August 2021

Report No. 21-023

Charles D. Baker
Governor

Karyn E. Polito
Lieutenant Governor

Jamey Tesler
MassDOT Secretary & CEO

Tracking the Energy and Emissions of MBTA Rapid Transit Vehicles

Principal Investigator (s)
Dr. Jimi Oke
Dr. Eric Gonzales
Dr. Eleni Christofa

University of Massachusetts Amherst

Research and Technology Transfer Section
MassDOT Office of Transportation Planning

U.S. Department of Transportation
Federal Highway Administration

Technical Report Document Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
21-023	n/a	n/a

4. Title and Subtitle	Tracking the Energy and Emissions of MBTA Rapid Transit Vehicles
5. Report Date	August 2021
6. Performing Organization Code	
7. Author(s)	Jimi Oke, Eleni Christofa, Eric Gonzales, Zhuo Han
8. Performing Organization Report No.	21-023
9. Performing Organization Name and Address	University of Massachusetts Amherst 130 Natural Resources Road, Amherst, MA 01003
10. Work Unit No. (TRAIS)	n/a
11. Contract or Grant No.	
12. Sponsoring Agency Name and Address	Massachusetts Department of Transportation Office of Transportation Planning Ten Park Plaza, Suite 4150, Boston, MA 02116
13. Type of Report and Period Covered	Final Report - August 2021 [May 2020-August 2021]
14. Sponsoring Agency Code	n/a
15. Supplementary Notes	Project Champion - Sean Donaghy, Massachusetts Bay Transportation Authority (MBTA)

16. Abstract

Rapid transit systems are critical components of public transportation networks in urban areas, in terms of their impact on person mobility and on monetary, energy, and environmental costs associated with their operations. To facilitate effective planning for current and future needs, a framework is required that provides important consumption metrics and explains the various contributors to energy consumption and their interactions. This study presents a model that utilizes operational and ridership data for the Massachusetts Bay Transportation Authority's rapid transit system, as well as ambient temperature, to accurately predict system-wide electricity consumption. Linear regression and random forests were used to estimate energy consumption, which could explain a 0.93 and 0.95 variance in the data set, respectively. The model was trained with data from 2019 and tested with data from 2020. The linear regression model provided predictions with RMSE of 2.7 MWh and MAPE of 4.68%, and the random forest model resulted in a RMSE of 2.94 MWh and MAPE of 5.01%. These models are robust and perform well, even with the significant impacts to transit operations associated with the COVID-19 pandemic. This project also investigated COVID-19 impact on the whole system by exploring effects on ridership, energy consumption, cost, and train movement metrics, before and during the pandemic.

17. Key Word	18. Distribution Statement
rapid transit, energy consumption, train trajectory	unrestricted

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
unclassified	unclassified	57	n/a

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

This page left blank intentionally.

Tracking and Reducing the Energy Consumption and Emissions of MBTA Rapid Transit Vehicles (TREEM)

Final Report

Prepared By:

Jimi Oke, Assistant Professor
Principal Investigator

Eleni Christofa, **Associate Professor**
Co-Principal Investigator

Eric J. Gonzales, **Associate Professor**
Co-Principal Investigator

Zhuo Han
Graduate Research Assistant

University of Massachusetts Amherst
130 Natural Resources Road, Amherst, MA 01003

Prepared For:
Massachusetts Department of Transportation
Office of Transportation Planning
Ten Park Plaza, Suite 4150
Boston, MA 02116

August 2021

This page left blank intentionally.

Acknowledgments

Prepared in cooperation with the Massachusetts Department of Transportation, Office of Transportation Planning, and the United States Department of Transportation, Federal Highway Administration.

The Project Team would like to acknowledge the efforts of all state transportation agencies and metropolitan planning organizations who participated in the interviews and shared information with the research team.

Disclaimer

This page left blank intentionally.

Executive Summary

This study of Tracking and Reducing the Energy Consumption and Emissions of MBTA Rapid Transit Vehicles (TREEM) was undertaken as part of the Massachusetts Department of Transportation (MassDOT) Research Program. This program is funded with Federal Highway Administration (FHWA) State Planning and Research (SPR) funds. Through this program, applied research is conducted on topics of importance to the Commonwealth of Massachusetts transportation agencies.

The Massachusetts Bay Transportation Authority (MBTA) operates the public transportation network that serves the Boston metropolitan area. The MBTA's rapid transit network is the fourth busiest in the United States by passenger ridership and includes a light rail line (Green Line) and three heavy rail lines (Red, Orange, and Blue Lines). The energy consumption of the rapid transit system is significant, with the MBTA spending approximately $38 million per year 422 GWh of electricity for the system, including vehicle traction power, signal systems, and stations. The MBTA currently has meters at electric substations throughout the system but no direct measurements of electricity consumption by trains.

This study addresses a need for understanding the contributors to energy consumption in the rapid transit system. This information is useful for planning and predicting energy consumption, which is important for making decisions about purchasing electricity, operating trains, and managing facilities. To this end, the study addresses the following questions:

- What is the relationship between energy consumption and train movement of MBTA rapid transit vehicles?
- Can the energy consumption of the system be reliably predicted?
- What were the impacts of the COVID-19 pandemic on operations and energy consumption?

In light of these questions, this study has two main objectives:

1. To analyze the relationship between energy (including costs and emissions) and rapid transit train movement.
2. To estimate a model to predict systemwide energy consumption at an hourly level.

During the course of the study, the COVID-19 pandemic had a profound impact on rapid transit ridership and operations, which led to greater variation in ridership and train operations than would normally be expected. These changes prompted a third objective:

3. To evaluate the impact of the pandemic systemwide energy consumption.

Methodology

The research method had three main parts. First, an exploratory analysis of the available data from the MBTA on energy consumption and train operations provided general insights about the trends over time and the relationships between the data. Energy consumption data from the MBTA were aggregated hourly for each substation in the system, which accounted for electricity consumption of trains, stations, and other related equipment in the vicinity of the station. Altogether, this accounted for the total electricity consumption of the rapid transit system. Comprehensive data on train movements and passenger ridership were obtained from the MBTA Research Database. This data included time-stamped records of train locations from each vehicle as it moved through the system. An automated process was developed to construct detailed trajectories for each vehicle's movement through the system, including calculation of speed and acceleration, which are important determinants of the energy required for tractive power. These data were visualized in a trajectory dashboard and used to calculate planning metrics that are relevant for monitoring the performance of the MBTA heavy rail system in terms of operations and energy consumption.

Second, a regression model was developed and estimated to relate explanatory factors to the systemwide energy consumption of the MBTA heavy rail system. An exploratory analysis of correlations between the potential explanatory variables and the systemwide energy consumption was used to identify the potential parameters for a regression model. A multiple linear regression (MLR) was then estimated to relate the explanatory variables to the energy consumption. Since the physical relationship between speed, acceleration, and energy is not linear, observations of speed and acceleration were grouped by quantile into six bins for each (allowing for 36 interaction terms). Three techniques were used to select the most relevant variables for the model: lasso regression was used to extract relevant features; the variance inflation factor (VIF) was used to identify multicollinear variables; and the correlation coefficients were inspected to identify collinear variables. These methods were used to develop a model that provides strong explanatory power using only variables that have a statistically significant relationship with energy consumption.

Third, the performance of the model was tested by using 80% of the observations from 2019 as training data and the remaining 20% of the observations for validation. The same model was then tested with data from 2020, which is particularly revealing, because the COVID-19 pandemic has led to significant reductions in rapid transit ridership and train operations, which tested the ability of the model to accurately capture the effect on energy consumption.

Results

Planning Metrics

Annual energy consumption for MBTA rapid transit operations has averaged 422 GWh from 2009 to 2020, with very little fluctuation until 2019. The changes in ridership and operations associated with the COVID-19 pandemic resulted in a reduction in energy consumption by 7% from 2019 (414 GWh) to 2020 (384 GWh). Energy consumption varied by month (with peak consumption in January associated with the coldest weather), day of the week (with fewer trains in operation on weekends), and time of day (again, with fewer trains operating in off-peak hours and no revenue service over night). The main planning metrics for energy consumption and rapid transit operations were identified and are summarized for 2019 in Table 1.

Table 1: Planning metrics

Planning metrics	Mean
Energy per vehicle-mile (MWh/veh-mi)	0.04
Energy per vehicle-hour (MWh/veh-hr)	0.27
Cost per vehicle-mile ($/veh-mi)	1.56
Cost per vehicle-hour($/veh-hr)	10.82

Trajectory Dashboard

The processed train-tracking records were used to construct trajectories for each train that operates in MBTA's rapid transit system. The data for each train can be visualized in a trajectory dashboard (e.g., Figure 1), which summarizes the distance traveled and elapsed time between observations, as well as computed speed and acceleration. The histograms show the distributions of observed speeds and accelerations, and a map shows the locations of observations in the network.

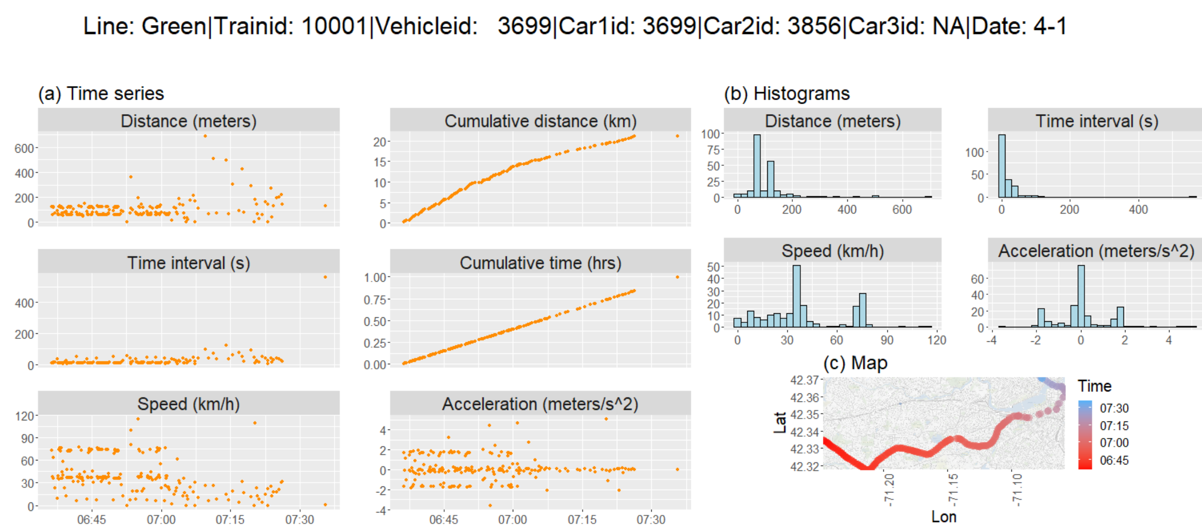


Figure 1: Trajectory dashboard for Green Line, Vehicle ID 3699

Energy Consumption Model

The regression analysis identified the variables with the strongest explanatory power for predicting the systemwide energy consumption. The final model predicts hourly energy consumption as a linear combination of the following explanatory variables:

- Number of trains operating within the hour
- Tap-in rapid transit ridership within the hour
- Daily mean temperature
- Daily mean temperature squared
- Monthly dummy variables
- 26 speed-acceleration interaction terms (out of 36 total)

The estimated model has indicated a good fit between the explanatory variables and the observed energy consumption data. The variables that had the biggest effect on variations in energy consumption are the temperature and the number of operating vehicles. Of the variables included in the model, ridership had the smallest effect.

The performance of the model was first evaluated with parameters estimated to fit a training data set from 80% of the observations in 2019. Applying this model to the remaining 2019 data, the root mean squared error (RMSE) was 2.19 MWh and mean average percentage error (MAPE) was 3.32%, all indicating high predictive accuracy (Figure 2). The model was then tested on data for the entire year 2020, which resulted in an RMSE of 2.7 MWh and MAPE of 4.68% (Figure 3). Although there was some variation in the energy consumption that the model did not fully capture, the predictions are robust even though the model was trained with pre-pandemic data from 2019. The model still produced accurate energy consumption estimates based on observed temperature, operations, and ridership data in 2020, the latter two of which were significantly impacted by the COVID-19 pandemic.

