

Estimating the Energy and Emissions Impacts of a Commuter Rail System in North Carolina

By

Zhengqi Jiao, MEM – Energy and Environment

Miaojun Pang, MEM – Energy and Environment

Chia-Shen (Jia-Shen) Tsai, MEM – Environmental Economics and Policy

Xinyi (Wendy) Wen, MEM – Energy and Environment & MEng

Dr. Timothy Johnson, Co-advisor

Alexander Yoshizumi, Co-advisor

*Masters project submitted in partial fulfillment of the requirements for the Master of Environmental
Management degree in the Nicholas School of the Environment of Duke University*

April 26, 2024

Executive Summary

The Transportation sector stands as the largest contributor to greenhouse gases (GHGs) emissions in the United States, representing approximately 36% of all emissions nationwide and 29% in North Carolina specifically. Within North Carolina, on-road light-duty gasoline vehicles alone accounted for 72% of the total GHG emissions from the Transportation sector in 2018. When considering emission reduction per passenger, public transit has a critical role to play. While commuter rail presents itself as a viable option for improving emission efficiency in North Carolina, its implementation poses significant challenges. Track-based systems like commuter rail require substantial funding for the development of physical infrastructure, including tracks, stations, and other facilities, as well as securing the right of way. This presents a considerable risk of sunk costs in the event of project abandonment or failure to meet ridership expectations. Given these challenges, decision-making bodies in the region, such as metropolitan planning organizations and county governments, must be well-informed about the emission performance of potential transportation investments.

We have developed a spreadsheet model capable of estimating the tank-to-wheel energy consumption and GHGs emissions of a commuter rail system. This model allows users to input parameters such as desired schedules, fuel types, and train weights, enabling them to compare and determine best practices. Its flexibility in rapid reparameterization and transparency enables users to assess how changes in those factors affect energy consumption and emissions. Our model is centered around the Greater Triangle Commuter Rail project, which spans highly populated areas from Durham County to Wake County in North Carolina. We selected this region for our pilot analysis due to the comprehensive feasibility study conducted and the potential for realization by 2040, particularly because it leverages existing intercity rail corridors. In our model, we have included diesel-electric and hydrogen fuel cell options as the primary fuel types, omitting electric trains due to their higher cost associated with the significant infrastructure required.

To determine the energy required and estimate emissions, we calculated the necessary energy as the initial step. The energy led to the fuel required, enabling us to estimate emissions by applying the emission factor. We adopted a microscopic consumption-based method to estimate the kinetic energy needed to move a train at a 50-foot resolution level. This analysis involved breaking down the energy into resistances encountered by the train, including aerodynamic, rolling, acceleration, and gradient resistances. Four essential elements — the train's speed, acceleration, weight, and gradient along the route — were crucial to determine the resistances. We developed a machine learning model to simulate speed based on empirical running data for intercity rail in the same corridor. The locomotive and train car weights were determined from other rail systems and literature. We also allocated GoTriangle's ridership estimates to different proposed schedules between each station. The gradient analysis utilized the 3-foot grid scale DEM data from NCDOT, characterizing the entire route with ArcGIS.

After compiling all the data, we inputted them into a Google spreadsheet to estimate the resistances, energy consumption, and consequently, CO₂ emissions for every 50-foot segment along the route. To enable a preliminary comparison with the road transportation emission, we calculated the emissions from single-occupancy vehicles using the same passenger-mile metric as the train. We extrapolated the average fuel

economy of passenger vehicles by the year 2040 based on both current performance and the estimated improvements mandated by regulations. The comparison assumes that passengers are either taking a train or driving a car alone. Due to rail's fixed route nature that led to longer travel distance, we also apply a distance adjustment factor to the road emission calculations.

The Results and Discussion section illustrates the model's application by varying passenger numbers or fuel technologies while keeping one set of parameters constant. We explored the emission reductions achievable by transitioning to hydrogen locomotives from diesel and analyzed the ridership needed for commuter rail to ensure carbon emission reductions compared to road transportation. Two key conclusions emerge:

1. The efficiency of hydrogen locomotives results in even the dirtiest hydrogen utilization emitting fewer greenhouse gases than the cleanest diesel as the fuel, with a threefold higher locomotive efficiency.
2. There exist specific thresholds for commuter rail to effectively reduce emissions by replacing single-occupancy vehicles. Notably, in the most realistic scenario, the hydrogen threshold aligns with the moderate levels of ridership, while the diesel threshold requires high levels of ridership to achieve reductions in emissions per passenger when compared to single-occupancy vehicles.

