	14	2.70	-4.77	37.10	74.80
23	Collector	7	1.49	-3.65	37.83	77.40
41	Arterial	10	1.81	-3.17	51.58	96.60
44	Arterial	29	2.22	-5.06	56.10	109.90
Min.	-	-	1.46	-5.06	37.10	64.80
Max.	-	-	2.70	-2.63	56.10	109.90
Mean	-	-	2.04	-3.94	42.58	80.33
St. D	-	-	0.53	0.95	7.12	16.17

Figure 5-2. The speed profiles for bus route 44, extracted from the GPS data and AVL data, respectively

While a second dataset (instantaneous speed profile) is collected using a GPS device (SX-Blue II) for the same bus routes during the same time periods. The SX-Blue II is a handheld tool connected to an android cell phone using Bluetooth and the SXblue toolbox application. We collected the data during the daily peak hours (7 am to 10 am and 1 pm to 4 pm). The device recorded more than 10,000 data points with an accuracy of 60 cm. The data points were analyzed using the GPS tool-track analyzer, which generates the longitude/latitude, trip distance, and elevation for each data point. Using these data points, we calculated the driver behaviour parameters (i.e. acceleration/deceleration rates and average speed) and the routeing topology (i.e. road gradient, stops/km, distance). The instantaneous dataset was used in the simulation model (detailed in section 5.2.3.) to simulate the BEB's E_C.

5.2.1.2. Automatic Passenger Count (APC) Data

The APC data is collected through sensors mounted at the front and rear doors of the bus.

Given that front doors are used for boarding and rear doors are used for alighting, bus occupancy could be easily estimated between any given stops. As such, we have analyzed the APC data for the selected routes during the same time periods to extract trip-level occupancy values. Figure 5.3 shows the bus stops' distribution along the selected bus routes.
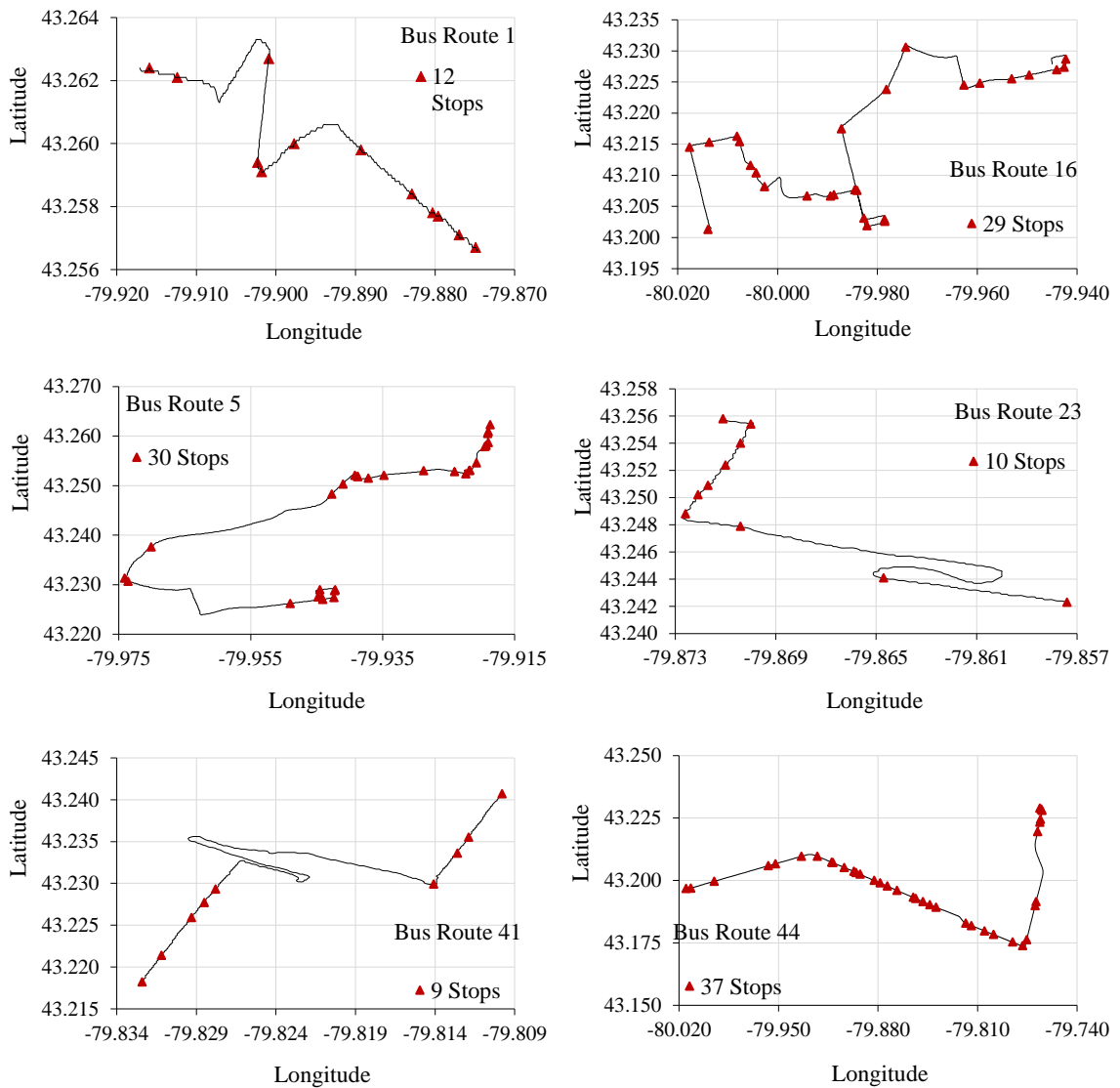

Figure 5-3. The layout of the selected bus routes and the bus stops' location for each route

The APC data for the selected routes at the daily peak hours are plotted in Figure

5.4. Each point in Figure 5.4 represents the number of passengers between every two stops for the six bus routes. The values vary between 0 passengers and 70 passengers. Therefore, we divided the number of passengers into eight levels: 0, 10, 20, 30, 40, 50, 60, and 70 passengers.
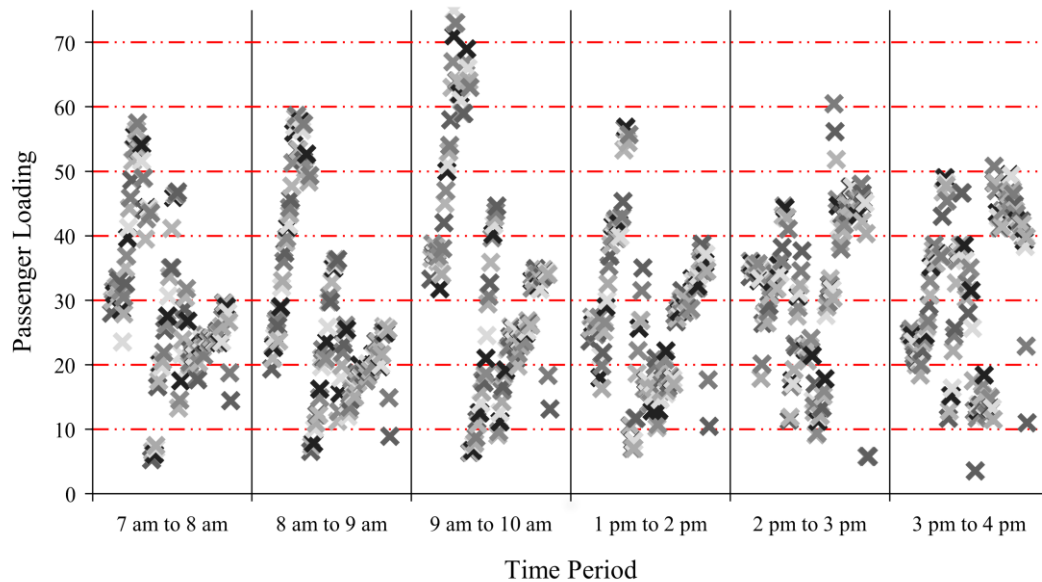

Figure 5-4. The passenger numbers between the stops on the selected bus routes at the peak hours

5.2.1.3. Topology Data

We imported the AVL-based coordinates into CanElevation to attain the elevations for each point on the selected bus routes. The elevations along the route length are plotted in Figure 5.5. Then, we used the elevation differences to calculate the road gradient. Due to the continuous changes in the road gradient along the bus route, we divide each bus route into sections, corresponding to the AVL data, where each section represents a constant road gradient. The average road gradient for bus routes 1, 16, 5, 23, 41, and 44 reached -2.11%,