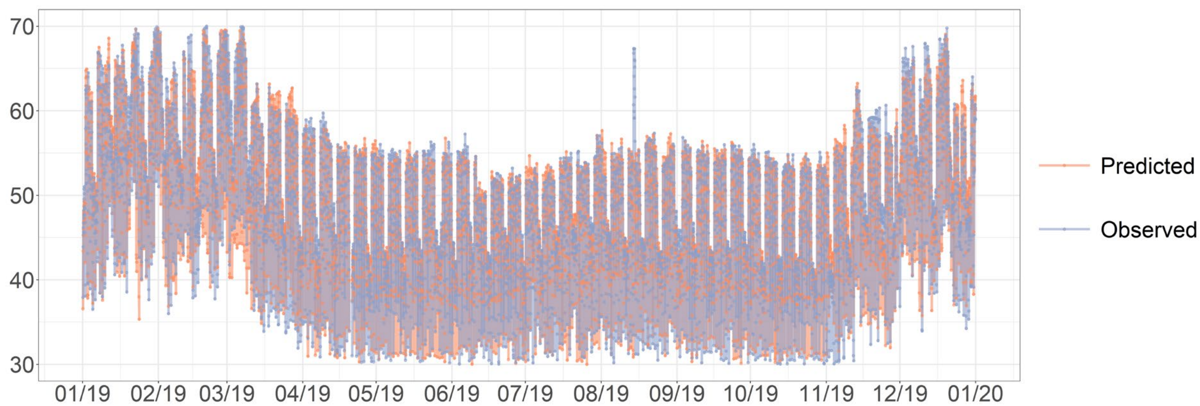


Figure 2: Linear regression model performance on train set and validation set, 2019

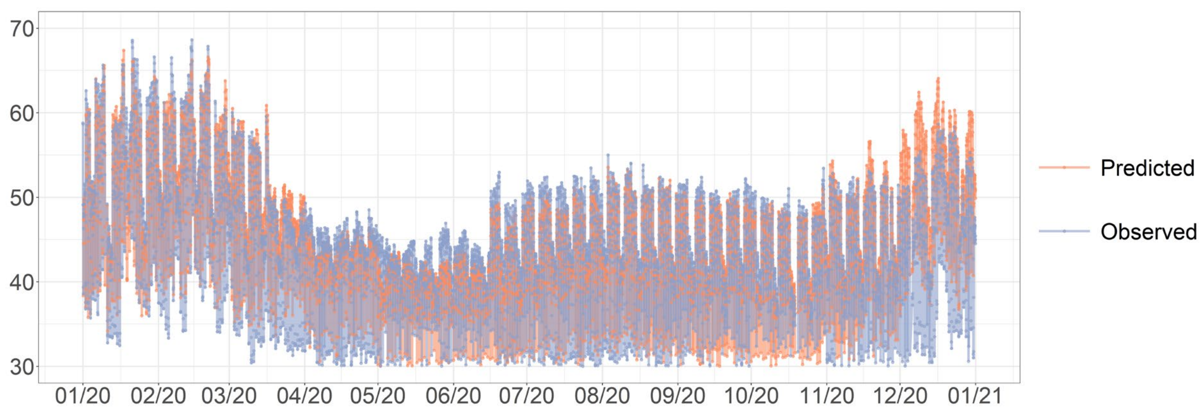


Figure 3: Linear regression model performance on test set, 2020

A random forests model was also estimated to predict energy consumption in this project. The optimal hyperparameters for the Random Forests were 500 estimators (trees) and a random splitting-variable subset size of 40. The entire set of 2019 hourly observations was used for training the model (noting that about 37% of these observations are expected to be in the OOB sample). The research team reserved observations from the year 2020 for predictive performance testing. The random forests model resulted in a RMSE of 2.94 MWh and a MAPE of 5.01%, shown in Figure 4.

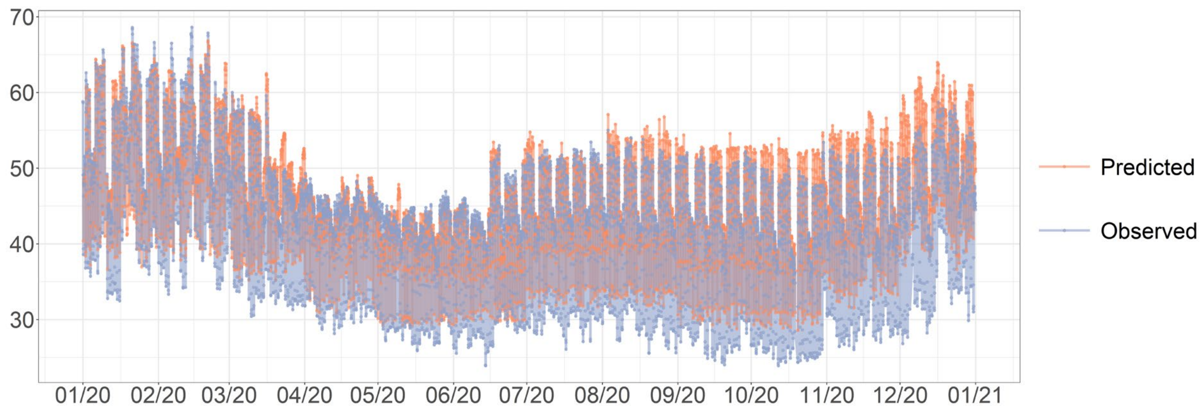


Figure 4: Random forests model performance on test set, 2020

Impacts of COVID-19

The COVID-19 pandemic has brought about significant changes to MBTA rapid transit use and operations, as shutdowns have been implemented to protect public health. Figure 5 shows time series of energy consumption, rapid transit ridership, and train operations (vehicle miles and vehicle hours) in 2019 and 2020. Although ridership plummeted following the lockdown orders in March 2020, and vehicle operations were reduced by almost half, the impacts on energy consumption were relatively low. Train service was restored to normal levels in July 2020, but ridership remained depressed through the end of the year. The model accounts for these changes in demand and operations in estimating energy consumption for 2020, and the low impact of ridership is reflected in the small coefficient in the model. The data and model show that energy consumption is largely driven by fixed components of the system.

Conclusions and Future Work

This project utilized data from MBTA rapid transit operations, ridership, and ambient temperature to accurately predict systemwide electricity consumption. The study resulted in four main contributions and findings:

1. Detailed train trajectories were computed from the MBTA's train tracking data, which characterize the movements of trains that are correlated with energy consumption.
2. A high-performance energy consumption model was developed that is robust to large disruptions, such as the COVID-19 pandemic.
3. The key drivers of system-wide energy consumption were identified as temperature, baseline energy consumption by facilities, and train operations.
4. Ridership was shown to have a very small impact on energy consumption.

Ongoing research efforts include equipping individual trains with accelerometers to collect high-resolution, train-specific data to identify more detailed relationships between train motion and energy consumption. The trajectory analysis tool that was developed can be extended to a real-time trajectory-energy dashboard. The dashboard and model can be used for scenario planning and analysis.

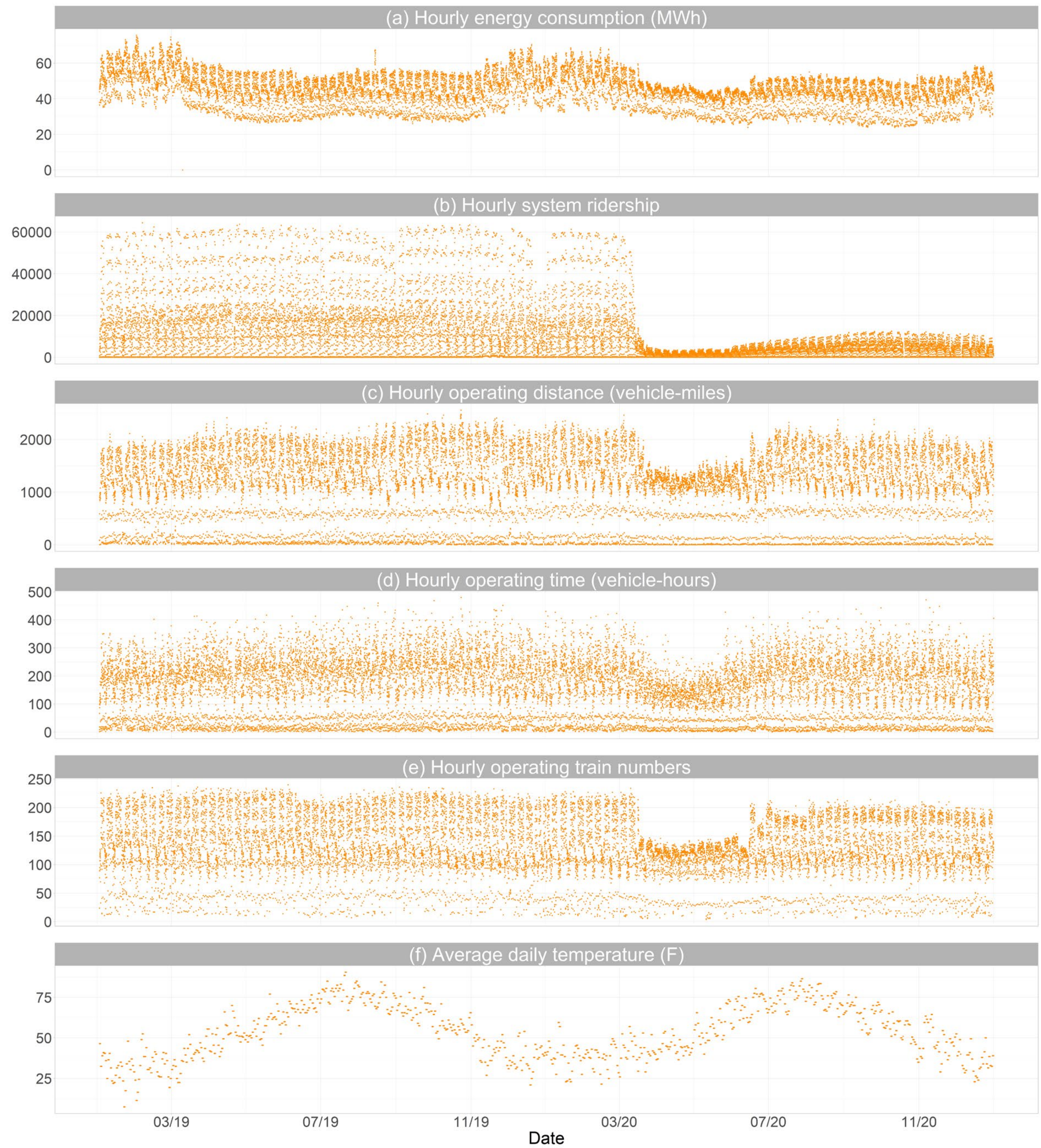


Figure 5: Time series plots of energy, ridership, vehicle-miles and vehicle-hours, Jan. 2019–Dec. 2020, for MBTA rapid transit system

Table of Contents

List of Tables

This page left blank intentionally.

List of Figures

This page left blank intentionally.

List of Acronyms

Acronym	Expansion
COVID-19	Coronavirus Disease
DOT	Department of Transportation
EIA	Energy Information Administration
EPA	Environmental Protection Agency
FHWA	Federal Highway Administration
GHG	Greenhouse Gas
GIS	Geographic Information System
MAPE	Mean Absolute Percentage Error
MassDOT	Massachusetts Department of Transportation
MBTA	Massachusetts Bay Transportation Authority
MLR	Multiple Linear Regression
MSE	Mean Squared Error
MAPE	Mean Absolute Percentage Errors
NOAA	National Oceanic and Atmospheric Administration
RMSE	Root Mean Squared Error
RTV	Rapid Transit Vehicle
VIF	Variance Inflation Factor

This page left blank intentionally.

1.0 Introduction

Urban rapid transit systems are critical for public transportation networks. In transit-oriented high-density cities such as New York, Tokyo, and Beijing, rapid transit is one of the most popular and convenient modes of transportation. According to the International Transport Forum *(1)*, the demand for urban passenger rail is expected to increase from 1.12 trillion passenger-kilometers traveled (pkm) in 2015 to 1.80 trillion pkm in 2050—an increase of 60.7% (Figure 6).

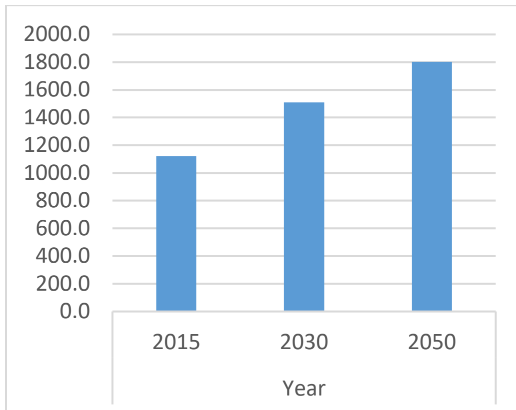


Figure 6: Urban rail passenger-kilometers by mode (billions)

Also, CO_2 emissions due to urban passenger transport were estimated at 2200 million tCO_2e in 2015. This is expected to increase by 37% to 3000 million tCO_2e in 2050 (see Figure 7) *(2)*. Data on energy consumption of urban rapid transit systems, however, are not readily available.

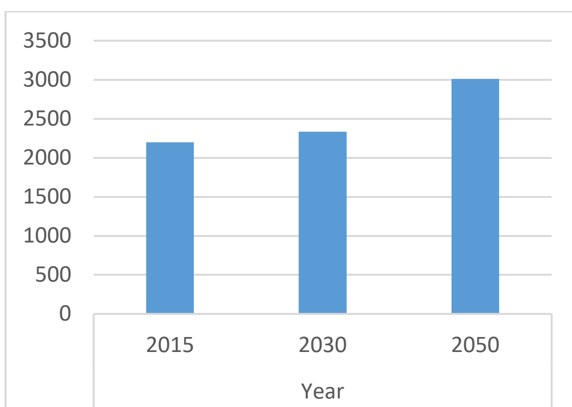


Figure 7: Global urban passenger CO_2 emissions (million tCO_2e)

The heavy- and light-rail (rapid transit vehicle) system of the Massachusetts Bay Transportation Authority (MBTA) is the fourth busiest in the United States (Wikipedia, n.d.). In 2019, the MBTA RTV system operated trains over 1.5 million vehicle-hours and 10.5 million vehicle-miles. In addition, 150 million riders tapped into the system for their mobility needs in the same year. To provide traction power for this system, the MBTA spends $38 million annually. According to MBTA urban transit annual energy consumption statistics (Figure 8), annual consumption has averaged 422 GWh over the past 11 years. This energy consumption has been coupled with increased costs and rising greenhouse gas (GHG) emissions.