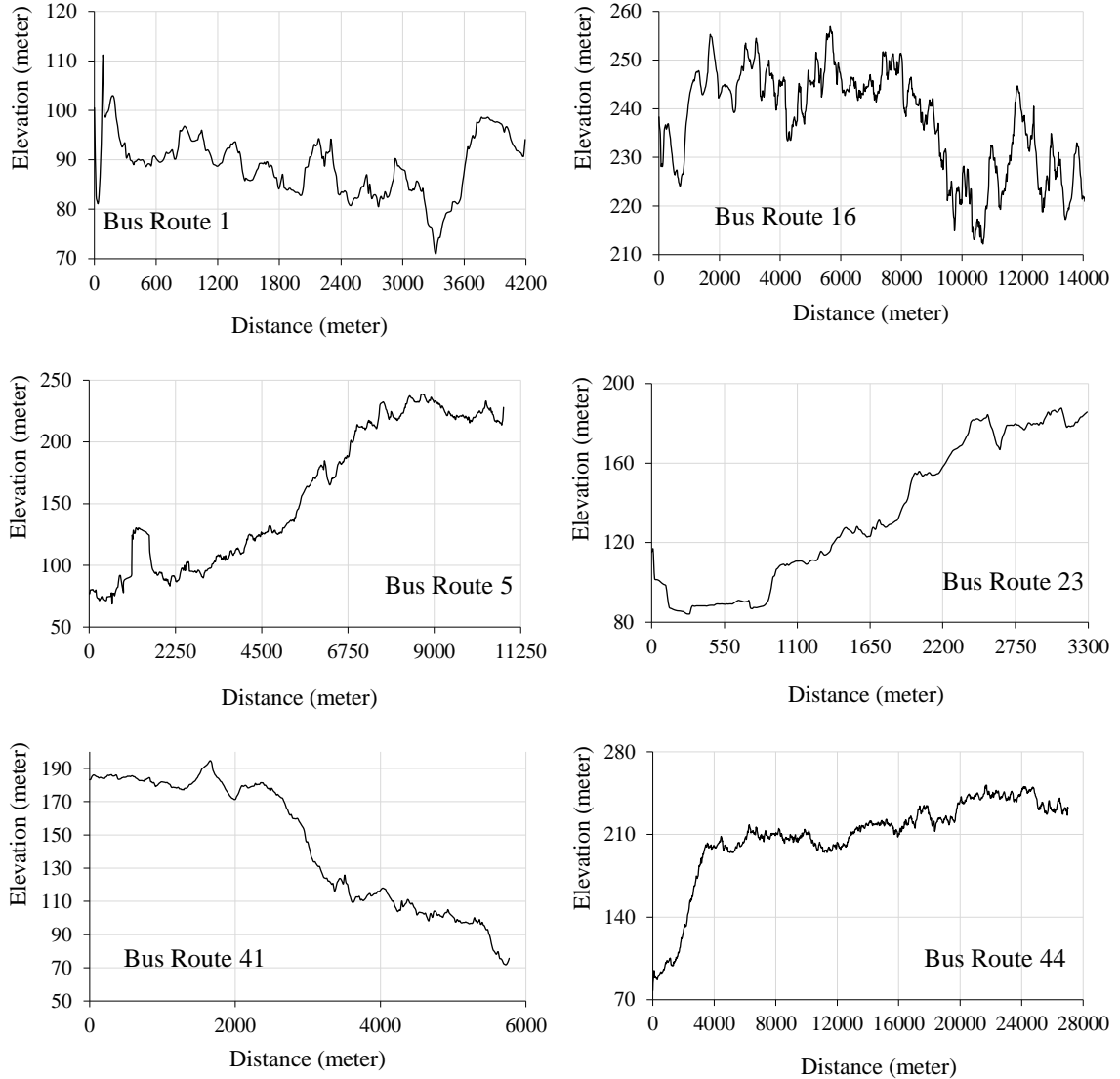4.44%, 0.29%, 1.19%, -0.52%, and -0.56%, respectively.

Figure 5-5. The route elevation via distance for each bus route

5.2.1.4. Weather Data

We collected historical weather data for the last 20 years in Ontario, Canada, as shown in

Figure 5.6 (Environment-Canada, 2021). It shows the maximum and minimum temperature

ranges for every year. The mean temperature ranges between 30 $^{\circ}$C and -20 $^{\circ}$C.

Furthermore, weather data during the analysis period was collected. However, the variation in the ambient temperature was marginal. Therefore, we divided the mean temperature degrees into six levels: -20 ℃, -10 ℃, 0 ℃, 10 ℃, 20 ℃, and 30 ℃.
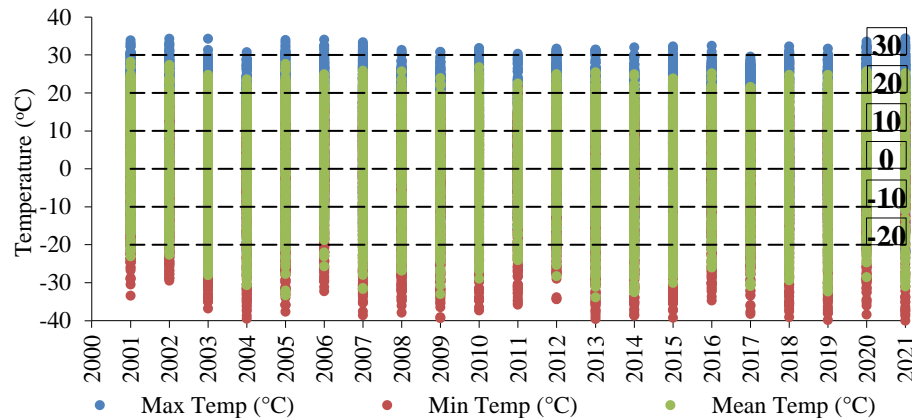

Figure 5-6. The historical weather data for the last 20 years in Ontario, Canada

5.2.2. Experimental Design

Due to the uncertainty in the vehicular, topological, and external parameters such as the road condition, SoC_i, passenger loading, and HVAC parameters, we designed a full-factorial experiment that contains all possible combinations between those parameters and the operational parameters that we extracted from the speed profiles. Although the open-source datasets reflect limited operational scenarios, we have developed a full-factorial experiment based on the unique combination of all variables extracted from the open-source datasets.

For the operational parameters and the speed profile data, we selected the total bus mass according to the curb weight and the passenger loading. The passenger loading is calculated based on the automatic passenger count (APC) collected by transit providers,

ranged 0-70 passengers during the trip. Thus, we divided the number of passengers into eight levels, with an average weight of 75 kg for each passenger (Lajunen et al., 2018).

Regarding the vehicular parameters, we included four levels for the initial state of charge (70% till 100%, with 10% intervals). The rolling resistance coefficient is considered according to the literature (three levels ranging from 0.006 to 0.02) (Gallet et al., 2018; Vepsäläinen et al., 2018a; Vepsäläinen et al., 2019).

The topological parameters (i.e. road gradient) are collected using the digital elevation model (HRDEM-CanElevation Series) (CanElevation, 2021), as explained in section 5.2.1.3. The external parameters included the HVAC power calculated based on the ambient temperature, following the model developed by (Lajunen and Tammi, 2016) and modified by (Vepsäläinen et al., 2019). We used the six temperature levels in section 5.2.1.4. to estimate the HVAC power.

As a result of the full-factorial experimental design, we obtained 1,008 single bus trips for each bus route, as shown in Figure 5.7. The selected single bus trips are employed in estimating the E_C rates using the simulation model and the data-driven model.
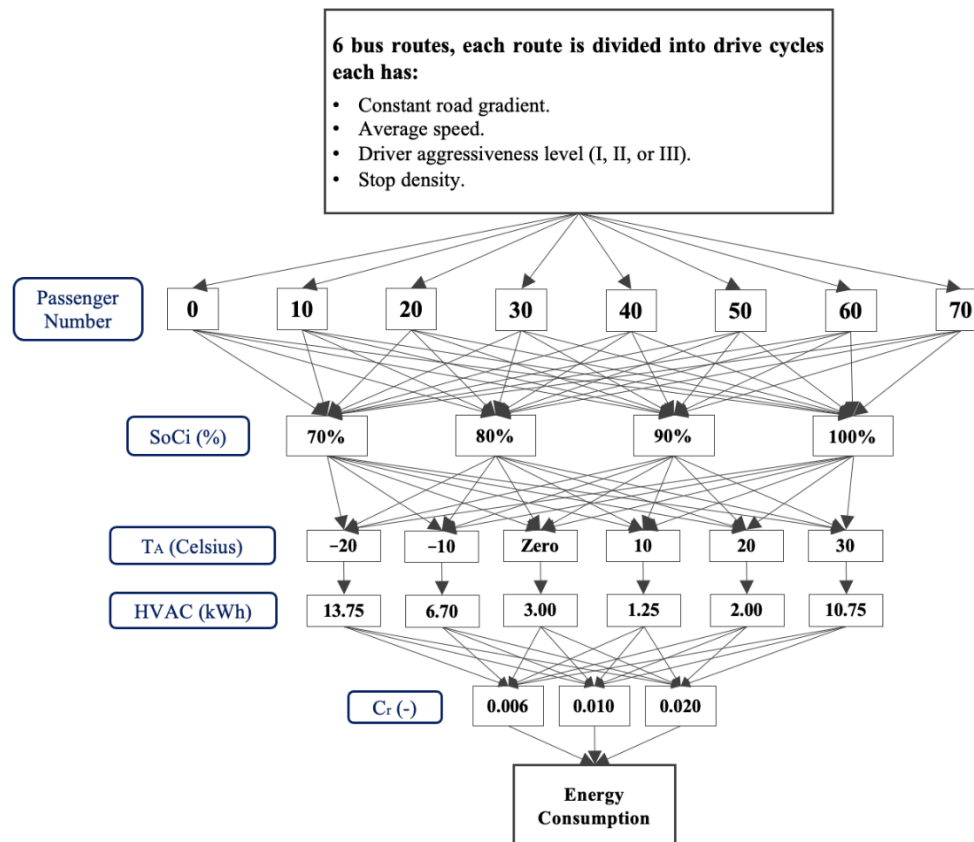
Figure 5-7. A flow chart showing the full-factorial experimental design

5.2.3. Validated E_C Simulation Model

The estimation of the E_C rates through field observation seems to be the most accurate approach since it is based on real-world interaction data. However, it is not bias-free unless there is an extensive dataset available. Suppose the external validity of the data is not guaranteed; in that case, its result can be highly subjective due to the possibility of human error and the hardness of controlling the impact of external factors (e.g. ambient temperature). The BEB literature has solved this challenge with the simulation-based approach since it allows the simulation of all the possible scenarios to obtain clear insights

into complex models as well as controlling the impact of external factors (Franca, 2015; Kivekäs et al., 2018; Vepsäläinen et al., 2018b).

A simulation model developed by (Abdelaty et al., 2021) is used to predict the E_C of BEB, using a MATLAB Simulink model and considering the block designs used in the advanced vehicle simulator (ADVISOR), as shown in Figure 5.8.
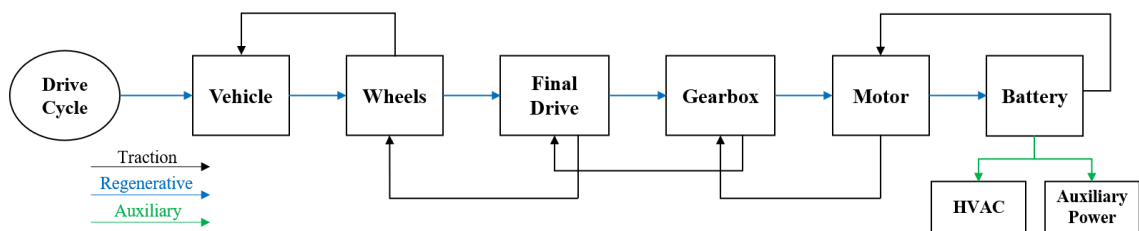
Figure 5-8. The model configuration of the BEB's powertrain (Abdelaty and Mohamed, 2021b)

The utilized simulation model is developed using the New Flyer XE40 electric bus features, which is a standard 40-foot city transit bus. The main input parameters for the model, including the vehicular and operational parameters, are listed in Table 5.2.

Table 5-2. The main parameters' values for the utilized BEB in this study (Baseline Model)

Parameter	Value
Curb weight (kg)	14932
Bus frontal area (m^2)	8.32
Maximum torque (N.m.)	2500
The tire's dynamic radius (m)	0.5
Drag coefficient	0.6
Air density (kg/m^3)	1.27
Auxiliary power (kWh)	7
Battery capacity (kWh)	200

Altoona test results for New Flyer XE40 were used to validate the simulation model (Altoona, 2015). The Altoona test involves three driving cycles, comprising the Arterial

(ART) cycle, Central Business District (CBD) cycle, and Commuter (COM) cycle. These driving cycles are used in the simulation model to calculate the E_C, and then we compared the simulated results with the Altoona results. The results demonstrate that the used simulation model calculated the E_C within a 5% error (0.05 kWh/km), as shown in Figure 5.9. Thus, with confidence, we can claim that this simulation model for BEB energy usage is highly accurate.
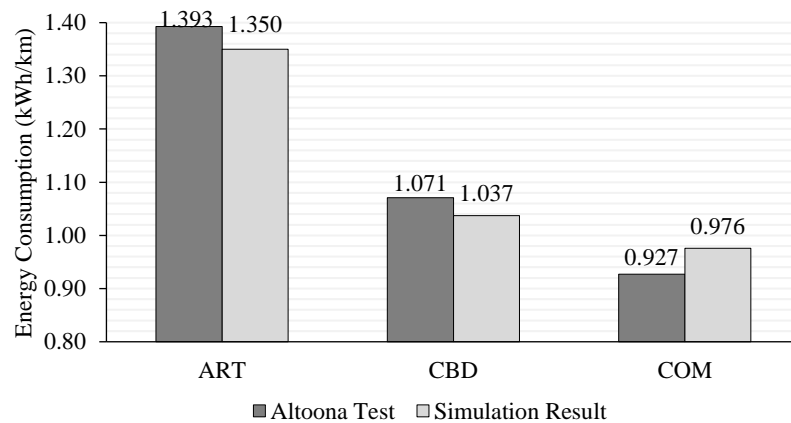
Figure 5-9. Comparison between the E_C results for the Altoona test and our simulation model (Abdelaty et al., 2021)

5.2.4. Data-Driven Prediction Model

A data-driven model was developed to predict the BEB's E_C using Multiple Linear Regression Analysis (MLR). This model was developed by (Abdelaty et al., 2021) using a training dataset of 907,199 operation scenarios and its accuracy was evaluated using a testing dataset of 169,344 operation scenarios.

The model's goodness of fit (Eq. 5.1) is evaluated using the coefficient of determination (R^2), and root mean squared error (RMSE). The selected model demonstrates

a vigorous relationship between the BEB's E_C and the input parameters, with an R^2 of 0.961. The testing dataset implying that the prediction model explains about 94% of the variance occurring in the predicted E_C.

$$E_C = -0.885 + 0.380g + 0.012SoC_i + 0.260R_C + 0.036HVAC + 0.005P_L +$$

$$0.065D_{Agg.} + 0.128S_D + 0.007V_a + 0.173C_D \qquad \text{Eq. 5.1}$$

The developed model predicts the E_C rates based on nine vehicular, operational, topological, and external parameters. These parameters include:

1) Road gradient (g) represents the constant upgrade and downgrade of the road segment. In this study, we used the Digital Elevation Model (HRDEM) which is a part of the CanElevation Series (CanElevation, 2021) open-source data.

2) The initial state of charge (SoC_i) is the battery state of charge in percentages at the beginning of each trip.

3) Road condition (R_C) includes three levels I, II, and III, determined by a constant bus tyre pressure and a rolling resistance coefficient, whereas condition Level I refers to good dry road condition that represents the rolling resistance coefficients ≤ 0.006. Level II refers to a fair wet road condition for rolling resistance coefficients ≥ 0.01 and > 0.02, while level III refers to a poor icy road condition (slush) that includes rolling resistance coefficients ≥ 0.02.

4) HVAC (kWh) can be accurately estimated using the ambient temperature data collected from open-source weather data (Environment-Canada, 2021) and following the model developed by (Lajunen and Tammi, 2016) and modified by

(Vepsäläinen et al., 2019).

5) Passenger loading (P_L) represents the passenger-hour load profile during the bus trip from the automatic passenger count (APC) data.

6) Driver aggressiveness (D_{Agg}) is derived from the average speed and the acceleration/deceleration rates of the bus drivers. It is split into three levels; Level I represents the slow driving behaviour with acceleration rates < 0.5 m/s^2 to and deceleration rates < 1.5 m/s^2, level II represents the normal driving behaviour with acceleration rates from 1 m/s^2 to 1.5 m/s^2 and deceleration rates from 2 m/s^2 and 2.5 m/s^2. Level III represents the aggressive driving behaviour with acceleration rates > 1.5 m/s^2 and deceleration rates > 3 m/s^2. It can be drawn from the speed profiles data collected through the automatic vehicle location (AVL) data or the GPS devices.

7) Stop density (S_D) represents the number of stops per kilometre along the bus trip. It is calculated from the AVL data and validated through the General Transit Feed Specification (GTFS)

8) Average speed (V_a) represents the BEB's average speed (km/h) during the trip. It is estimated from the automatic vehicle location (AVL) data.

9) Drag coefficient (C_D) contains three levels selected based on the previous studies and ranges from 0.5 to 0.7 (Gallet et al., 2018; Lajunen, 2018; Vepsäläinen et al., 2019).

5.3. Prediction Model Performance

The low-resolution open-source-based data-driven prediction model performance has been assessed through a three-step validation process. This includes: 1) analyzing the predicted EC for each single bus trip for the one-minute drive cycles; 2) examining the model validity on full bus trips; and 3) analyzing the estimated error for each parameter.

5.3.1. One-Minute Single Trip Validation

For the first dataset collected using the open-source data, the E_C is predicted using the data-driven prediction model for each one-minute single trip on the selected routes (n = 53,424). Each trip represents different average speeds, stop density, road gradient, ambient temperature, the initial state of charge, passenger loading, and road condition. These are compared to the observed E_C to assess the performance of the prediction model, as shown in Figure 5.10.
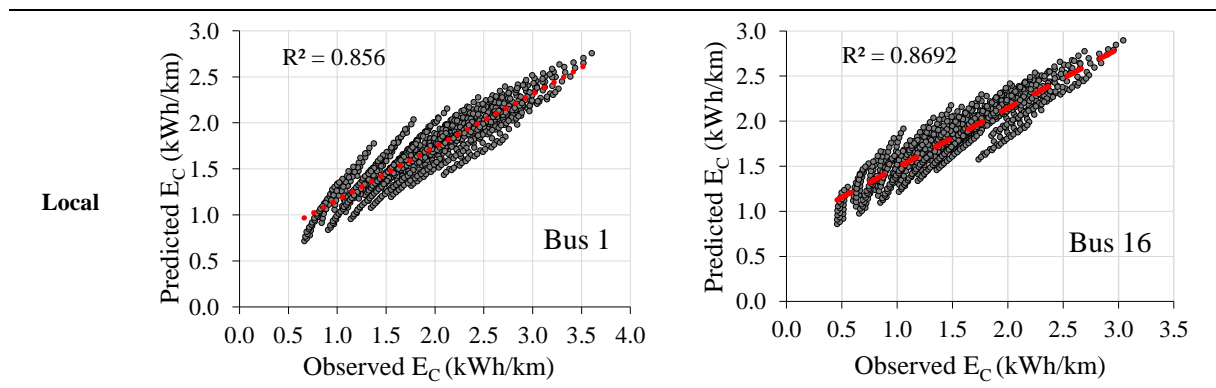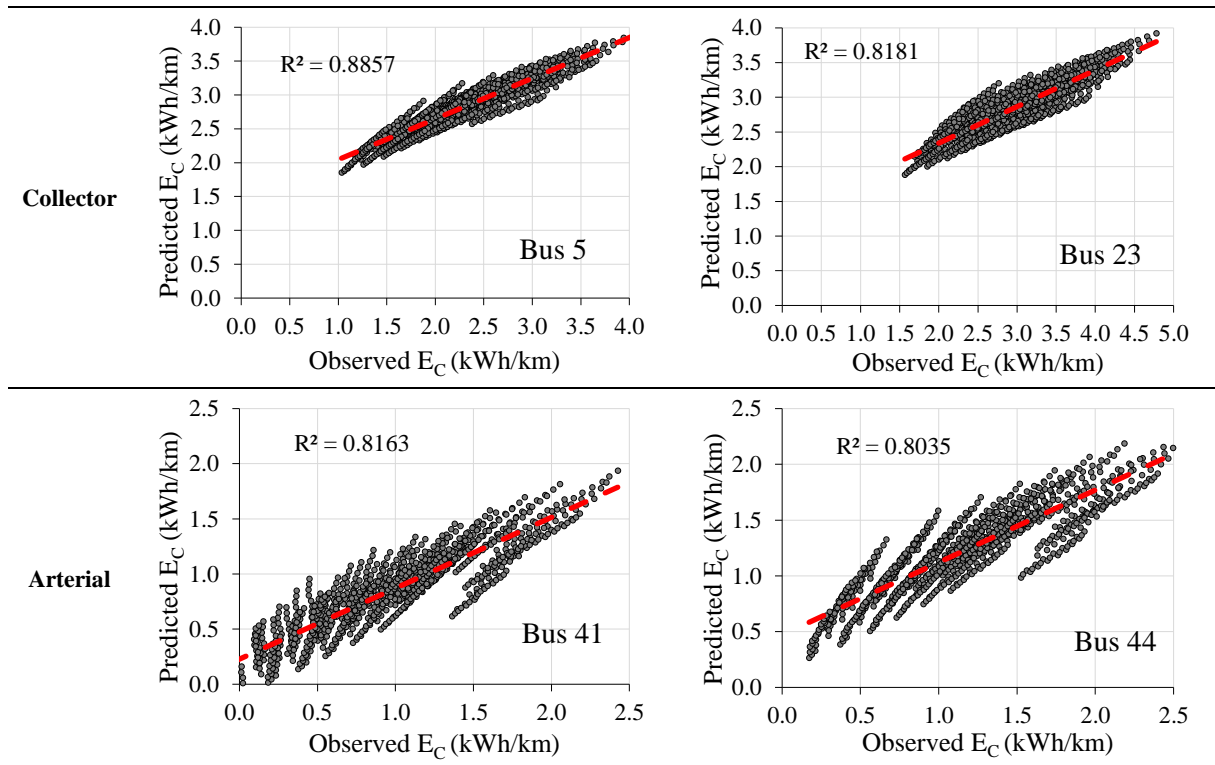
Figure 5-10. Scatterplot of the predicted E_C per one-minute trip compared to the observed E_C for each bus route