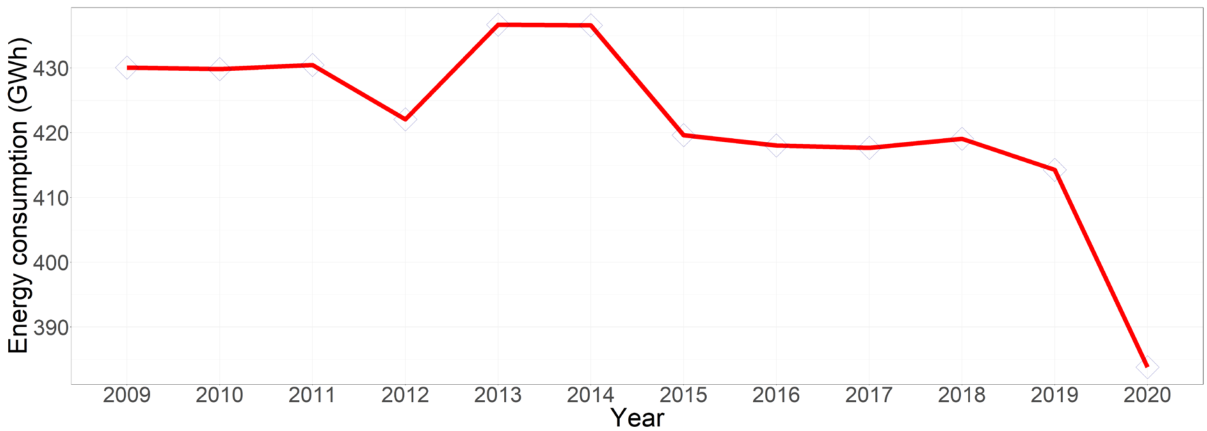


Figure 8: Annual energy consumption of MBTA urban rapid transit system

In order to facilitate effective planning for current and future needs, the MBTA requires a framework that not only provides important consumption metrics but also explains the various contributors to energy consumption and their interactions. Furthermore, this framework should also be useful for predicting energy usage in order to evaluate the relevant impacts of proposed strategy decisions, particularly in response to disruptive events or financial constraints. Ultimately, there is a critical need to reduce costs while still meeting the mobility needs of the surrounding communities in the Boston area.

1.1 Research Questions and Objectives

In order to frame these practical challenges, the research team developed the following research questions:

- What is the relationship between energy consumption and train movement of MBTA rapid transit vehicles?
- Can the energy consumption of the system be reliably predicted?
- What were the impacts of COVID-19 on the system, and what can be learned from the implemented response?

The corresponding research objectives are as follows:

- Analyze the relationship between energy (along with costs and emissions) and train movement.
- Estimate a model to predict system-wide energy consumption.

- Analyze the impacts of the service and demand changes due to COVID-19 on the system and use these to validate the estimated model.

1.2 Report Structure

The rest of the report is organized as follows. Chapter 2 presents a survey of the existing literature on transit system energy modeling. Chapter 3 describes the data structures and sources used in the project. Chapter 4 presents the research framework and methods, followed by the results in Chapter 5. These consist of the planning metrics estimates, model coefficients, and model performance. Chapter 6 analyzes the impacts of COVID-19 on the MBTA RTV system. Chapter 7 concludes with a summary of study findings and possible future directions for research, practice, and knowledge transfer.

This page left blank intentionally.

2.0 Literature Review

2.1 Overview

Various approaches have been implemented in recent years to address the challenges relating to energy consumption and consequent costs and emissions in rapid transit systems. These efforts have ranged from simulation platforms to optimization and machine learning models. Currently, systems researchers have developed effective methods to optimize energy consumption. Yet, these approaches have their limitations when applied to the real world, as they rarely integrate the objectives of reducing energy, emissions, and costs, which are important for service providers and regulators.

2.2 Energy Consumption Simulation Approaches

Some researchers proceeded from the network's operating system to conduct simulations to find ways to reduce energy consumption. Higuera et al. *(2)* proposed that during the acceleration of the metropolitan railway, the stored kinetic energy will be greatly consumed to avoid DC-link perturbation voltage. They established a Voltage Compensation System based on simulation analysis of six railway stations and used this system to evaluate the energy savings. Mao, Mao, and Yu *(3)* built a traction power supply model for a virtual metro network using the Simulink *(4)* environment. The study demonstrated that adjusting the departure interval between metro trains is a viable approach for energy consumption reduction. Ruigang et al. *(5)* designed a new onboard energy storage system (ESS), and this system was improved for energy recovery of the metro vehicle braking. It simulated metro vehicle traction conditions and tested the characteristics of the ESS.

Su, Tang, and Wang *(6)* analyzed how the factors in an optimal train control model influence the traction energy consumption. They established the relationship between energy reduction strategies and energy system design. The energy strategies were estimated by the data of the Beijing Yizhuang metro line. The model indicated that the energy could reduce by 1.5% to 15% if the factors in the model were appropriately adjusted. An electric train energy consumption modeling framework was created by Wang and Rakha *(7)*, which takes instantaneous regenerative braking efficiency into account to support railway simulation systems. The model was calibrated by data from the Metropolitan Area Express Blue Line, using an unconstrained nonlinear optimization program, and was validated using data from Chicago, Illinois. It was confirmed that the energy recovery of the tested Chicago route could reduce the overall power consumption by 20%.

Sanchis and Zuriaga *(8)* developed a computer model for calculating the speed curve of subway vehicles to minimize energy consumption. The model considered the behavior of a single vehicle under manual driving and is limited by various factors to calculate the different commands that the driver wants to execute by the system. The proposed solution can reduce net energy consumption by about 19%. Another example used a neural network model to simulate the non-linear relationship between energy consumption and other variables such as site space design, meteorological factors, and the usage of 19 selected stations. This study provides a useful methodology to reduce the electric energy consumption of new stations *(9)*.

Other efforts have been made to determine the factors responsible for energy consumption in rapid transit systems. For example, one study *(10)* introduced a multicriteria decision method for this purpose and demonstrated that the introduction of renewable energy is important for reducing consumption. Andrade and D'Agosto *(11)* evaluated the energy used in the life cycle of a new subway network in Rio de Janeiro and the emissions generated and avoided. They confirmed that the increase in the share of renewable energy in power generation and the improvement in cement and steel production are key factors in reducing emissions during the life cycle.

Some field surveys and measurement data are utilized to investigate the electricity consumption of underground subway stations. A study conducted by Arikan and Cam *(10)* showed that the lighting system dominated the underground station's energy consumption. This study shows that complete information of non-traction energy usage is effective on the energy consumption reduction and operation costs. Wang et al. *(12)* developed a continuous railway transportation simulator for multimodal energy-efficient routing applications. The simulator was calibrated through an offline optimization program to optimize three model parameters to match the preprogrammed railway schedule. The pilot test of the simulator implementation was conducted to prove that it supports the feasibility of energy-saving travel.

Machine learning theory also has many applications in reducing energy consumption. Fernandez, Roman, and Franco *(13)* introduced an energy consumption model based on a neural network to estimate the energy consumption of electric trains using one line of the Valencia metro network. The inputs used were speed, acceleration, and track grade, and the optimum neural network size was 15 neurons in the hidden layer. The result showed that a neural network can reliably estimate vehicle consumption along a specific route. Additionally, the deep convolutional neural networks proposed by Modi, Bhattacharya, and Basak *(14)* were used to estimate the real-time energy consumption of electric vehicles to reduce driver anxiety.

Malikopoulos and Aguilar *(15)* investigated the driving style factors that have a significant impact on fuel economy. An optimization framework was proposed to optimize the driving style for these driving factors. A set of polynomial meta-models was constructed to reflect the response generated by changes in driving factors. The optimization method used here facilitates a better understanding of the potential energy benefits of conservative driving. Oettich, Albrecht, and Scholz *(16)* proposed three methods to reduce the energy consumption of trains. The first is to adjust the route and schedule of single trains. The second approach is to exploit the energy regenerated during braking in newer trains. The last method is helping the transportation system meet demand by using small vehicles and a flexible itinerary.

2.3 Emissions of Urban Transit Systems

Emissions impacts of urban rapid transit are also critical to sustainability. For example, the life cycle assessment (LCA) method was used to define the system boundary of the Shanghai Metro life cycle, and the relevant resource input and emission output are counted based on actual observation data. The results showed that the total carbon emissions in the construction phases of the Shanghai Metro amounted to 2.68 MtCO2e *(17)*. Zhang, Long, and Chen *(18)* analyzed 18 cities in China using a backward analysis method to estimate the proportion of coal used for rail transit in various cities from 2015 to 2017.

Studies have shown that transportation demand is directly proportional to the carbon emission reduction potential of rail transit. Promoting the development of rail transit technology and reducing energy consumption per unit of travel distance per capita are the fundamental ways to increase the potential of rail transit to reduce emissions.

2.4 Summary

Given the differences in transit networks worldwide and the diversity of their influencing factors, there is currently no perfect model for system energy consumption prediction and inference. At present, most researchers have conducted in-depth research on energy consumption at the level of simulation theory and have produced very good results. However, the exploration of the factors affecting system-wide energy consumption and the development of models describing the relationship between energy consumption and influencing factors in order to render them useful for decision making remain an avenue for further research. Also, there have only been a few pilot projects demonstrating these gains in real-world situations. The TREEM project aims to fill some of these gaps by developing a system-wide energy consumption model that (a) explains the various sources responsible for energy consumption and (b) can be used as a decision-support tool for sustainability planning and disruptive event response in the MBTA RTV system.

This page left blank intentionally.

3.0 Data Summary

The research team utilized the following types of data and sources for this project:

1. Energy consumption data from 2008 through 2020 (provided by the MBTA as an Excel spreadsheet).
2. Time-stamped train location (latitude and longitude) data from the MBTA Research Database (for light and heavy rail vehicles): Tables obtained for 2019 and 2020.
3. Time-stamped tap-in ridership, also from the MBTA Research Database: Tables obtained for 2019 and 2020.
4. Energy costs for the system from 2008 through 2020 (provided by the MBTA as an Excel spreadsheet).
5. Daily average temperature for Boston for 2019 and 2020 (obtained from National Oceanic and Atmospheric Administration [NOAA] historical records).

3.1 Energy Consumption

The MBTA provided a spreadsheet with the hourly energy consumption data of the system for the years 2008 through 2020. In this section, data are visualized and described at the annual, monthly, daily, and hourly timescales.

3.1.1 Annual Patterns

On average, the MBTA rapid transit system consumes 422 GWh of energy (calculated for the period 2008 through 2020). The annual consumption and year-on-year change are shown in Table 2. From 2016 onward, energy use was largely stable at just under 420 GWh. In 2020, it decreased by 7%, due to the service reductions as a result of the COVID-19 pandemic.

Table 2: Annual energy consumption and percentage change, year-on-year

Year	Energy Consumption (GWh)	% Change
2009	430.06	
2010	429.81	0
2011	430.47	0
2012	422.07	-2
2013	436.68	3
2014	436.58	0
2015	419.62	-4
2016	418.04	0
2017	417.66	0
2018	419.07	0
2019	414.27	-1
2020	383.82	-7

3.1.2 Monthly Patterns

At the month level, energy use is greatest in the winter months (Figure 9). Peak usage is observed in January (average of 39 GWh). From April through November, energy consumption is visibly lower, presumably due to reduced third-rail heating needs. A smaller peak, however, is observed in July, which is typically the hottest month of the year. Thus, weather is shown to be a significant explanatory variable for energy consumption.

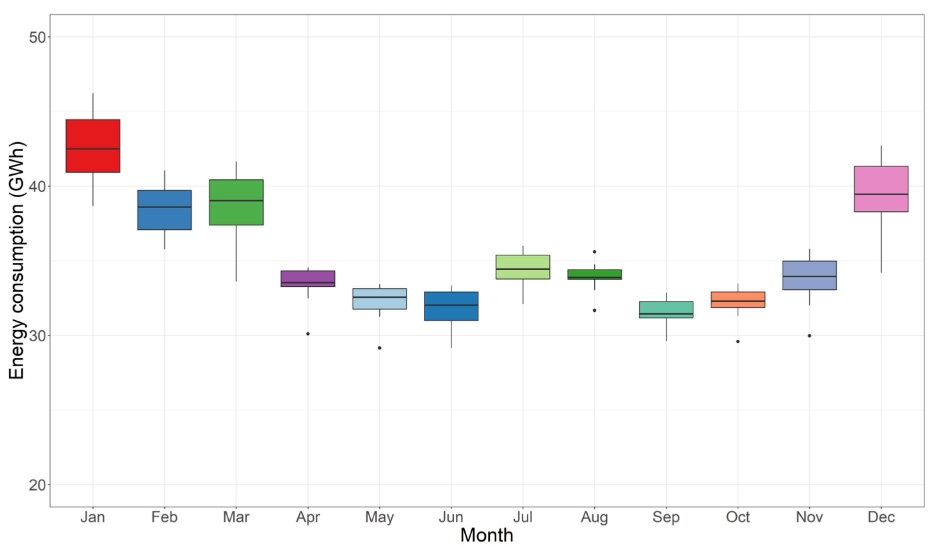


Figure 9: Boxplot of monthly energy usage

3.1.3 Daily Patterns

At the day level, the average energy usage is just over 1 GWh (Table 3). As seen in Figure 10, there is a marked difference between weekday consumption and that of weekend days. The seasonality is even more strongly observed, as the peak days occur during the winter (heating needs), and shorter second peaks occur in the middle of the summer (cooling needs).

Table 3: Daily energy consumption statistics

Year	Mean (GWh)	Maximum (GWh)	Minimum (GWh)	Median (GWh)
2009	1.18	1.66	0.84	1.13
2010	1.18	1.60	0.89	1.16
2011	1.18	1.66	0.74	1.15
2012	1.15	1.49	0.85	1.16
2013	1.20	1.59	0.87	1.18
2014	1.20	1.67	0.88	1.16
2015	1.15	1.60	0.83	1.12
2016	1.14	1.54	0.83	1.12
2017	1.14	1.59	0.85	1.11
2018	1.15	1.59	0.83	1.13
2019	1.13	1.62	0.87	1.09
2020	1.05	1.43	0.82	1.03

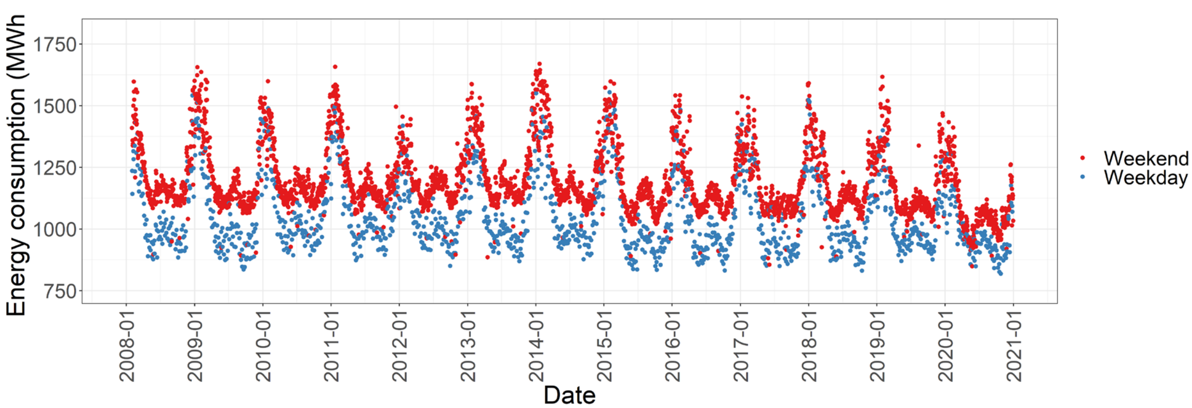


Figure 10: Time series of daily energy consumption

3.1.4 Hourly Energy Consumption

The hourly energy distribution indicates that the peak consumption period occurs from 7 to 9 AM and from 4 to 6 PM (Figure 11). The average hourly peak energy consumption in 2019 was 53 MWh, and the lowest value was 33 MWh. "Overnight" is the period during which there is minimal train movement and no ridership. This occurs between 2 and 4 AM. Yet, there is a baseline average energy consumption of 34 MWh during this period. The overall average hourly energy consumption from 2008 to 2010 is 48 MWh (Table 4).

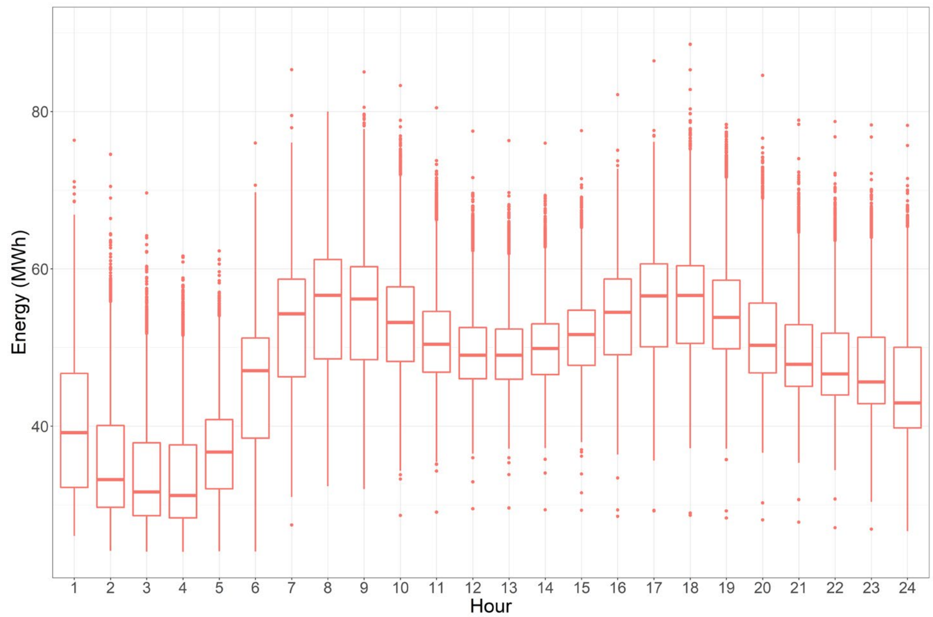


Figure 11: Boxplot of hourly energy consumption

Table 4: Hourly energy consumption in different periods

Period of day	Average hourly energy consumption (MWh)
Morning peak	54
Afternoon peak	55
Overnight	34
Overall	48

3.2 Train Operations

The research team obtained train location data from the MBTA Research Database. Due to the size of the tables, only data for 2019 and 2020 were downloaded and processed. The number of unique trains running in each hour is shown for the four lines (Red, Blue, Green, and Orange Lines) in Figure 12. As will be discussed in the next chapter, a pipeline was developed to compute trajectories from the time-stamped locations. Furthermore, the trajectory details for each unique train were visualized in a format that could potentially serve as a future online dashboard monitoring platform for the system.

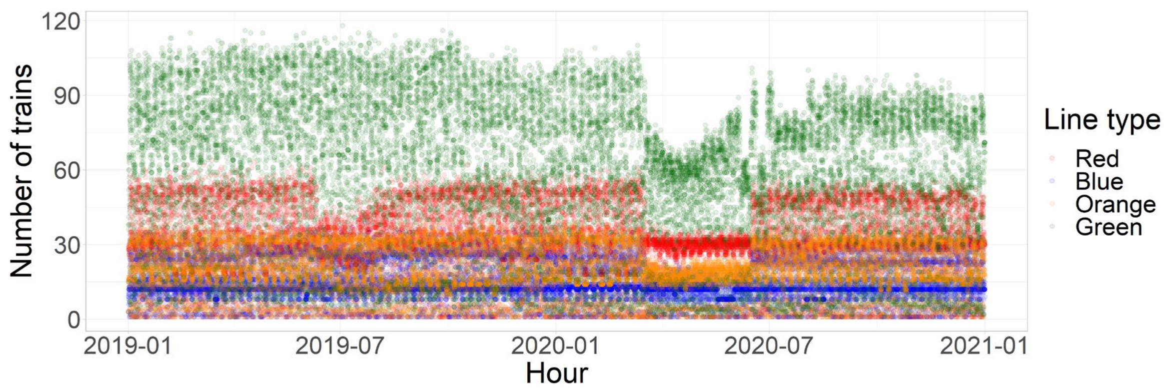


Figure 12: Hourly operating train numbers

3.3 Ridership

The research team obtained system-wide tap-in ridership from the MBTA Research Database for 2019 and 2020, aggregating the observations at the hourly level. The time series is shown Figure 13. A steep decline in ridership began in March 2020 as a result of the COVID-19 pandemic and lockdown policies.

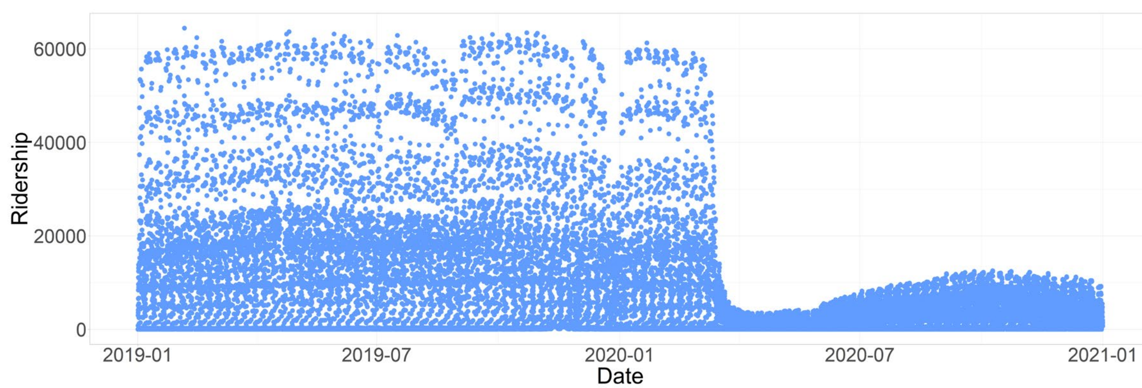


Figure 13: Tap-in ridership counts for MBTA RTV system, Jan. 2019–Dec. 2020

3.2 Energy Cost

In 2019, the energy bill of the MBTA rapid transit system was \$16 million (Figure 14). The average monthly cost was \$1.3 million. The monthly cost between January and June was significantly higher (\$1.55 million on average) than the remaining months in the year (\$1.17 million on average). The peak monthly cost occurred in April (\$1.62 million). Another smaller peak occurred in September (\$1.20 million).

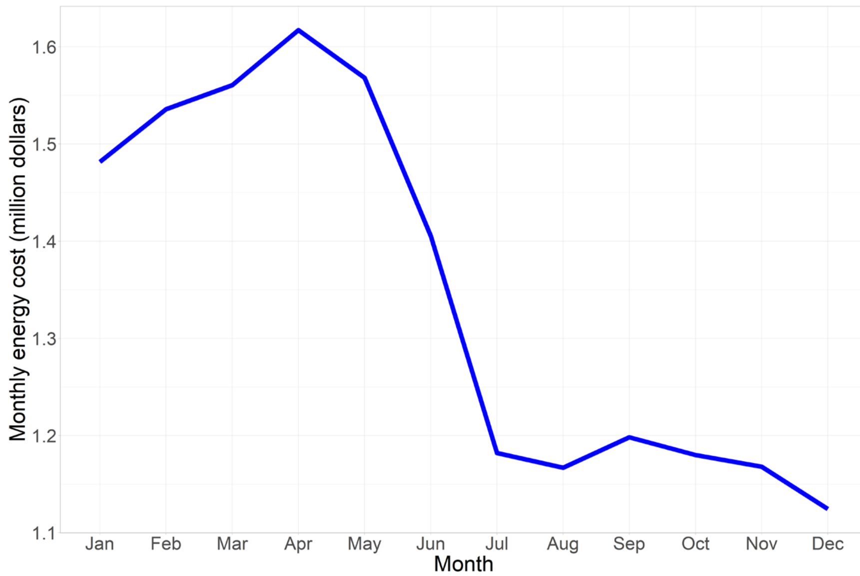
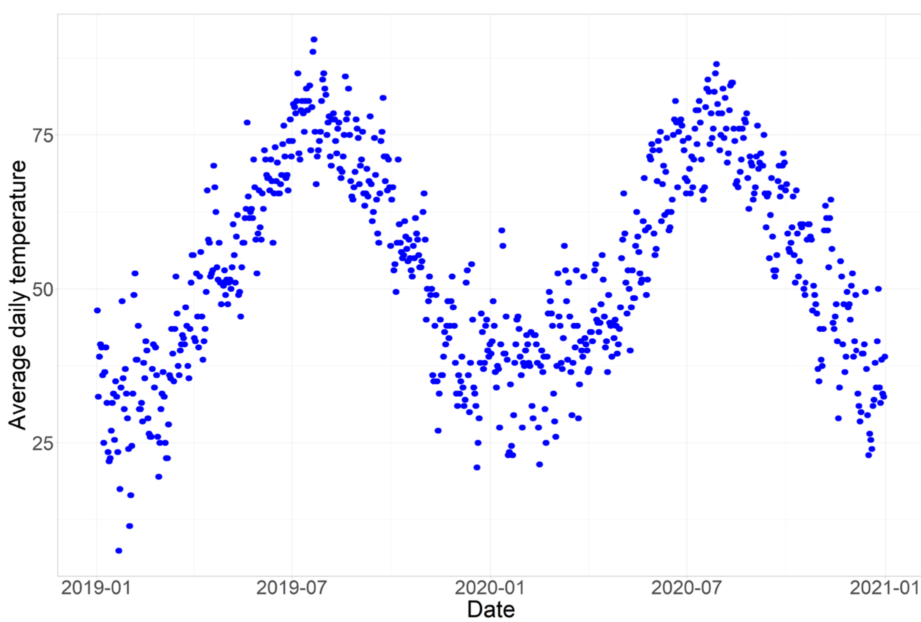


Figure 14: Monthly energy cost of RTV operations in MBTA, 2019

3.3 Weather

The research team obtained average daily temperatures in Boston from the National Oceanic and Atmospheric Administration (NOAA) database for the years 2008 through 2020. Figure 15 shows the time series of the temperatures from 2019 to 2020. As expected, a seasonal pattern was observed in the data. The lowest temperatures are recorded in January (average: 34°F). The highest temperatures are recorded in July (average: 77°F).



Source: NOAA

Figure 15: Average daily temperature in Boston, Jan. 2019–Dec. 2020

4.0 Research Methodology

4.1 Research Framework

The research team developed an integrated framework to (a) obtain and process data from a variety of sources; (b) compute train trajectories (distance, time, speed, and acceleration); (c) generate model input variables at the hourly level; (d) estimate planning metrics; and (e) model and predict hourly energy consumption. The framework is depicted in Figure 16.

The team created a pipeline to query the MBTA Research Database and download heavy and light rail location tables, along with tap-in ridership tables, for a given month and year. Location data were used to compute trajectories and other train operation variables. Daily average temperature data were obtained from NOAA and spreadsheets for hourly energy consumption and monthly costs from the MBTA. All these variables were then integrated into a combined table and aggregated at the desired temporal level (hourly). Using this integrated table, the team computed planning metrics for energy, cost, and emissions. Then, the team used these as training and validation data for the estimated system energy model.