The goodness-of-fit is assessed using the coefficient of determination (R^2). As shown in Table 5.3, the results of R^2 across all routes ranged between 0.8035 and 0.8857, which expresses a high accuracy of the prediction model. Besides, the error is estimated for the prediction model through different measures, including Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), and Mean Absolute Percentage Error (MAPE). The MAD values ranged from 0.0965 to 0.37336, which indicates the low spread of the data. That shows the convergence between the predicted E_C and the observed E_C. The MSE (0.0500 - 0.2273) and RMSE (0.2237- 0.4768) demonstrates the low distances between the E_C data points and the regression line of the prediction model. These findings, in turn, suggest the high performance of the prediction model. The MAPE

results show that the error percentage differed between 10 % and 18% for all route types used in our analysis, confirming the good fitting of the prediction model.

Table 5-3. Comparison of the prediction model performance on different route types

Route Type	Bus Route	R^2	MSE	RMSE	MAD	MAPE
Local	Bus 1	0.8560	0.22730	0.47676	0.30128	18.39596
	Bus 16	0.8692	0.05283	0.22985	0.15718	12.27805
Collector	Bus 5	0.8857	0.16088	0.40110	0.34670	14.84315
	Bus 23	0.8181	0.18877	0.43448	0.37364	10.54224
Arterial	Bus 41	0.8163	0.08521	0.24769	0.20530	15.24210
	Bus 44	0.8035	0.05004	0.22371	0.09652	10.04563

5.3.2. Trip-Level Validation (Per Parameter)

This part validates the prediction model's performance on full bus trips for the selected bus routes (n = 6,048). This validation is using the observed dataset that represents instantaneous speed profiles collected using GPS device. Figure 5.11 shows the relation between the predicted and the observed E_C, where each dot represents a full bus trip, and the colours represent the different values of the parameters mentioned above.
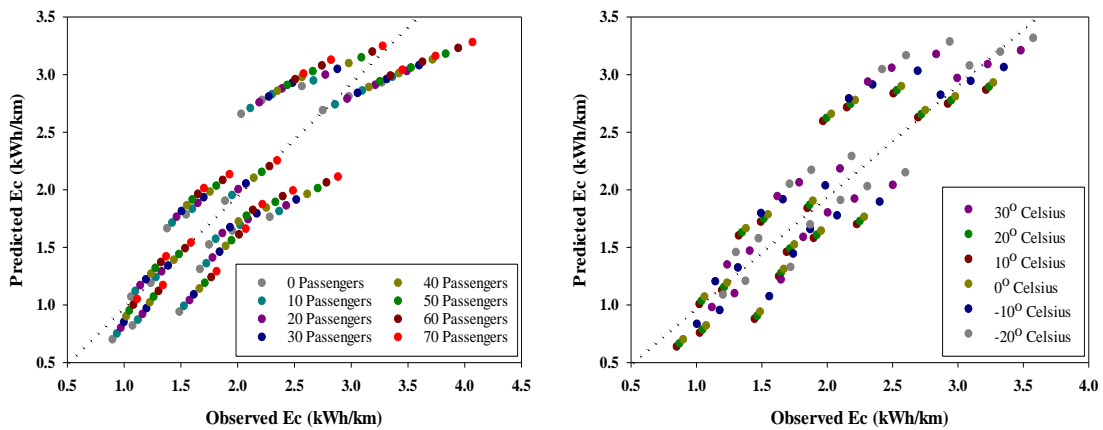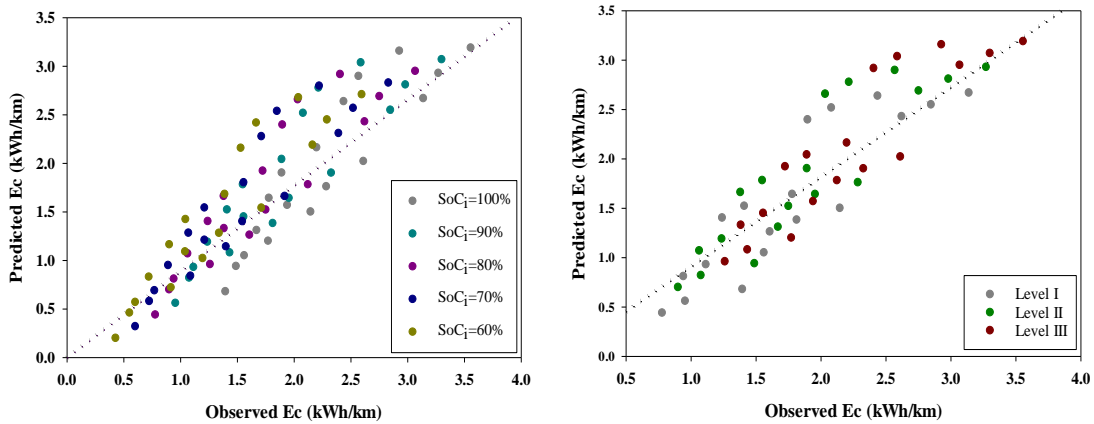
Figure 5-11. A parameter-level comparison between the observed E_C and the E_C predicted from a low-resolution data

The distances between the regression line and E_C rates of the bus trips are low for passengers' occupancy between 0 and 70. The R^2 and the standard error of the estimate (SEE) are 0.9699 and 0.3444, respectively. That indicates that the prediction model fits the E_C resulted from low-resolution data of passenger loading. The model also shows high accuracy in predicting the E_C from different temperatures (-20 ºC, -10 ºC, 0 ºC, 10 ºC, 20 ºC, and 30 ºC), with $R^2 = 0.9672$ and SEE = 0.3489. Moreover, the prediction model explains more than 97% of the variation occurs in the E_C predicted from the initial state of charge ($R^2 = 0.9719$ and SEE = 0.3528). The validation process yielded a very accurate prediction with an R^2 of 0.9582 and a SEE value of 0.3570 for all the E_C rates resulted from different levels of the road condition, including Levels I, II, and III.

Furthermore, the predicted E_C rates are compared to the observed E_C to assess the prediction model's performance for the full bus trips on the selected bus routes, using the validation dataset, as shown in Figure 5.12. The goodness-of-fit is assessed using the coefficient of determination (R^2) that ranged between 0.8158 and 0.9491.

Figure 5-12. Scatterplot of the predicted E_C per full trip compared to the observed E_C, for each bus route
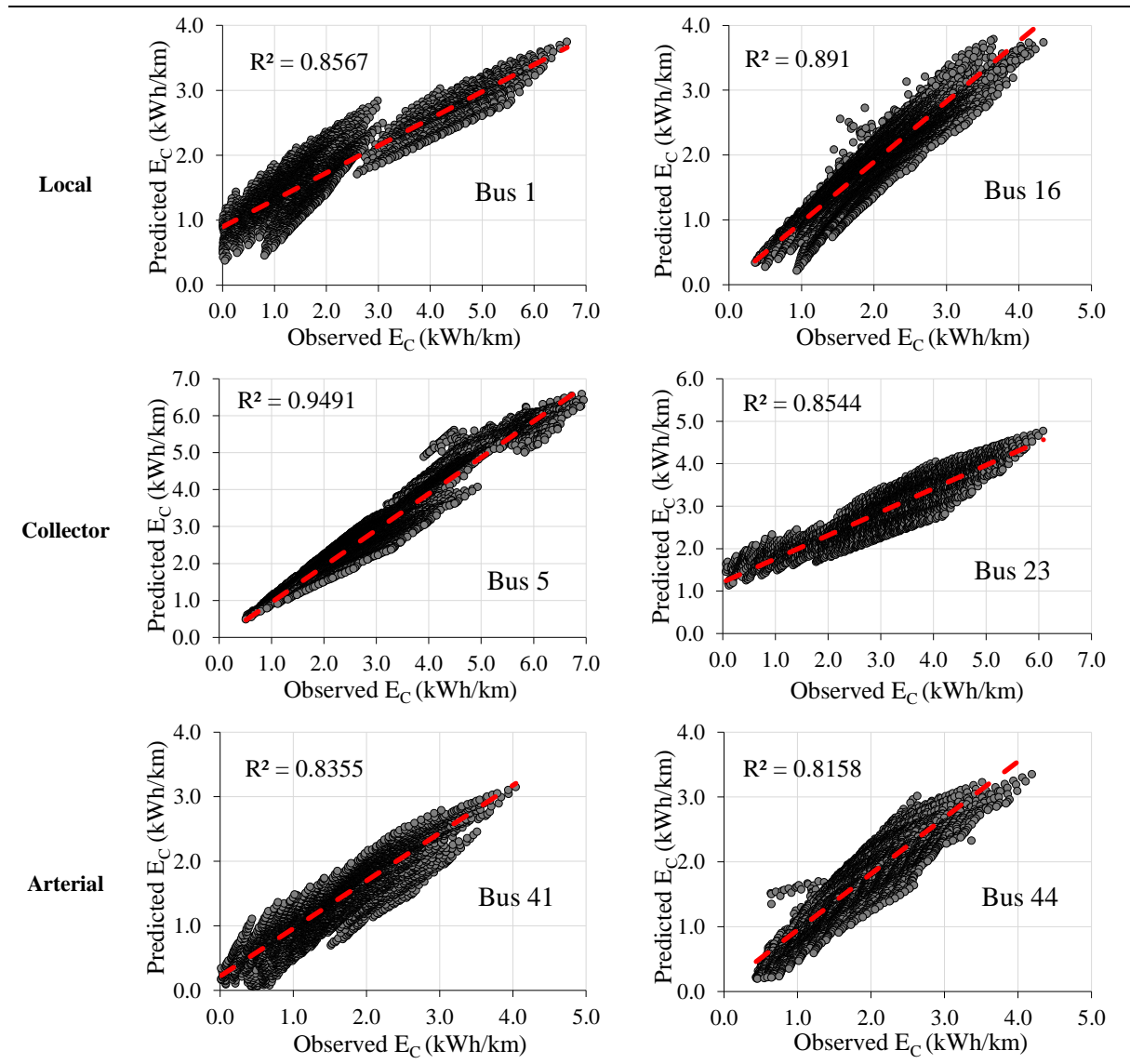
5.3.3. Error Analysis

The residuals between the predicted and observed E_C for the bus trips are used to estimate the error percentage for the local, collector, and arterial routes. The histograms resemble a normal distribution with different forms for each route type, as shown in Figure 5.13.

The estimated error of the predicted E_C on the local routes shows a double-peaked distribution with errors from -5% to -1% and from 12% to 17%. The estimated error on the collector routes shows a right-skewed distribution, indicating that the data-driven prediction model overestimates the E_C rates with 5% to 16% of the observed E_C. For the arterial routes, the error distribution shows many peaks close together ranging from -2% to -18%, where the top of the distribution resembles a plateau.
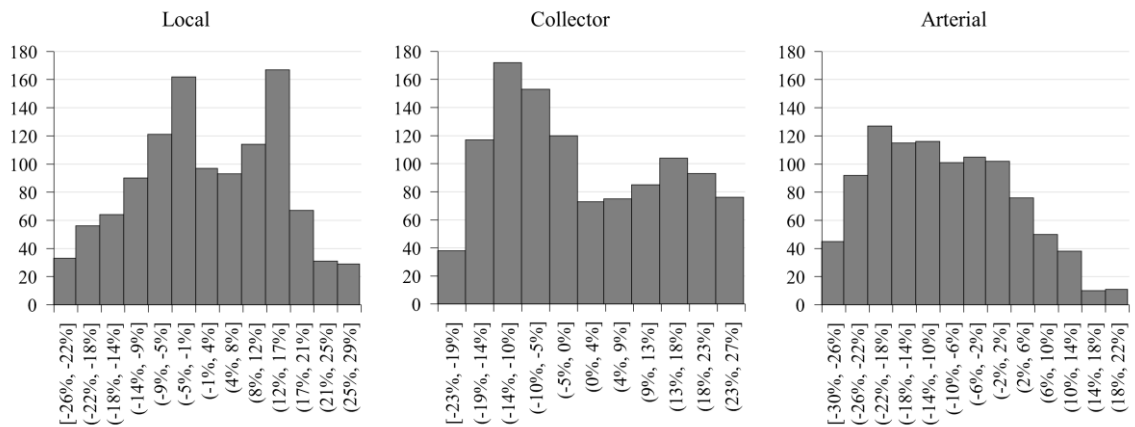
Figure 5-13. Histograms of the estimated errors on the local, collector, and arterial routes

The estimated error of all bus trips is categorized by parameter values using four low-resolution data parameters. The parameters include the initial state of charge, passenger loading, road condition, and weather condition. Then, these errors are plotted using boxplots (Figure 5.14) to show the spread and centers of the error for all route types, as well as indicating the interquartile range where the bulk of the values lie.

The arterial routes have the highest error range among all route types, where most errors are negative errors, indicating that the data-driven model overestimates the predicted E_C. In contrast, the performance of the data-driven model on the local and collector routes

tends to predict E_C equal or lower than the observed values. However, the interquartile ranges of the maximum estimated error are between +21% and -21% for all route types.

Figure 5.14.a. displays the impact of the initial state of charge on the performance of the data-driven energy prediction model. The interquartile range for the estimated error is smaller for the bus trips operating with a full SoC_i (100%). The error ranges increase gradually as the SoC_i decrease until they reach the highest ranges at $SoC_i = 70\%$.

Figure 5.14.b. shows that the interquartile range of the estimated error is similar for all the passenger loading levels. The data-driven model's error on the local route is within -12% and 10% for all passenger loadings between zero and 70 passengers. While for the collector routes, the error differs from -10% to 12% for all the passenger loading levels. The predicted E_C of the bus trips on the arterial routes shows errors within 5% and -21%.

Figure 5.14.c. presents the similarity of the estimated error behaviour among all the levels of the road condition where the interquartile range is between -12% and 15% for the bus trips on local and collector route types. While the bus trips on the arterial route have higher negative error ranges that reach -19%.

Figure 5.14.d. shows that the interquartile range of the estimated error is varying according to the weather temperature. The temperature degree peaks (-20° and 30°) hold the lowest error percentage (±12%) for all route types. The error boundary is getting higher for those bus trips that are operating in weather temperatures closer to zero degrees.

a) Error variation due to SoC$_i$ levels

b) Error variation due to passenger loading levels

c) Error variation due to road condition levels

d) Error variation due to weather temperature levels
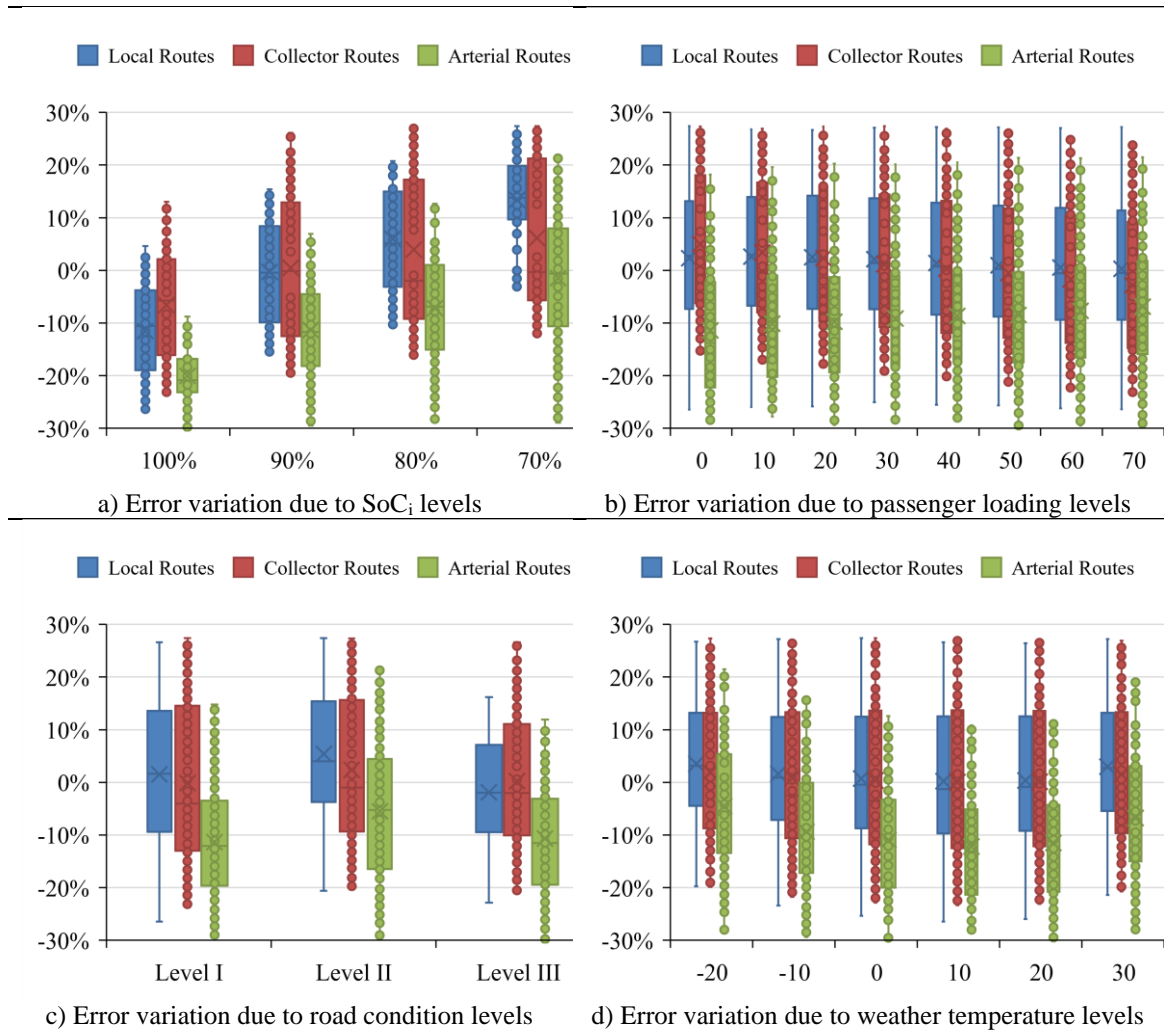
Figure 5-14. Error variation for all route types according to the low-resolution data parameters

5.4. Discussion

5.4.1. Discussion of the Results

The model's performance analysis indicates superior fitting of the E$_C$ results derived from the data-driven prediction model (predicted E$_C$) and the validated simulation model (observed E$_C$). A one-minute single trip validation resulted in a promising goodness-of-fit. The R^2 (up to 0.8857) indicates that the prediction model explains the variation in the E$_C$

rates on the selected bus routes. Further, the lowest R^2 value for the one-minute segments is 0.8035. The results from MAPE reveal that the error percentage for all route types included in our assessment was between 10% and 18%, which proves that the prediction model was well fitted.

Another validation is performed for the full bus trips. This considers the impact of the parameters derived from the low-resolution data, including weather condition, road condition, passenger loading, and the initial state of charge. The collected passenger loading data (0 passengers to 70 passengers) and the weather condition data (-20 ^{o}C to 30 ^{o}C), road condition (level I to level III) yielded a very accurate prediction with an R^2 of 0.9699, 0.9672, and 0.9582, respectively. In addition, the initial state of charge resulted in a superior goodness-of-fit with an R^2 of 0.9719.

We analyzed the error estimated from the residuals between the predicted and the observed E_C. The interquartile range of the estimated error varying between +21% and -21% for all bus trips. Among all route types, arterial routes have the greatest error range, which indicates that the data-driven model has an overprediction of E_C. Conversely, on local and collector routes, the data-driven model's performance tends to anticipate E_C values equal or lower than those observed.