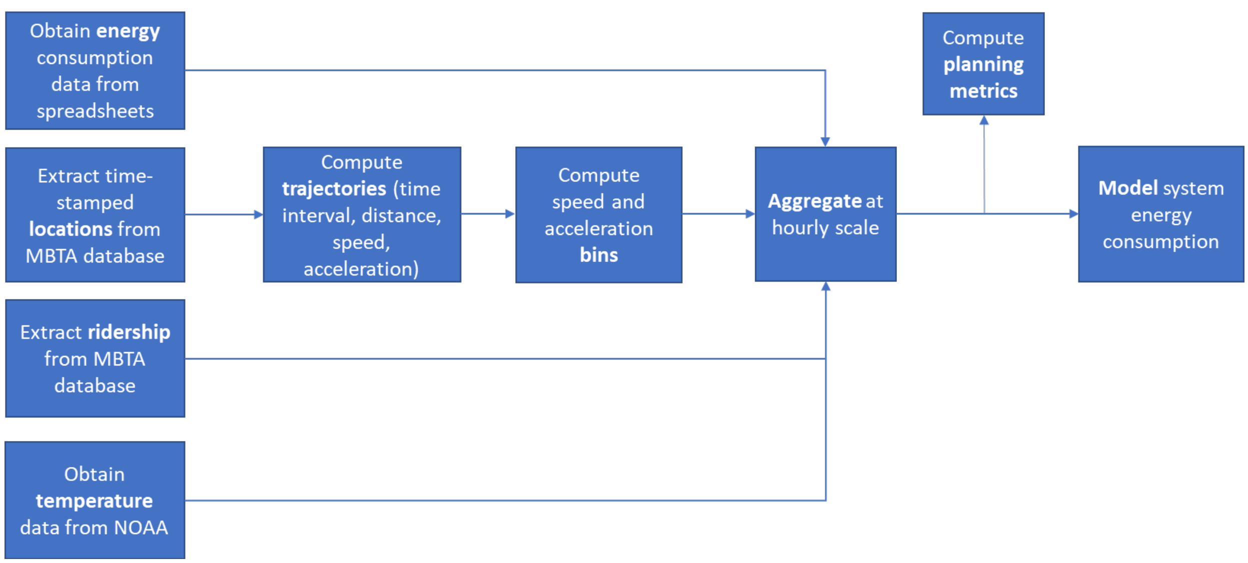


Figure 16: Research framework flowchart

4.2 Trajectory Computation and Dashboard

Using location data, the research team calculated the distance, time interval, speed, and acceleration for each unique train ID in a given day. Together, these measures constitute the trajectory of the given train. An example of a Green Line trajectory is shown in Figure 17. Based on the coordinate data in the raw trajectory table, the team computed the distance between successive pairs of time records and also computed the corresponding time interval. Then the speed, acceleration, cumulative distance, and cumulative time could be calculated accordingly. All of these measurements are visualized by time series in the dashboard (Figure 17[a]). Part (b) of the dashboard indicates the distribution of the above computation results. This will help viewers to better understand the operation status of one train, such as how long it operates at a high speed or high acceleration. Finally, Part (c) spatially depicts the train trajectory on a map, color-coded by time.

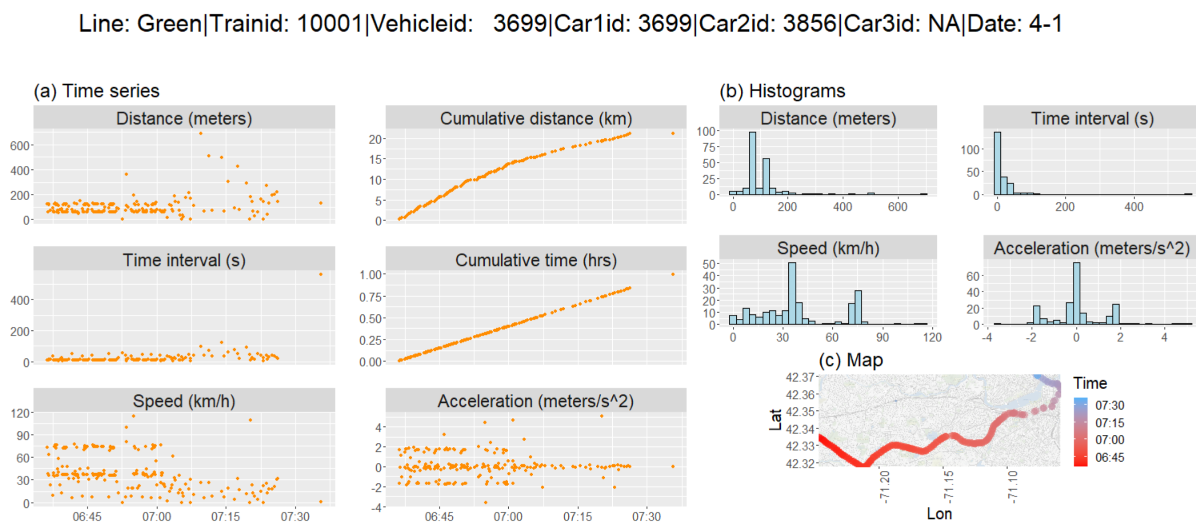


Figure 17: Green Line trajectory dashboard

Further sample dashboards for the other three lines are shown in Appendix 8.0.

4.3 Computation of Planning Metrics

System-wide estimates of energy, costs, and emissions are critical for sustainable planning and management. The objective of this task was to therefore calculate these planning metrics based on both vehicle distances and vehicle operating times. Energy metrics were estimated using monthly data provided by MBTA for 2019. Cost metrics were also estimated based on monthly billing data for the same year. In order to estimate emissions, the research team used emissions intensities for the Commonwealth of Massachusetts provided by the Energy Information Administration (EIA). The following equation was used to obtain the 95% confidence intervals for each of the six metrics:

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

where:

CI: confidence interval
$\bar{x}$: sample mean
z: critical value (equal to 1.96 for 95% confidence level)
s: standard deviation
n: sample size (in this case, 12)

4.4 Speed and Acceleration Binning

The energy consumption of trains fundamentally depends on not just their mass but also on their speeds and rates of speed change (acceleration). Given the high-level system model objective, the contribution of train movements to energy consumption can be captured by observing how much time is spent at various speeds and accelerations. This allows for a fewer number of variables (depending on how many intervals are used) and, therefore, less uncertainty in the model parameters.

In this case, the project team created equal probability bins for speed and acceleration. After testing with different bin numbers, the team selected Bin 6 for both speed and acceleration. These are described in Table 5. Each bin is initially taken as an indicator variable (1 or 0) for each time interval trajectory observation for each vehicle. It is then converted to a time variable as the length of the corresponding interval in which it is observed.

Table 5: Speed and acceleration bins

Bin Number	Percentile Range	Speed (miles/h)	Acceleration (m/s^2)
1	[0, 16.7)	[0, 3.8)	[-5.0, -0.4)
2	[16.7, 33.3)	[3.8, 11.0)	[-0.4, 0)
3	[33.3, 50)	[11.0, 15.5)	[-0.1, 0)
4	[50, 66.7)	[15.5, 22.6)	[0, 0.1)
5	[66.7, 88.3)	[22.6, 30.1)	[0.1, 0.7)
6	[88.3, 100]	[30.1, 104]	[0.7, 5.0]

Based on the fundamental physics of train energy consumption, both the speed and acceleration of a vehicle are contributors to the energy demand. Thus, the team further computed speed-acceleration bin-time variables using each of the bins. This resulted in 36 combinations of speed and acceleration levels. Each interval observation would then have an indicator corresponding to the matching speed-acceleration bin it represents. When the data were aggregated at the hour level, the resulting speed-acceleration interaction variables then denoted the total time collectively spent at a given speed, acceleration, or speed-acceleration interval.

For convenience, the notation "SxAy" represents the speed-acceleration interaction time variable. Thus, the symbol "S1A1" represents the speed-acceleration interaction time for Speed Bin 1 (0 to 3.8 mph) and Acceleration Bin 1 (-5.0 to -0.4 m/s^2).

4.5 Linear Regression

The project team used a multiple linear regression (MLR) approach to predict hourly energy consumption based on the available potential explanatory variables. MLR has the advantage of being highly interpretable due to its simplicity, without sacrificing performance. the team used a variety of techniques to select the most relevant variables for the system model:

- Lasso regression for feature extraction.
- Variable inflation factor (VIF) to identify multicollinear variables.
- Correlation coefficient inspection to identify collinear variables.

Applying the above methods, the team compared the fitness of candidate models using the R^2 statistic, as well as assessing the validation and test performances.

The model equation is given by:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i$$

where:

Y is the hourly energy
β_i are the coefficients
X_i are the explanatory variables

For training purposes, the research team used 80% of the hourly observations in 2019, reserving the remaining 20% for validation. 2020 observations were used to test the performance of the final model, in addition to analyzing the impacts of COVID-19.

4.6 Random Forests

A random forests (RF) model was also estimated to predict energy consumption in this study. Random forests is an ensemble learning approach that estimates multiple regression trees based on respective bootstrap samples of the data. It mitigates noise and bias by using a random fixed-size subset of variables at each branching (node-splitting) step of tree partitioning.

The two hyperparameters the modeler must select are the number of estimators (trees) and size of the subset of random features to be considered for tree partitioning.

The goodness of fit of a random forests model is determined based on error metrics computed on the out-of-bag (OOB) sample. OOB observations are those that are not present in any of the bootstrap samples. Thus, the OOB metrics serve as an estimate of the validation error of the model. Ensemble models, such as random forests, are less interpretable than parametric approaches. However, in random forests, variable importance can be computed from the tree partitioning process, which ranks the relevance of each of the explanatory variables to the dependent variable.

In this application, the best hyperparameters for the random forests model were 500 estimators (trees) and a random splitting-variable subset size of 40.

The entire set of 2019 hourly observations was used for training the model (noting that about 37% of these observations are expected to be in the OOB sample). The research team reserved observations from the year 2020 for predictive performance testing.

This page left blank intentionally.

5.0 Results

5.1 Planning Metrics

The planning metrics (estimated on 2019 energy and cost data) are shown in Table 6 and Figure 18. All these metrics are estimated at the month level. The project team observed that energy and cost (and, consequently, emissions) are more sensitive to operating time than to operating distance (Table 6). Specifically, the per vehicle-hour energy/cost/emissions is one order of magnitude greater than the corresponding per vehicle-mile metric. This demonstrates that operating time is more relevant than operating distance. This could be helpful for the decision makers as they assess strategies for adjusting train schedules or responding to events and constraints in order to save energy and reduce costs.

Table 6: Planning metrics

Planning metrics	Mean	Lower Confidence Bound	Upper Confidence Bound
Energy per vehicle-mile (MWh/veh-mi)	0.04	0.036	0.043
Energy per vehicle-hour (MWh/veh-hr)	0.27	0.25	0.30
Cost per vehicle-mile ($/veh-mi)	1.56	1.38	1.75
Cost per vehicle-hour($/veh-hr)	10.82	9.46	12.18
GHG emissions per vehicle-mile (tCO2e/veh-mi)	18.73	17.03	20.45
GHG emissions per vehicle-hour (tCO2e/veh-hr)	129.31	117.11	141.51

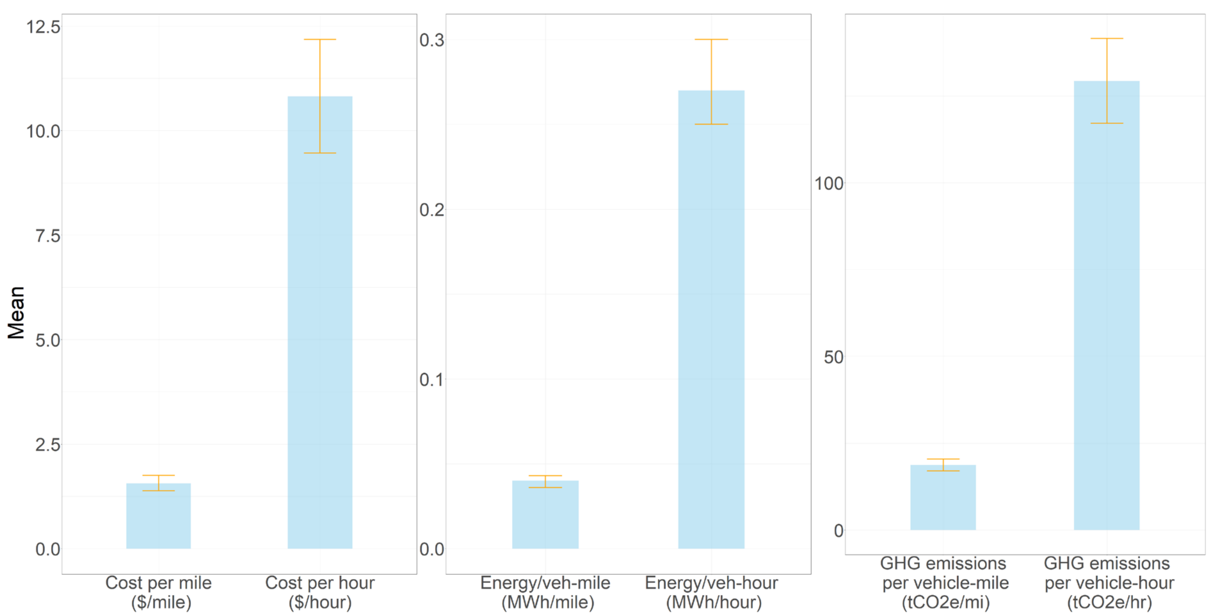


Figure 18: Planning metrics

5.2 Multiple Linear Regression Model

In the final selected model, the hourly energy is predicted on the following variable categories:

- Number of trains
- Tap-in ridership
- Temperature
- Monthly dummy variables
- Speed-acceleration interaction terms

The model has an R^2 of 0.93, indicating that it explains 93% of the variance in the data.