In short, the validation done on the data-driven prediction model suggests that the model can predict the BEBs' energy consumption with high accuracy. The computed sum of squared residuals, the errors, and the coefficient of determination indicate that the model has superior goodness-of-fit when we used low-resolution data for the passenger loading, road condition, weather condition, road topology, and low-resolution speed profiles.

However, error values between predicted and observed E_C can be explained by issues with the collected data, specifically the initial state of charge. That is because the data-driven prediction model does not consider the regenerative brake energy in predicting the E_C, while the validated simulation model does. To comprehend that point, we made a sensitivity analysis for the error to understand how the error varies with the parameters used in the prediction model.

5.4.2. Sensitivity Analysis

A sensitivity analysis is conducted to ascertain how the prediction model and the errors are affected based on the changes in the utilized parameters. A Sobol index, a form of global sensitivity, is used to determine how the model output and the errors depend upon each input parameter (first-order effect) or upon the interaction between the parameters (total-order effect).

5.4.2.1. Prediction Model's Sensitivity

The road gradient parameter has the greatest significance on all the generated models for both first-order and total-order effects. This is presented in Table 5.4. The initial state of charge has the second-highest influence on the BEB's E_C. In contrast, driver aggressiveness has the lowest effect on the BEB's E_C. The order effect of the other parameters are similar in both predicted and observed E_C data where the HVAC (weather condition), road condition, passenger loading, stop density, and average speed are affecting the energy consumption rate in the same order. These findings confirm and add to the results of (Abdelaty et al., 2021; Pardo, 2017; Vepsäläinen et al., 2019; Vepsäläinen et al., 2018b).

Table 5-4. The order effects of the input parameters for both the data-driven prediction model (predicted E_C) and the validated simulation model (observed E_C)

Order-effect	Predicted E_C		Observed E_C	
	First-order effect	Total-order effect	First-order effect	Total-order effect
g	0.9428	0.9455	0.8995	0.9035
SoC$_i$	0.0204	0.0237	0.0284	0.0341
HVAC	0.0096	0.0092	0.0161	0.0198
R$_C$	0.0094	0.0125	0.0151	0.0167
P$_L$	0.0068	0.0064	0.0132	0.0151
S$_D$	0.0018	0.0025	0.0087	0.0104
V$_a$	0.0018	0.0020	0.0075	0.0088
D$_{Agg}$	0.0005	0.0008	0.0098	0.0102

5.4.2.2. Error Sensitivity

For the estimated error, the Sobol indices values for the utilized parameters show that the initial state of charge has the highest impact on the error values (see Figure 5.15). The remaining parameters have small impacts (first-order and total-order effects) on the estimated errors where the stop density, road condition, and road gradient have the second, third, and fourth effects on the error.
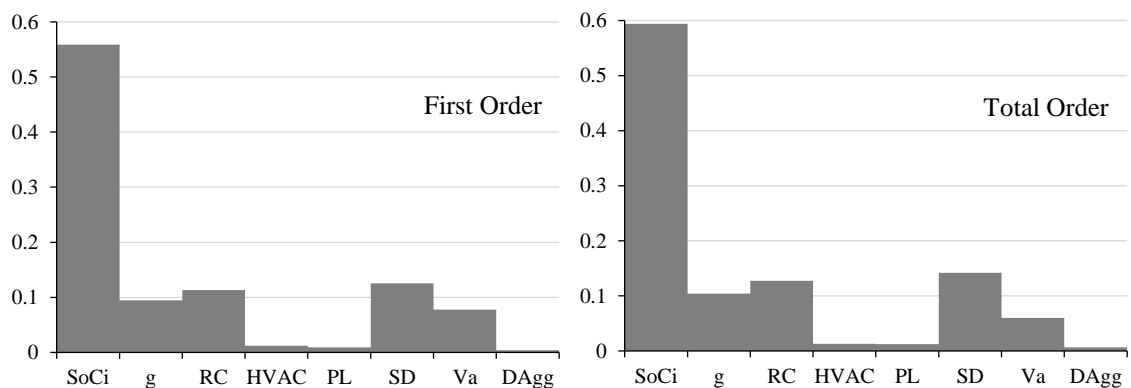

Figure 5-15. The Sobol indices of first-order and total-order effects for the estimated error

The high impact of the initial state of charge on the estimated error is because the validated simulation model counts the regenerative energy in calculating the E_C rates. In

contrast, the used data-driven prediction model does not consider regenerative energy in predicting the E_C rates. This is the leading cause of the deviation between the observed E_C and the predicted E_C. The error is estimated by subtracting the predicted E_C (calculated using the proposed data-driven prediction model) from the observed E_C (calculated using the validated simulation model). In most bus trips on the local, collector, and arterial routes, the error is negative, which means that the prediction model gives higher E_C rates than the observed E_C since it does not add the regenerative energy to the predicted energy.

On the other hand, the lowest range of the estimated error occur when the SoC_i is 100%, as shown in Figure 5.14.a. The reason is that while the $SoC_i = 100\%$, the regenerative energy that goes back to the battery is very low, which results in higher observed E_C rates compared to the prediction model. As long as the SoC_i is getting lower till the 70%, the regenerative energy is getting higher, reflecting on the observed E_C values, which in turn increases the error range of the prediction model. This affirms the findings of (Vepsäläinen et al., 2019) on the non-linear impact of SoC on E_C consumption.

Overall, these results indicate the potential of predicting the BEB's E_C based on a data-driven prediction model coupled with low-resolution open-source transit data with a reasonable error margin. The maximum error estimated in our model is within ±21%, which is the highest level of error in the one-minute segments across all route types.

5.5. Conclusion

This paper develops and assesses an open-source low-resolution data-based framework to estimate BEB's E_C. The framework integrates a data-driven prediction model developed

using the MLR technique based on various operational, topological, external, and vehicular parameters, and it relies on low-resolution open-source data collected by transit providers. The data is collected for six bus routes representing local, collector, and arterial routes to accommodate all the bus transit operations in Canada.

A low-resolution speed dataset is collected through the Automatic Vehicle Location (AVL) data, and another dataset (instantaneous speed profile) is collected using a GPS device for the same bus routes during the same time periods. Besides, the automatic passenger count (APC) is used to estimate the passenger loading throughout the bus trips. The topological parameters (e.g. road gradient) are collected from CanElevation, while external parameters (e.g. weather conditions) are collected from Environment-Canada.

On the other hand, a full-factorial experimental design is performed to generate all possible operation scenarios based on all parameters. As a result, we obtained about 1,008 single bus trips for each bus route with a total of 6,048 single trips. The energy consumption rates are calculated for these trips using a validated simulation model to generate the observed E_C. In addition, the E_C rates are predicted using the proposed data-driven prediction model, then the error is estimated from the deviation between the predicted and the observed E_C datasets.

A three-step validation process is used to check the performance of the proposed prediction model, including a) a one-minute single trip validation where the speed profiles of the selected routes were divided into one-minute drive cycles resulted in 53,424 single bus segments; b) a full bus trip validation based on the selected parameters; and c) a sensitivity analysis of the estimated error.

The validation findings show a powerful goodness-of-fit where the prediction model can explain more than 80.35% (the lowest value for the one-minute segments) from the variation that occurs in the E_C rates on the selected bus routes. The MAPE results show that the error percentage for all route types is from 10% to 18%, which proves the high accuracy of the proposed prediction model. Furthermore, the collected passenger loading data (0 passengers to 70 passengers) and the weather condition data (-20 ºC to 30 ºC), road condition (level I to level III) yielded a very accurate prediction with an R^2 of 0.9699, 0.9672, and 0.9582, respectively. At the same time, the initial state of charge contributes to a superior goodness-of-fit with an R^2 of 0.9719. In addition, errors estimated from the residuals between the predicted and the observed E_C indicate that the interquartile error range is varying between +21% and -21% for all bus trips.

The sensitivity analysis reveals that the initial state of charge has the highest impact on the estimated error. The error range is the lowest at SoC_i is 100%, and this error range increases with the decrease in the SoC_i. Therefore, future enhancement of the data-driven model should focus on accommodating regenerative brake variable as a predictor in the model. Despite neglecting the regenerative energy in the data-driven prediction model, the maximum error in the predicted E_C for most bus trips is within ±15% on all route types.

Compared to recent studies in the field, the work of (Hjelkrem et al., 2021) relied on several data fusion steps, hindering the automation of their proposed model. Further, their E_C model was not validated against real-world data. While the work of (Fiori et al., 2021) relied on high-resolution GPS-based speed data. Furthermore, the study by (Li et al., 2021) generates a stochastic speed profile in estimating the E_C of BEBs. These studies have

paved the way for the utilization of data-driven models in BEB planning. However, they fall short in providing transit agencies with an easy-to-implement open-source-based model that is compatible with transit data architecture and could be readily implemented in their daily operation. As such, the present study complements the current literature and provides sound contributions.

Overall, the proposed data-driven model can inform the optimum operation profile design that enhances the energy utilization in the BEB's transit systems. It also offers significant contributions to estimating the BEBs' energy consumption without relying on highly sophisticated simulation models or the need to collect real-world data. Our study clearly demonstrated that different route types exhibit varied energy consumption rates, which must be considered for successful BEB planning.

5.6. Acknowledgement

5.7. References

Abdelaty, H., Al-Obaidi, A., Mohamed, M., Farag, H.E.Z., 2021. Machine learning prediction models for battery-electric bus energy consumption in transit. Transportation Research Part D: Transport and Environment 96.

Abdelaty, H., Mohamed, M., 2021a. Energy Consumption Uncertainty Model For Battery-Electric Buses in Transit, 2021 IEEE Transportation Electrification Conference &

Expo (ITEC), pp. 1-5

Abdelaty, H., Mohamed, M., 2021b. A Prediction Model for Battery Electric Bus Energy Consumption in Transit. Energies 14(10), 2824.

Altoona, 2015. FEDERAL TRANSIT BUS TEST, Manufacturer: New Flyer, Model: XE40,

CanElevation, 2021. Open Government. High Resolution Digital Elevation Model (HRDEM) - CanElevation Series.https://open.canada.ca/en.