The estimated coefficients are summarized in Table 7. In order to provide a better sense of the contribution of each variable to the hourly energy consumption, the team included the average value of the variables from 2019. The team also showed the average effect, which was obtained by multiplying the coefficient and the average value of the corresponding variable. For reference, the average hourly energy in 2019 was 47.3 MWh. Thus, the intercept alone contributes a sizable net positive effect of 58.5 MWh. Temperature is next (with the linear and quadratic portions indicating the contributions of colder and hotter temperatures). In comparison, the net effect of ridership is quite small at 0.3 MWh.

As the research team was unable to explicitly capture the effects of third-rail heating, and also due to unknown interactions with temperature, the team included monthly dummy variables in the model to capture these variations. January is the baseline month. On average, there was a larger reduction in energy usage as the months proceeded from March through November. The greatest savings were seen in May and October, compared to the baseline of January. This behavior mirrors the pattern shown in Figure 9.

The team also analyzed how the energy varies with the speed-acceleration interaction variables. The coefficients of the relevant variables are also included in Table 7 and visualized using a matrix heatmap in Figure 19. Nearly all of the interaction terms involving Acceleration Bin 1 (A1) were negative, except for S3A1. One potential reason for this is that some trains don't process regenerative process. The negative sign indicates a net energy saving and is reflective of the trains in the system with regenerative braking capabilities. In contrast, the positive sign indicates that the train is accelerating. The project team observed average effect of S6A5 to energy consumption was 2.13MWh, which made the greatest contribution to energy consumption around all positive interaction terms. This also indicated that trains spent most of accelerating time on operating with Speed Bin 6 and Acceleration Bin 5. Average hourly interaction terms effect based on 2019 measurements was 0.94 MWh.

Table 7: Model coefficients and variables contributions to energy consumption

Category	Coefficient	Value	Corresponding variable	Description	2019 Average	Average effect
	β_0	58.48		Intercept		58.48
	β_1	-0.81	X_1	Temperature (F)	53.5	-43.3
	β_2	0.01	X_2	Sq. Temperature $(F)^2$	3,155.2	31.6
	β_3	0.06	X_3	Number of trains	129.8	7.8
	β_4	1.49×10^{-5}	X_4	Ridership	17,299.1	0.3
Monthly dummy	β_5	1.67	X_5	February		1.67
	β_6	-0.52	X_6	March		-0.52
	β_7	-3.92	X_7	April		-3.92
	β_8	-5.49	X_8	May		-5.49
	β_9	-4.86	X_9	June		-4.86
	β_{10}	-4.50	X_{10}	July		-4.50
	β_{11}	-3.37	X_{11}	August		-3.37
	β_{12}	-4.41	X_{12}	September		-4.41
	β_{13}	-5.44	X_{13}	October		-5.44
	β_{14}	-3.56	X_{14}	November		-3.56
	β_{15}	1.28	X_{15}	December		1.28
Speed-acceleration interaction time	β_{16}	-0.38	X_{16}	S1A1	0.92	-0.35
	β_{17}	-0.23	X_{17}	S2A1	1.99	-0.46
	β_{18}	0.59	X_{18}	S3A1	1.623	0.96
	β_{19}	-0.19	X_{19}	S4A1	1.617	-0.31
	β_{20}	-0.81	X_{20}	S5A1	1.28	-1.04
	β_{21}	0.13	X_{21}	S1A2	11.27	1.47
	β_{22}	0.14	X_{22}	S2A2	9.42	1.32
	β_{23}	1.23	X_{23}	S3A2	1.57	1.93
	β_{24}	0.99	X_{24}	S5A2	0.82	0.81
	β_{25}	0.03	X_{25}	S2A3	10.61	0.32
	β_{26}	0.27	X_{26}	S3A3	2.007	0.54
	β_{27}	0.52	X_{27}	S4A3	1.07	0.56
	β_{28}	0.59	X_{28}	S5A3	1.11	0.65
	β_{29}	-1.81	X_{29}	S6A3	0.59	-1.07
	β_{30}	0.02	X_{30}	S2A4	12.58	0.25
	β_{31}	-0.08	X_{31}	S4A4	3.49	-0.28
	β_{32}	0.27	X_{32}	S5A4	3.12	0.84
	β_{33}	0.88	X_{33}	S6A4	0.951	0.84
	β_{34}	0.63	X_{34}	S2A5	0.954	0.60
	β_{35}	-0.74	X_{35}	S3A5	1.92	-1.42
	β_{36}	-0.21	X_{36}	S4A5	3.12	-0.66
	β_{37}	1.01	X_{37}	S6A5	2.11	2.13

Category	Coefficient	Value	Corresponding variable	Description	2019 Average	Average effect
	β_{38}	2.02	X_{38}	S3A6	0.23	0.46
	β_{39}	-2.10	X_{39}	S5A6	1.46	-3.07
	β_{40}	0.46	X_{40}	S6A6	2.59	1.19

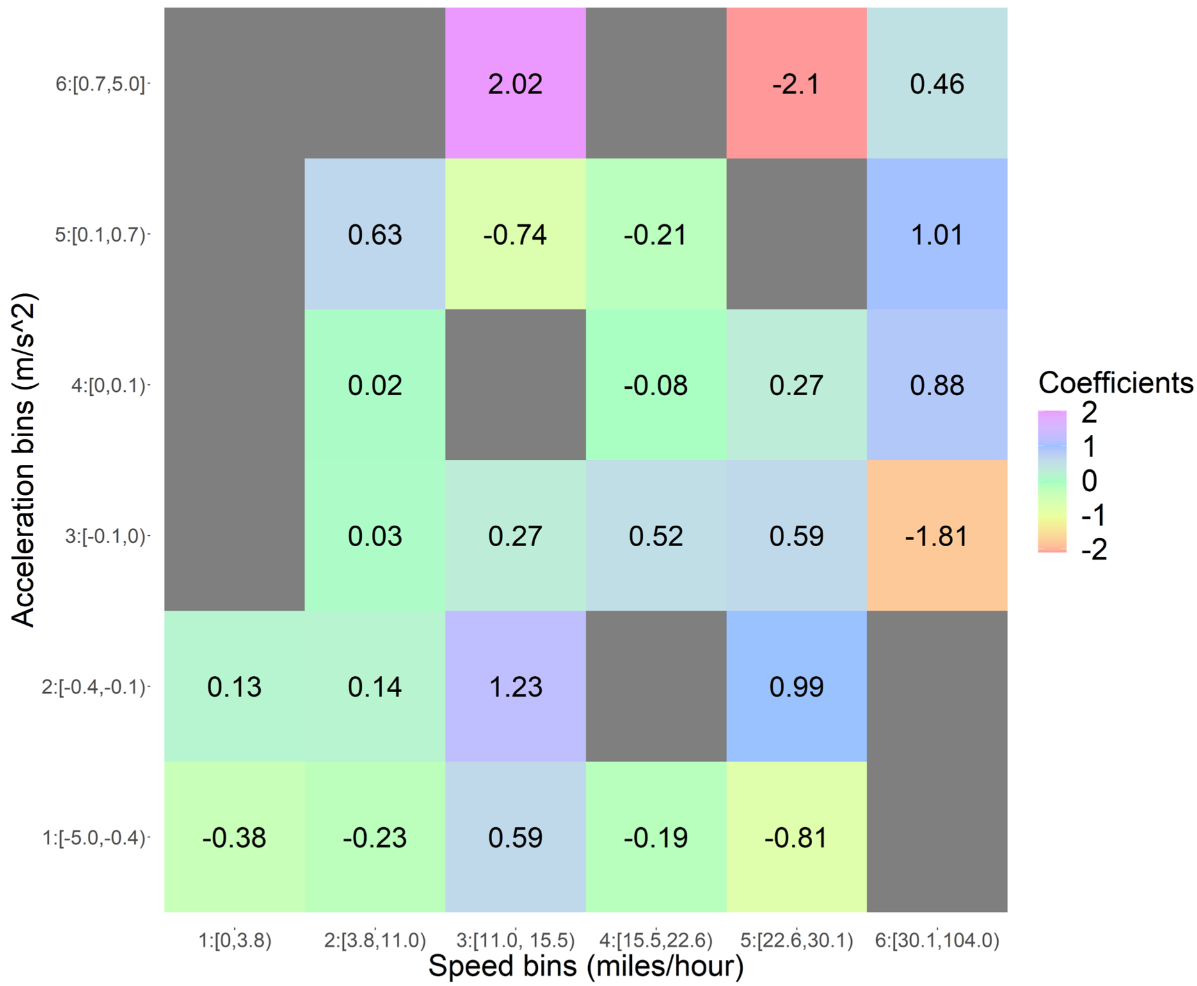


Figure 19: Coefficients of interaction terms

5.3 Random Forests

While the random forests model does not result in an explainable or parametric form, it allows us to understand the most relevant variables with respect to the predicted variable. This is given via the variable importance plot (Figure 20). The number of operating trains was the variable with the greatest impact on energy consumption. In addition, average daily temperature, ridership, operating time, monthly factors and some train movements (S6A5, S5A2, S3A2) were also very important for energy consumption. This result further confirmed that trains spent most of the time operating in Speed Bin S6 (30.1 to 104 mph) and Acceleration Bin A5 (0.1 to 0.7 m/s^2).

Generally, the importance rankings agreed with the coefficients and average effect sizes of the corresponding variables in the linear model.

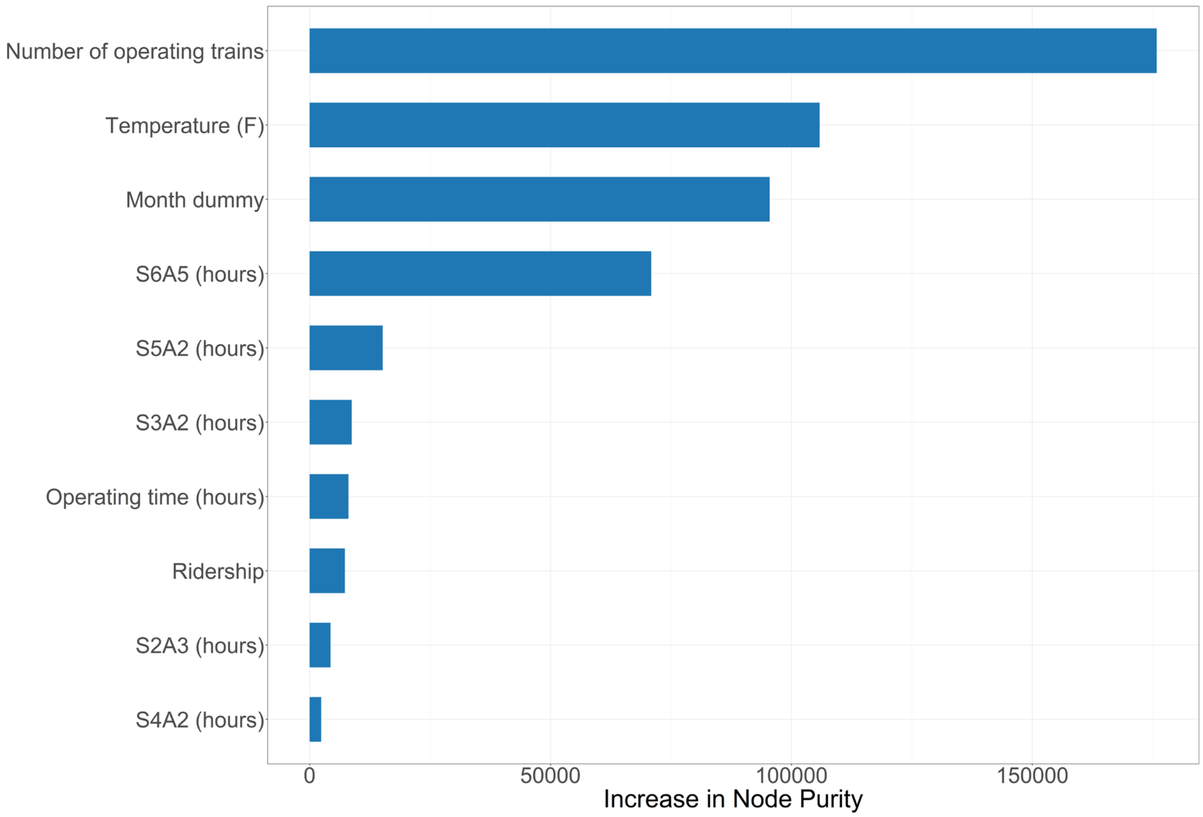


Figure 20: Top 10 important variables selected by random forests model

5.4 Model Performance

5.4.1 Linear Regression

The model performance on the validation set is shown in Figure 21. With an $R^2 = 0.93$, the team estimated a model that reasonably fits the data. The training RMSE was 2.19 MWh, and the MAPE was 3.32%, which indicate that this model provides high predictive accuracy in these two data sets.

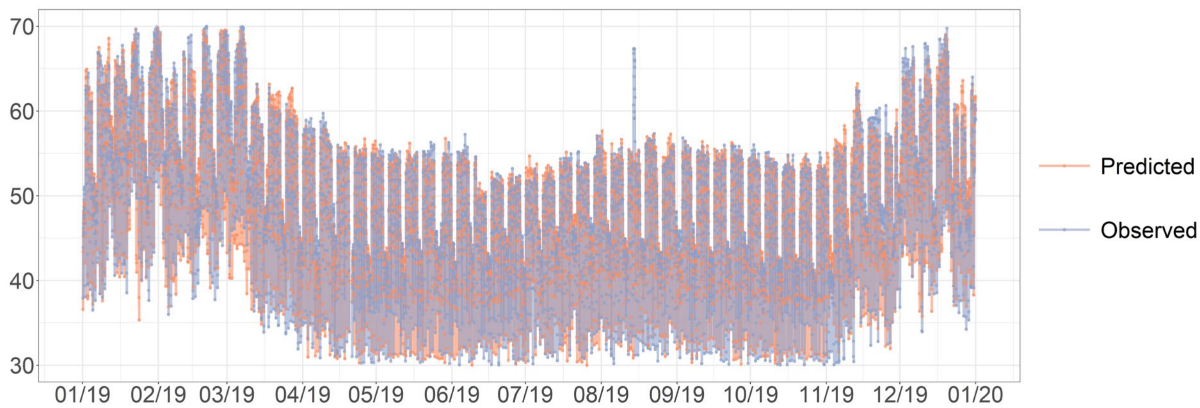


Figure 21: Linear regression model performance on train set and validation set

The research team tested the model on data from 2020. Based on the prediction over the year, a test RMSE of 2.7 MWh and test MAPE of 4.68% were obtained. Figure 22 shows that the prediction errors from May to December were greater than earlier in the year, indicating that there are some effects the model does not fully capture. Nevertheless, the model performed well in testing and was clearly robust to the disruptions that COVID-19 brought about.

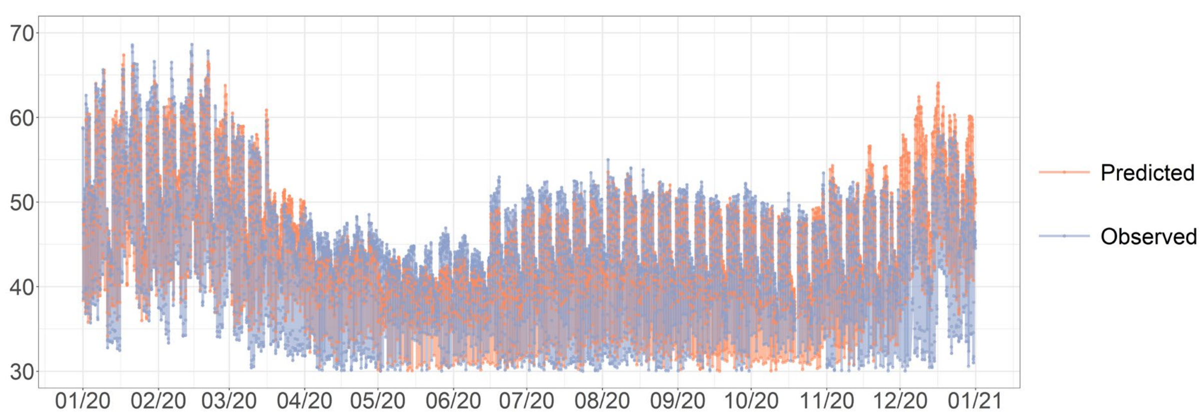


Figure 22: Linear regression model performance on test set

5.4.2 Random Forests

The random forests model was tested on the 2020 data set. The model performance is shown in Figure 23. Based on predictions in 2020, an RMSE of 2.94 MWh and MAPE of 5.01% were obtained. Random forests basically captured the effects in 2020, but the prediction errors from July to December were slightly higher than for other periods in 2020. Nevertheless, as seen from the residuals (Figure 24), both models performed similarly well in testing.

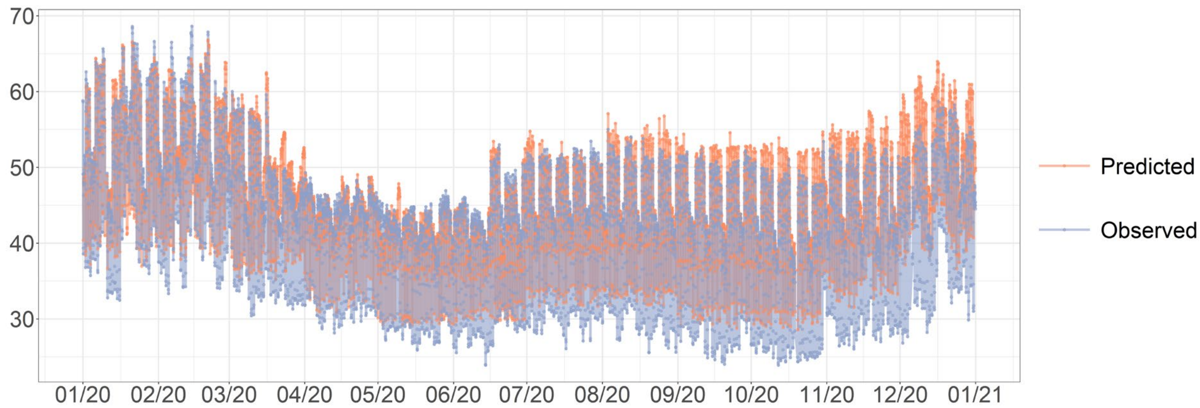


Figure 23: Random forests model performance on test set

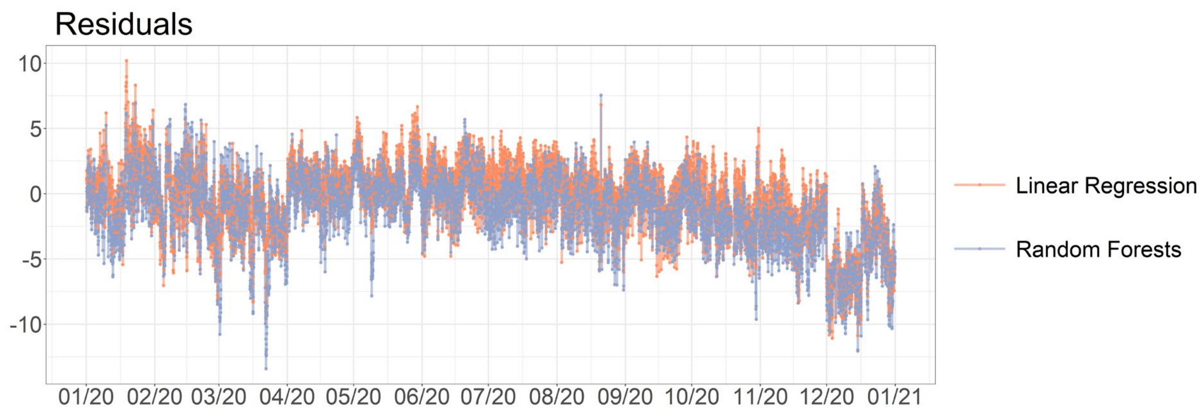


Figure 24: Residuals of linear regression and random forests

Table 8: Summary of model goodness of fit metrics, based on out-of-bag estimate

Model	**Training**		**Validation**		**Test**	
			RMSE (MWh)	MAPE (%)	RMSE (MWh)	MAPE (%)
Linear regression	R^2	0.93	3.32	3.32	2.70	4.68
Random forests	PVE	0.95	1.78	-	2.94	5.01

5.5 Impacts of COVID-19

COVID-19 was first reported in December 2019 and rapidly spread all over the world, reaching pandemic proportions in January 2020. It was declared a U.S. national emergency in March 2020. Subsequently, all nonessential travel was curtailed. Public places were closed, and working or schooling from home was imposed in many areas. Given the infectiousness of the disease, social distancing was imposed in many locales, and this significantly reduced transit usage across the United States. In response to these events and in order to cut costs as revenues declined, the MBTA reduced service on the Red, Orange, and Green lines by 20%, while reducing Blue Line service by 5% beginning March 14, 2020.

Later in the year, regular service was largely restored, even though ridership remained depressed throughout the rest of the year. Key metrics for the MBTA urban transit system in 2019 and 2020 are compared in Table 9. Ridership decreased by 66% from 150.3 million tap-ins in 2019 to 51.1 million tap-ins in 2020. This ridership decline is also shown in Figure 25(b). Compared to other metrics, ridership had the greatest change. It did not begin to climb to prior levels until the end of 2020.

The energy consumption in 2020 decreased by 7.6%, contributing to a cost decrease of 13.6%. The relatively low impact on energy demand, even though ridership declined so significantly, provided further evidence on the unimportance of ridership to overall energy consumption in the MBTA RTV system.

The service reductions resulted in declines in operating distance (vehicle-miles) and operating times (vehicle-hours): 5.7% and 13.3%, respectively. Regular MBTA rapid transit operations resumed in July 2020.

Table 9: Summary of COVID-19 impacts on MBTA

Year	**2019**	**2020**	**% change**
Cost (million $)	16.2	14	-13.60%
Energy consumption (GWh)	410.9	379.8	-7.60%
Vehicle-mile (million miles)	10.5	9.9	-5.70%
Vehicle-hour (million hours)	1.5	1.3	-13.30%
Ridership (million)	150.3	51.1	-66%
Energy per vehicle-mile (kWh/mile)	39.2	38.6	-1.50%
Energy per vehicle-hour (kWh/hour)	270.3	281.5	4.10%

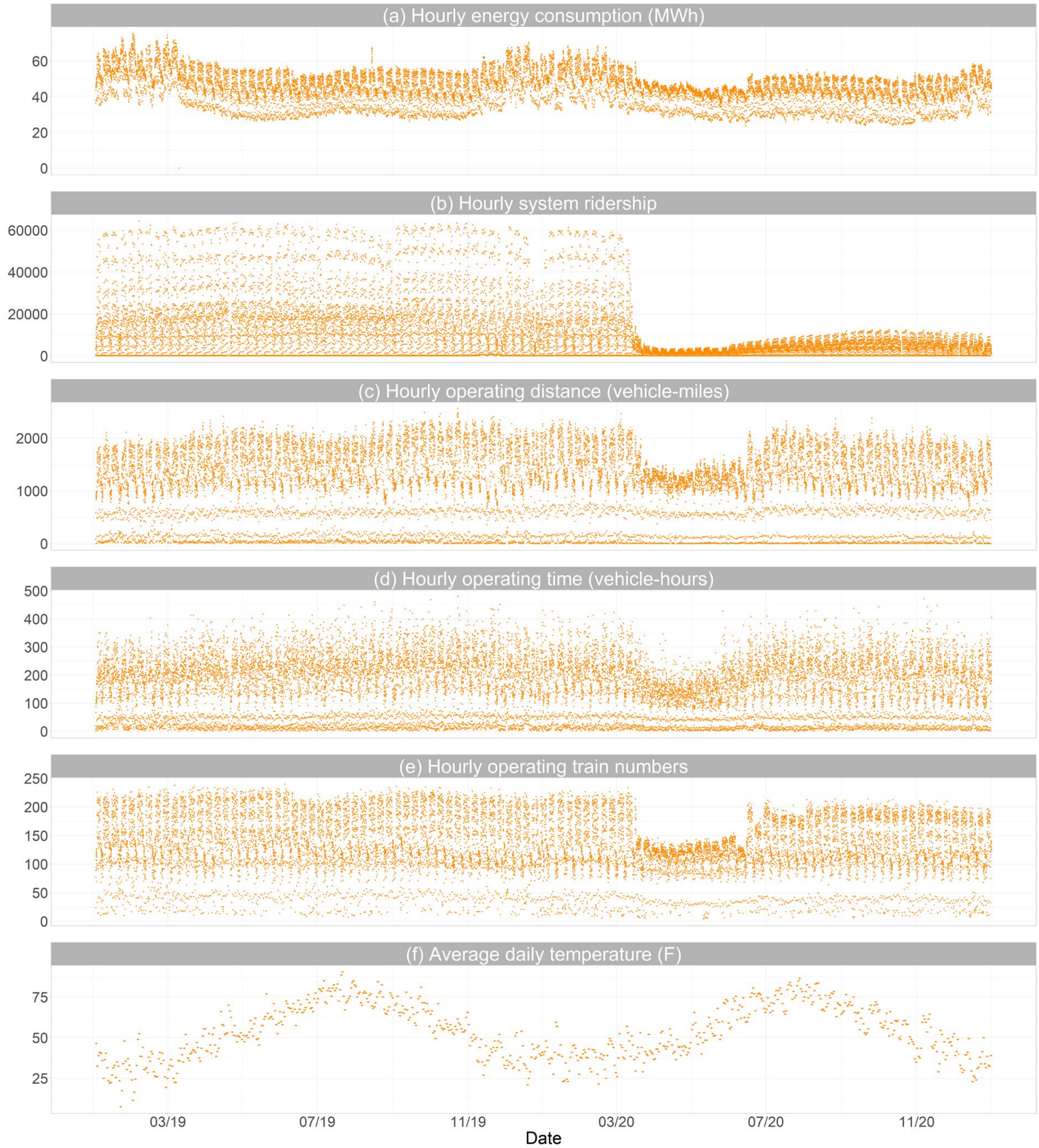


Figure 25: Time series plots of energy, ridership, vehicle-miles and vehicle-hours, Jan. 2019–Dec. 2020, for MBTA rapid transit system

This page left blank intentionally.

6.0 Conclusion

The TREEM project utilized data from train movement and operations, ridership, and ambient temperature to accurately predict and explain system-wide electricity consumption in the Massachusetts Bay Transportation Authority rapid transit vehicle network. First, the project team developed an integrated framework for data processing and trajectory computation. A proposed trajectory dashboard visualized train trajectory variables (distance, time, speed, acceleration), distributions of these variables, and a map showing the physical location of the given trajectory.

The project team then estimated a high-performance energy consumption multiple linear regression (MLR) model with an R^2 of 0.93, and random forests model (RF) with proportion of variance explained (PVE) of 0.95. On testing the model on 2020 data, these two models produced errors of less than 5.1%. The models provide insights into the driving factors of system energy consumption while also showing potential as a decision-support tool for future planning. The key drivers of the system-wide energy consumption were identified as temperature, baseline energy consumption by facility, and train operations, while ridership was found to have a very small impact on energy consumption. This is buttressed by the fact that the model predictions held up under COVID-19 disruptions.