De Cauwer, C., Verbeke, W., Coosemans, T., Faid, S., Van Mierlo, J., 2017. A Data-Driven Method for Energy Consumption Prediction and Energy-Efficient Routing of Electric Vehicles in Real-World Conditions. Energies 10(5).

El-Taweel, N.A., Farag, H.E.Z., Mohamed, M., 2020. Integrated Utility-Transit Model for Optimal Configuration of Battery Electric Bus Systems. IEEE Systems Journal 14(1), 738-748.

El-Taweel, N.A., Zidan, A., Farag, H.E.Z., 2021. Novel Electric Bus Energy Consumption Model Based on Probabilistic Synthetic Speed Profile Integrated With HVAC. IEEE Transactions on Intelligent Transportation Systems 22(3), 1517-1531.

Environment-Canada, 2021. Environment and Natural Resources, Weather, Climate and Hazard,

Ferguson, M., Mohamed, M., Maoh, H., 2019. On the Electrification of Canada's Vehicular Fleets: National-scale analysis shows that mindsets matter. IEEE Electrification

Magazine 7(3), 55-65.

Fiori, C., Montanino, M., Nielsen, S., Seredynski, M., Viti, F., 2021. Microscopic energy consumption modelling of electric buses: model development, calibration, and validation. Transportation Research Part D: Transport and Environment 98.

Franca, A., 2015. Electricity consumption and battery lifespan estimation for transit electric buses: drivetrain simulations and electrochemical modelling, Department of Mechanical Engineering. University of Victoria, p. 162

Gallet, M., Massier, T., Hamacher, T., 2018. Estimation of the energy demand of electric buses based on real-world data for large-scale public transport networks. Applied Energy 230, 344-356.

Gao, Z., Lin, Z., LaClair, T.J., Liu, C., Li, J.-M., Birky, A.K., Ward, J., 2017. Battery capacity and recharging needs for electric buses in city transit service. Energy 122, 588-600.

Hahn, B., Valentine, D., 2019. SIMULINK® Toolbox. In Essential MATLAB for Engineers and Scientists.

He, X., Zhang, S., Ke, W., Zheng, Y., Zhou, B., Liang, X., 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, 714-722.

Hjelkrem, O.A., Lervåg, K.Y., Babri, S., Lu, C., Södersten, C.-J., 2021. A battery electric bus energy consumption model for strategic purposes: Validation of a proposed

model structure with data from bus fleets in China and Norway. Transportation Research Part D: Transport and Environment 94.

Kivekäs, K., Lajunen, A., Vepsäläinen, J., Tammi, K., 2018. City Bus Powertrain Comparison: Driving Cycle Variation and Passenger Load Sensitivity Analysis. Energies 11(7).

Kivekas, K., Vepsäläinen, J., Tammi, K., 2018. Stochastic Driving Cycle Synthesis for Analyzing the Energy Consumption of a Battery Electric Bus. IEEE Access 6, 55586-55598.

Kivekas, K., Vepsäläinen, J., Tammi, K., Anttila, J., 2017. Influence of Driving Cycle Uncertainty on Electric City Bus Energy Consumption, 2017 IEEE Vehicle Power and Propulsion Conference (VPPC), pp. 1-5

Kunith, A., Mendelevitch, R., Goehlich, D., 2017. Electrification of a city bus network— An optimization model for cost-effective placing of charging infrastructure and battery sizing of fast-charging electric bus systems. International Journal of Sustainable Transportation 11(10), 707-720.

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

Lajunen, A., 2018. Lifecycle costs and charging requirements of electric buses with different charging methods. Journal of Cleaner Production 172, 56-67.

Lajunen, A., Kivekäs, K., Baldi, F., Vepsäläinen, J., Tammi, K., 2018. Different Approaches to Improve Energy Consumption of Battery Electric Buses, 2018 IEEE

Vehicle Power and Propulsion Conference (VPPC), pp. 1-6

Lajunen, A., Tammi, K., 2016. Energy consumption and carbon dioxide emission analysis for electric city buses, 29th World Electric Vehicle Symposium and Exhibition (EVS29)

Li, P., Zhang, Y., Zhang, Y., Zhang, Y., Zhang, K., 2021. Prediction of electric bus energy consumption with stochastic speed profile generation modelling and data driven method based on real-world big data. Applied Energy 298.

Ma, X., Miao, R., Wu, X., Liu, X., 2021. Examining influential factors on the energy consumption of electric and diesel buses: A data-driven analysis of large-scale public transit network in Beijing. Energy 216.

Mahmoud, M., Garnett, R., Ferguson, M., Kanaroglou, P., 2016. Electric buses: A review of alternative powertrains. Renewable and Sustainable Energy Reviews 62, 673-684.

Mohamed, M., Farag, H., El-Taweel, N., Ferguson, M., 2017. Simulation of electric buses on a full transit network: Operational feasibility and grid impact analysis. Electric Power Systems Research 142, 163-175.

Mohamed, M., Ferguson, M., Kanaroglou, P., 2018. What hinders adoption of the electric bus in Canadian transit? Perspectives of transit providers. Transportation Research Part D: Transport and Environment 64, 134-149.

Nikel, C., Eldeeb, G., Mohamed, M., 2020. Perceived Quality of Bus Transit Services: A Route-Level Analysis. Transportation Research Record: Journal of the

Transportation Research Board 2674(2), 79-91.

Pamuła, T., Pamuła, W., 2020. Estimation of the Energy Consumption of Battery Electric Buses for Public Transport Networks Using Real-World Data and Deep Learning. Energies 13(9).

Pardo, M.d.C.R., 2017. Uncertainty in Electric Bus Mass and its Influence in Energy Consumption, Mechanical engineering. AALTO UNIVERSITY, p. 63

Qi, X., Wu, G., Boriboonsomsin, K., Barth, M.J., 2018a. Data-driven decomposition analysis and estimation of link-level electric vehicle energy consumption under real-world traffic conditions. Transportation Research Part D: Transport and Environment 64, 36-52.

Qi, Z., Yang, J., Jia, R., Wang, F., 2018b. Investigating Real-World Energy Consumption of Electric Vehicles: A Case Study of Shanghai. Procedia Computer Science 131, 367-376.

Quarles, N., Kockelman, K.M., Mohamed, M., 2020. Costs and Benefits of Electrifying and Automating Bus Transit Fleets. Sustainability 12(10).

Rupp, M., Handschuh, N., Rieke, C., Kuperjans, I., 2019. Contribution of country-specific electricity mix and charging time to environmental impact of battery electric vehicles: A case study of electric buses in Germany. Applied Energy 237, 618-634.

Rupp, M., Rieke, C., Handschuh, N., Kuperjans, I., 2020. Economic and ecological optimization of electric bus charging considering variable electricity prices and $CO2eq$ intensities. Transportation Research Part D: Transport and Environment 81.

Tammi, K., Lajunen, A., 2016. Energy consumption and carbon dioxide emission analysis for electric city buses. Curran Associates, Inc.

Teoh, L.E., Khoo, H.L., Goh, S.Y., Chong, L.M., 2018. Scenario-based electric bus operation: A case study of Putrajaya, Malaysia. International Journal of Transportation Science and Technology 7(1), 10-25.

Vepsäläinen, J., Kivekäs, K., Otto, K., Lajunen, A., Tammi, K., 2018a. Development and validation of energy demand uncertainty model for electric city buses. Transportation Research Part D: Transport and Environment 63, 347-361.

Vepsäläinen, J., Otto, K., Lajunen, A., Tammi, K., 2019. Computationally efficient model for energy demand prediction of electric city bus in varying operating conditions. Energy 169, 433-443.

Vepsäläinen, J., Ritari, A., Lajunen, A., Kivekäs, K., Tammi, K., 2018b. Energy Uncertainty Analysis of Electric Buses. Energies 11(12).

CHAPTER SIX

6. CONCLUSION AND FUTURE RESEARCH

6.1. Summary

Battery electric buses (BEBs) transit system is an essential mobility choice ensuring accessibility and boosting energy efficiency. It is, nevertheless, critical for governments, transit agencies, and utilities to anticipate the effect of an all-electric fleet on the current energy infrastructure before phasing out the use of diesel buses. To that end, the BEB feasibility, including electrifying transit systems, emissions and feasibility, transit system's cost optimization, battery capacity/performance, charging station's spatial distribution, are at the forefront of transportation researcher and transportation agencies.

However, all the above-mentioned research domains are based mainly on the understanding of the energy consumption behaviour, as well as the accurate estimation of the BEB's energy uncertainty. This is a challenging task due to the vast operational range of the BEB's transit networks in real-world environment that constantly varies with the circumstances surrounding each bus trip on the transit network level and the route level. Such variation in operation results in a high-level of uncertainties in the BEB's E_C, attributed to the diversity in the driver behaviour, speed profiles, passenger-hour load profile, route topology, and weather conditions.

Accordingly, with the primary goal of providing a planning framework for BEB deployment in bus transit networks, this thesis aims to construct simulation, data-driven,

and low-resolution models using big data to quantify the BEB's E_C. The thesis aim is achieved due to four major studies that are contained in chapters 2-5. The following subsections present the main findings of each study (i.e. chapter).

6.2. Conclusion of Chapter Two

This study aims to understand the behaviour of bus transit networks in normal conditions and under disruptions. That is to provide us with a reasonable understanding of the bus transit network behaviour under disruptive events such as electricity outages in BEBs.

Toward that end, 1) The complex network theory (CNT) is used to quantify, assess, and comprehend the operational and topological characteristics of the bus transit network. Four topological measures are quantified for 40 Canadian bus transit networks, including degree centrality, betweenness centrality, clustering coefficient, and average path length. 2) The bus transit network behaviours are assessed through static-robustness and dynamic-robustness measures. The static robustness is assessed through three measures applied from the literature, including robustness metric, critical threshold, and the robustness indicator. The dynamic-robustness is assessed through three targeted removal strategies: targeted removal of bus stations with the highest degree centrality, targeted removal of bus stations with the highest betweenness centrality, targeted removal of bus links with the highest weight. 3) A cluster analysis is conducted for classifying the Canadian bus transit networks according to their behaviour under several disruption scenarios.

The main findings of chapter two:

▫ The topological results show no general network classification (e.g. small world, random, or scale-free) noted from the results. Thus, there is a shortage of a network classification that could be holistically applied to Canada's entire bus transit networks. In turn, that shows the significant variation between the bus transit networks and necessitates the investigation of the operation behaviour resulting from any cascading loss in each bus transit network under disruptions.

▫ There were significant contradictions between the three static-robustness measures used in the literature, which strongly argue that their utilization leads to misleading outcomes, as proved in this study. Moreover, these measures fall short in accommodating the unique operational characteristics associated with different bus transit networks.

▫ The dynamic-robustness results show that the targeted removal strategies related to the bus stations cause the highest severe cascading effect on the selected bus transit networks. In comparison, the bus network behaviour after removing the bus links is, to a certain degree, similar to a steady negative relationship.

▫ The cluster analysis resulted in three distinct network clusters that can attribute to the distribution of hub stations and multiple links on the bus transit networks. The network clusters (i.e. network types) resulted from this analysis are: 1) small station sensitive networks, which demonstrated a significant degree of sensitivity associated with removing bus stations with the highest attribute.

These networks clusters feature relatively fewer hub stations and fewer multiple links. Therefore, in this cluster, the impact of station removal is severe. 2) small link-sensitive networks, that feature relatively a higher number of multiple bus routes operating on the same links. Therefore, there is a significant impact on operation associated with removing links with the highest weight. 3) medium-size networks, that are relatively less sensitive to the removal of links or bus stations.

6.3. Conclusion of Chapter Three

This study aims at developing a robust simulation model for estimating accurate rates of the BEB's E_C. Besides, it aims at studying the impact of the vehicular operational, topological, and external parameters on the BEB's E_C. That is to find the parameters that have high significance on the variation that occurs in the consumed energy of the BEB during bus trips. Moreover, this study presents a prediction model for BEB energy consumption to inform the optimal bus route design that diminishes the energy utilization.

Toward that end, 1) A simulation model had been developed to estimate BEB's E_C rates using MATLAB Simulink to overcome the challenges of collecting real-world E_C data. Then, the model is validated using Altoona real-world test results. 2) A full-factorial experiment design is conducted to comprise all the possible operation scenarios that can face the BEB in real-world operation. 3) A regression analysis is performed on the big data of the operation scenarios using a multiple linear regression model (MLR) to identify E_C parameters' significance and study the intertwined relationship between the parameters and

the E_C. 4) A fractional-factorial experiment design is conducted to generate random operation scenarios to test the regression model's validity. 5) we developed five hypothetical scenarios to apprise the optimum transit operation profile design that can enhance energy savings.

The main findings of chapter three:

▫ The validation of the developed simulation model resulted in a superior accuracy within ±5% error in the estimated E_C.

▫ The regression analysis results revealed a significant relationship between the BEB's E_C and the following parameters: road gradient, the initial state of charge, HVAC, road condition, driver aggressiveness, average speed, passenger loading, and stop density.

▫ The results show that the main parameter driving the variation in the E_C rates is the road grade, while the stop density had the lowest impact.

▫ The prediction model was validated using the goodness-of-fit, which shows that the model can explain more than 96.1% of the variation in the E_C rates. Besides, the testing dataset, used to validate the model, resulted in a very accurate E_C prediction with an R^2 of 0.9439.

▫ The result of the five hypothetical scenarios recommends paying attention to the road gradient while planning the bus routes electrification in the transit networks. In addition, bus routes with lower stop density should be considered for transit electrification.

6.4. Conclusion of Chapter Four

This study aims to use the significant parameters from the previous study to develop several data-driven prediction models using big data of the BEB operation scenarios. Then, compression is done between the developed data-driven models to assess and evaluate the techniques with the superior fit.

Toward that end, 1) the validated simulation model developed in chapter three is used to estimate the energy consumption rates of the BEBs. 2) The training dataset (907,199 operation scenarios) generated from the full-factorial experiment design is used in developing the prediction models. 3) we developed seven data-driven modelling techniques, including multiple linear regression analysis (MLR), interpolation method (RBF), decision tree (DT), gradient boosting decision tree (GBDT), support vector machine (SVM), multilayer perception neural network (MLP-NN), and radial-basis neural network (RBNN). 4) The accuracy of the models is assessed using the testing dataset (169,344 operation scenarios). 5) a Sobol sensitivity analysis is performed on the seven data-driven models to evaluate and compare the impact of each parameter in the energy consumption variation.

The main findings of chapter four:

- The results show the high significance of the road gradient, the initial state of charge, road condition, HVAC, passenger loading, driver aggressiveness, average speed, stop density and drag coefficient on the BEB's energy consumption, with different magnitudes.

- The developed data-driven models show a powerful goodness-of-fit with R^2 of 0.96 and 0.89 for training and testing datasets.

- The validation is performed using the testing datasets for the seven data-driven prediction models and shows that the MLR and SVM models are deemed the most accurate models.

- The sensitivity analysis confirms that the road gradient has the most significance on the E_C. Also, the results show that driver aggressiveness and the drag coefficient have the lowest impact on the E_C, for both the first-order and total-order effects.

- We compared the Sobol sensitivity for the developed models with the sample-based sensitivity for the training and testing datasets. We found that the MLR model has the closest behaviour to the training and testing datasets. This, in turn, confirms the high accuracy of the MLR model in predicting the BEB's E_C.

- Both MLR and SVM techniques provide various opportunities to practitioners and scholars to develop accurate and reliable data-driven prediction models for the studies related to BEB's E_C.

6.5. Conclusion of Chapter Five

This study aims at developing and assessing an open-source low-resolution data-based framework to estimate BEB's E_C. The framework integrates a data-driven prediction model and relies on low-resolution open-source data collected by transit providers. This study applies an extensive validation process on the utilized E_C data-driven prediction model.

Toward that end, 1) low-resolution data is collected for six bus routes in Hamilton bus transit network, Canada, representing local, collector, and arterial routes. The collected data include: vehicle positions and instantaneous speed profiles using Automatic Vehicle Location (AVL) and a GPS device; passenger loading, using the automatic passenger count (APC); topological parameters, using CanElevation; weather condition, using Environment Canada historical data. 2) a full-factorial experiment design is performed to generate all possible operation scenarios on the six bus routes, which resulted in 6,048 single bus trips. 3) the validated simulation model developed in chapter three is used to estimate the observed E_C for these bus trips. 4) we used the MLR data-driven prediction model developed in chapter four in this study to obtain the predicted E_C for these bus trips. 5) A three-step validation process is used to check the performance of the proposed E_C data-driven prediction model, based on the error difference between the observed E_C and the predicted E_C. 6) a sensitivity analysis is performed on the data-driven prediction model and the estimated error to check how they are affected based on the changes in the utilized parameters.

The main findings of chapter five:

▫ The validation findings show a powerful goodness-of-fit where the prediction model can explain more than 80.35% (the lowest value for the one-minute segments) from the variation in the E_C rates on the selected bus routes.

▫ The MAPE results show an error percentage from 10% to 18% for all route types, which proves the high accuracy of the proposed prediction model.

- ▫ The passenger loading data (from 0 passengers to 70 passengers) and the weather condition data (between -20 ºC and 30 ºC), road condition (level I to level III) yielded a very accurate prediction with an R^2 of 0.9699, 0.9672, and 0.9582, respectively.

- ▫ The errors estimated from the residuals indicate that the interquartile error range varies between +21% and -21% for all bus trips.

- ▫ The sensitivity analysis shows that the initial state of charge has the highest impact on the estimated error. The lowest error range is at $SoC_i = 100\%$, and this error range increases with the decrease in the SoC_i.

- ▫ Despite neglecting the regenerative brake energy in the data-driven prediction model, the maximum error in the predicted E_C is within ±15% for most bus trips.

- ▫ The proposed framework offers significant contributions for estimating the BEB's E_C without relying on highly sophisticated simulation models or the need to collect real-world data.

6.6. Overall Conclusions

Overall the thesis provides a set of remarks emerged from the findings of all chapters. These could be considered as higher-level conclusions directed at different stakeholders' groups.

For policy-makers, the thesis:

- ▫ Explains the behaviour of the bus transit networks under several disruption scenarios, intending to understand their robustness.

□ Provides the optimum operating profile for the BEB's transit systems that enhances BEB energy utilization.

□ Contributes to understanding the relationship between transit networks' operational, topological, and external features and the BEB's energy consumption to better plan BEB adoption in transit.

□ Provides the means to plan, assess, and optimize the deployment of BEBs in transit.

For transit-agencies, the thesis:

□ Presents a clear picture of the holistic behaviour of the Canadian bus transit networks that can be readily assessed.

□ Encourages the transit planners to pay attention to the road gradient while planning the bus routes electrification.

□ Provides a quick and efficient E_C prediction model for implementing the BEB in transit networks without relying on a sophisticated and technically advanced simulation model.

□ Informs the possibility of using low-resolution data in predicting the BEB's E_C with an acceptable error margin.

For scholars, the thesis:

□ Informs the practitioners and scholars with avenues to quantify the dynamic-robustness of bus transit networks.

▫ Provides various opportunities to develop accurate, prompt, and reliable data-driven models in BEB energy consumption studies.

▫ Provides the scholars with a planned framework to predict the BEB's energy consumption using open-source low-resolution data.

6.7. Limitations and Future Work

Due to data access limitations, the present study did not represent large transit networks in Canada in quantifying the topological measures of the complex network theory (CNT). Hence, it is not considered in assessing the static-robustness or the dynamic-robustness of the Canadian bus transit network. Besides, timetables and passenger flow were not considered in determining the link weight in the CNT assessment.

Although the BEB's E_C prediction model results are in line with previous studies, we did not study the impact of operational parameters on the change of the regenerative braking energy recovery. In addition, the auxiliary power was used in our model as a constant rate, which is sensitive to the operation conditions (e.g. boarding/dwelling). Therefore, future research will focus on accommodating the regenerative brake variable as a predictor in the data-driven model.

We also encourage investigating the proposed models' sensitivity to different BEBs types with varying powertrain configurations. This will indeed yield more clarification on the sensitivity of the bus itself to vehicular, operational, topological, and external parameters.

Therefore, future research will focus on studying the combined effect of BEB energy uncertainty and the associated e-transit network robustness and vulnerability in a single model. As such, this model will provide essential validation to the current BEB transit system optimization and will inform a robust means for resource allocation.