One limitation of the current approach is that movement variables are aggregated without accounting for rail type, i.e., heavy, or light rail. Given the clear differences in speeds and vehicle mass between these two types, the models could be improved by computing these variables (such as bin times, train numbers, etc.) as type-specific or line-specific. Ongoing research efforts include equipping individual trains with accelerometers to calibrate physical models of electric train energy consumption using their high-resolution data. These models can then be upscaled for better system-wide energy explanation, and potentially shed more light on the contribution of other operational factors. In addition, the team plans to further develop the trajectory analysis tool into an online trajectory-energy dashboard. The dashboard can be used for real-time monitoring in order to pinpoint areas or vehicles in network with significant changes in energy use patterns.

Ultimately, while the models estimated have shed more light on the relevant variables for energy consumption, they would be most impactful as high-level decision-support tools that can guide planning efforts in light of future budgetary constraints or in response to disruptive events. By implementing learning procedures to map these low-level movement and operation variables to high-level planning metrics, generative processes can be estimated to produce valid synthetic data in response to any proposed policy response. Thus, future strategies can be readily assessed for energy and costs impacts and save valuable resources by providing reliable estimates to guide decision choices.

This page left blank intentionally.

7.0 References

1. ITF Transport Outlook 2019. https://www.oecd-ilibrary.org/transport/itf-transport-outlook-2019_transp_outlook-en-2019-en. Accessed July 8, 2020.
2. Higuera, A., L. Claudio, L. Vela, L. Hernandez, and J. Valdez. Energy Performance Analysis in an Electrical Subway Traction System. *IEEE Lat. Am. Trans.*, Vol. 14, No. 2, Feb. 2016, pp. 729–736. doi: 10.1109/TLA.2016.7437216
3. Mao, F., Z. Mao, and K. Yu. The Modeling and Simulation of DC Traction Power Supply Network for Urban Rail Transit Based on Simulink. *J. Phys. Conf. Ser.*, Vol. 1087, Sept. 2018, p. 042058. doi: 10.1088/1742-6596/1087/4/042058
4. Simulink Documentation. https://www.mathworks.com/help/simulink/. Accessed July 8, 2020.
5. Ruigang, S., Y. Tianchen, Y. Jian, and H. Hao. Simulation of Braking Energy Recovery for the Metro Vehicles Based on the Traction Experiment System. *Simulation*, Aug. 2017. doi: 10.1177/0037549717726146
6. Su, S., T. Tang, and Y. Wang. Evaluation of Strategies to Reducing Traction Energy Consumption of Metro Systems Using an Optimal Train Control Simulation Model. 2016. doi: 10.3390/en9020105
7. Wang, J. and H. A. Rakha. Electric Train Energy Consumption Modeling. *Appl. Energy*, Vol. 193, May 2017, pp. 346–355. doi: 10.1016/j.apenergy.2017.02.058
8. Sanchis, I. V., and P. S. Zuriaga. An Energy-efficient Metro Speed Profiles for Energy Savings: Application to the Valencia Metro. *Transp. Res. Procedia*, Vol. 18, Jan. 2016, pp. 226–233. doi: 10.1016/j.trpro.2016.12.031
9. Leung, P. C. M., and E. W. M. Lee. Estimation of Electrical Power Consumption in Subway Station Design by Intelligent Approach. *Appl. Energy*, Vol. 101, Jan. 2013, pp. 634–643. doi: 10.1016/j.apenergy.2012.07.017
10. Arikan, Y., and E. Cam. Assessment of Factors of Energy Consumption in Railways with the AHP Method. *ResearchGate*. https://www.researchgate.net/publication/322274321_Assessment_of_Factors_of_Energy_Consumption_in_Railways_with_the_AHP_Method. Accessed April 17, 2020.
11. De Andrade, C. E. S., and M. de A. D’Agosto. Energy Use and Carbon Dioxide Emissions Assessment in the Lifecycle of Passenger Rail Systems: The case of the Rio de Janeiro Metro. *J. Clean. Prod.*, Vol. 126, July 2016, pp. 526–536. doi: 10.1016/j.jclepro.2016.03.094
12. Wang, J., A. Ghanem, H. Rakha, and J. Du. A Rail Transit Simulation System for Multi-modal Energy-efficient Routing Applications. *Int. J. Sustain. Transp.*, Vol. 0, No. 0, Mar. 2020, pp. 1–16. doi: 10.1080/15568318.2020.1718809
13. Fernández, P. M., C. G. Román, and R. I. Franco. Modelling Electric Trains' Energy Consumption Using Neural Networks. *Transp. Res. Procedia*, Vol. 18, Jan. 2016, pp. 59–65. doi: 10.1016/j.trpro.2016.12.008
14. Modi, S., J. Bhattacharya, and P. Basak. Estimation of Energy Consumption of Electric Vehicles Using Deep Convolutional Neural Network to Reduce Driver’s Range Anxiety. *ISA Trans.*, Vol. 98, Mar. 2020, pp. 454–470. doi: 10.1016/j.isatra.2019.08.055

15. Malikopoulos, A., and J. P. Aguilar. Optimization of Driving Styles for Fuel Economy Improvement. Oak Ridge National Lab (ORNL), Oak Ridge, TN, National Transportation Research Center, Jan. 2012. https://www.osti.gov/biblio/1045844-optimization-driving-styles-fuel-economy-improvement. Accessed Apr. 17, 2020.
16. Oettich, S., T. Albrecht, and S. Scholz. Improvements of Energy Efficiency of Urban Rapid Rail Systems. *Urban Transp. X*, 2004, p. 10.
17. Li, Y., Q. He, X. Luo, Y. Zhang, and L. Dong. Calculation of Life-cycle Greenhouse Gas Emissions of Urban Rail Transit Systems: A case study of Shanghai Metro. *Resour. Conserv. Recycl.*, Vol. 128, Jan. 2018, pp. 451–457. doi: 10.1016/j.resconrec.2016.03.007
18. Zhang, L., R. Long, and H. Chen. Carbon Emission Reduction Potential of Urban Rail Transit in China Based on Electricity Consumption Structure. *Resour. Conserv. Recycl.*, Vol. 142, Mar. 2019, pp. 113–121. doi: 10.1016/j.resconrec.2018.11.019

8.0 Appendix: Prototype Trajectory Dashboards

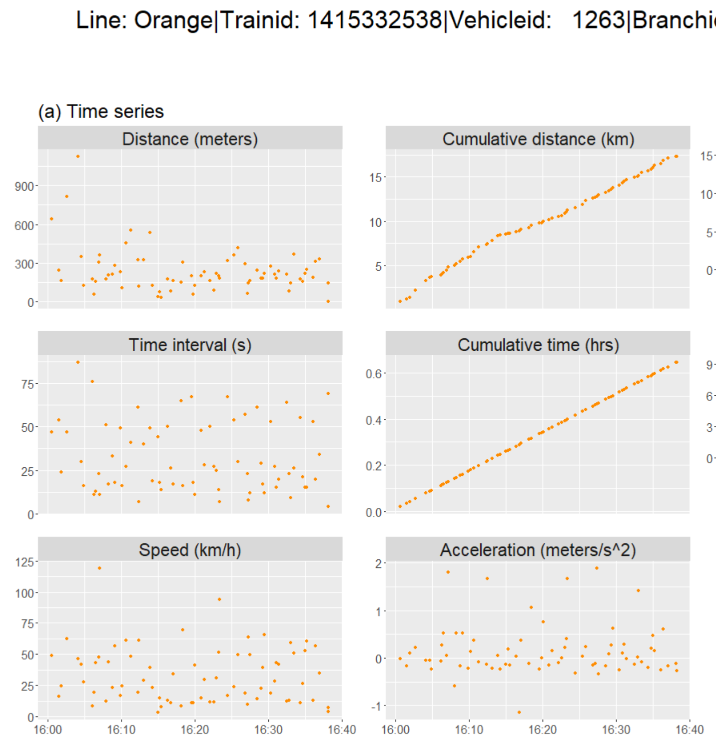


Figure 26: Orange Line dashboard example

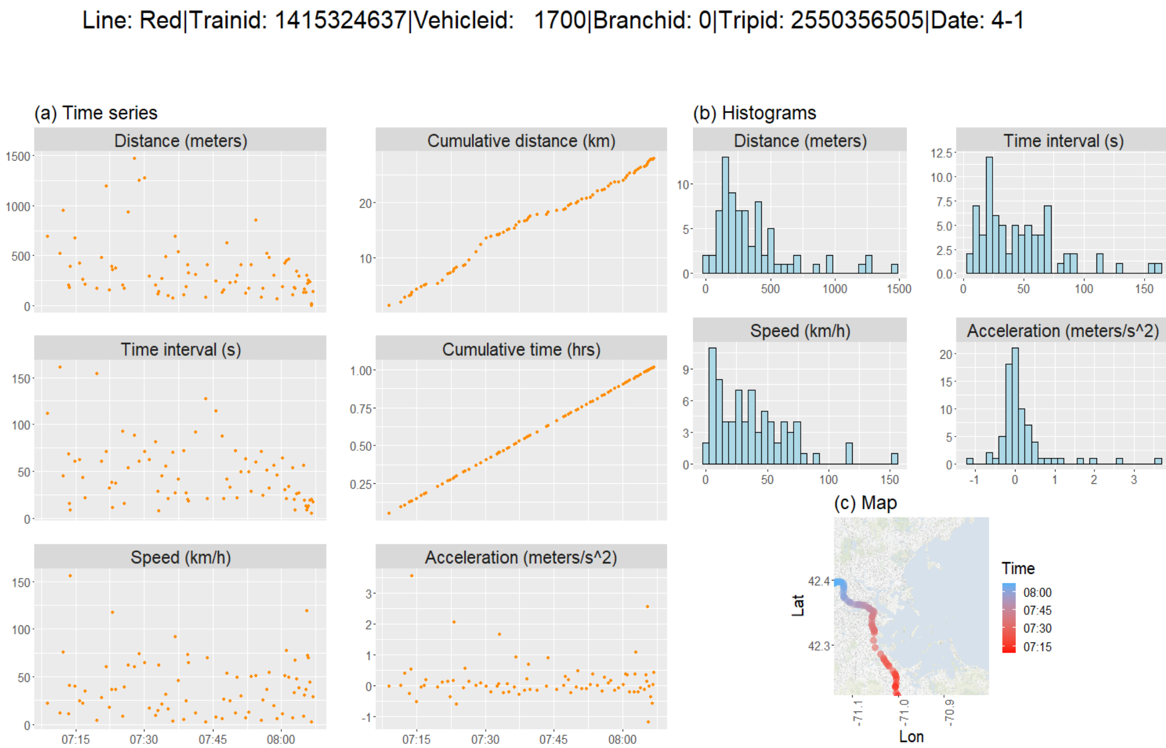


Figure 27: Red Line dashboard example

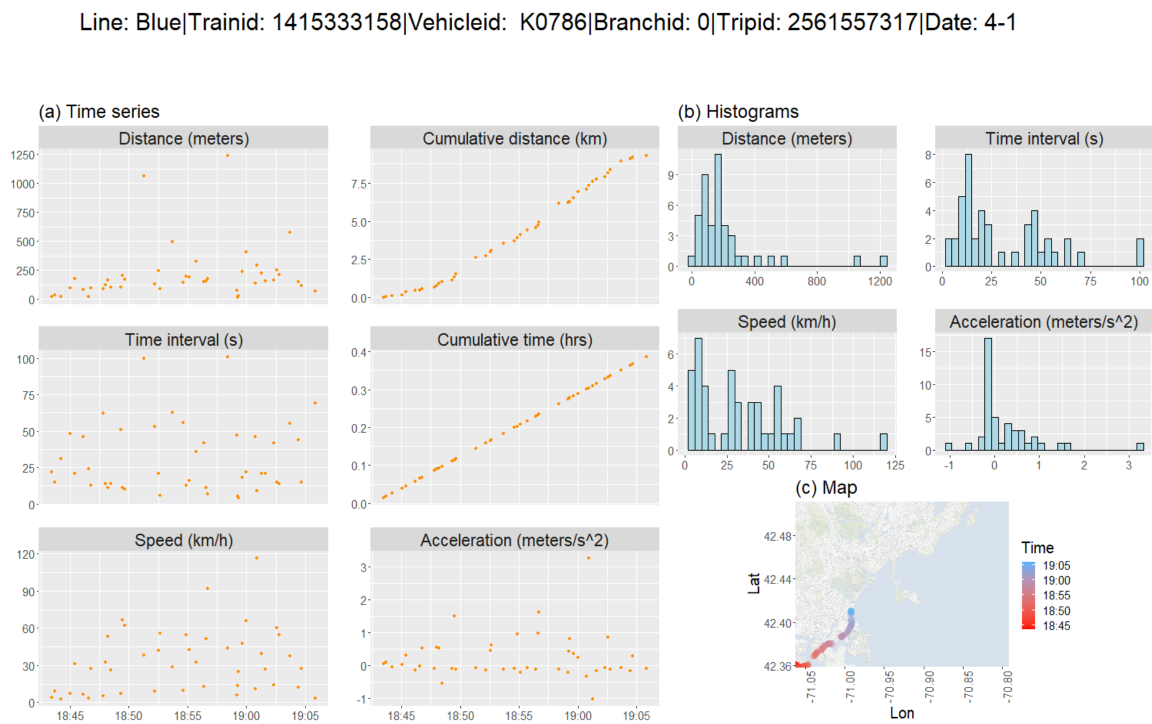


Figure 28: Blue Line dashboard example