These findings underscore the significance of public support and strategic planning, including policies like transit-oriented development, in expanding commuter rail systems. In summary, the modeling results highlight the necessary conditions for the commuter rail system to achieve emissions reductions, which can serve to guide decision-making and policy formulation. The existence of the model provides a framework for decision-makers to identify key factors influencing emission levels in a proposed commuter rail system and enables evidence-based goal-setting for ancillary policies towards a decarbonized transportation future.

Table of Content

Executive Summary	1
1. Introduction.....	5
1.1 Background.....	5
1.2 Motivation.....	6
1.3 Area of Interest	6
2. Methods	9
2.1 Literature Review	9
2.2 Overview of the Microscopic Model for the Emission Estimation	10
2.3 Data Collection and Compilation	13
2.3.1 Train Data	14
2.3.2 Ridership Data	15
2.3.3 Speed data.....	18
2.3.3.1 Data Preparation	18
2.3.3.2 Random Forest Model Setting and Result.....	19
2.3.4 Emission Factors.....	21
2.4 Scenario Modeling Development	22
2.4.1 Train Parameters.....	22
2.4.2 Trip Parameters.....	22
2.4.2.1 Weekday Total Flow.....	22
2.4.2.2 Service Schedule and Peak/Off-peak Ratio.....	22
2.4.3 Station.....	23
3. Results and Discussions.....	23
3.1 Compositions of Driving Resistance	23
3.2 Emissions from Different Fuel Types.....	25
3.3 Emissions Comparison with Road Transportation	26
4. Limitation & Future Work.....	28
Acknowledgement	30
Appendices	31
Appendix 1: Existing Emission Estimation Models.....	31
Appendix 2: Typical Value of Rolling and Aerodynamic Resistances Coefficients.....	31
Appendix 3: Model Sheet Structure	32

Appendix 4: Variables for the Speed Machine-Learning Model.....	34
Appendix 5: Scripts for Data Processing.....	35
References.....	36

Figures

Figure 1. Route map of the proposed Greater Triangle Commuter Rail.....	8
Figure 2. Entity Relationship Diagram of the model.....	14
Figure 3. Modeling area with trip information from on-board transit surveys and CTPP JTW data.	15
Figure 4. Weekday passenger flow and ridership for peak-hour trips within the 8-2-8-2 schedule..	17
Figure 5. Average speed calculated from empirical data for each 0.1 mile on the Amtrak service using the H-line.	18
Figure 6. Distribution of minimal depth and its mean of each variable in the Random Forest model for the speed estimation.....	20
Figure 7. The estimated operating speed along the GTCR.....	20
Figure 8. The composition of driving resistance for every two adjacent stations.....	24
Figure 9. The daily GHG emissions of GTCR service utilizing several fuel types within the 8-2-8-2 schedule.....	25
Figure 10. The comparison of GHG emissions from GTCR service and road transportation (single- occupancy passenger vehicle) by passenger number.....	27
Figure A1. The modeling spreadsheets and their attributes for the microscopic emission model.....	32
Figure A2. The lookup tables and their attributes for the microscopic emission model.	33

1. Introduction

1.1 Background

Making up 11.2% of the global total in 2022, the United States is the second largest source of greenhouse gas (GHG) emissions following China (European Commission, 2023). The transportation sector is one of the largest sources of GHG emissions (~29% of total emissions) in the United States, with more than 80% coming from passenger vehicles and trucks on the road. While the rate of population growth is projected to slow over time at the national level (Congressional Budget Office, 2024), some geographies are expected to experience dramatic increases in population. At the state level, the North Carolina Office of State Budget and Management projects that the State will add 3.5 million people between 2021 and 2050 (NC OSBM, 2022). Zooming in further, forecasts by GoTriangle, a regional transit authority in North Carolina, project that over one million of those 3.5 million individuals will inhabit the Triangle region by 2050, potentially leading to a substantial increase in vehicles on the road (Duncan, 2023). Given the mounting pressures of climate change, it is critical to focus on reducing emissions from the transportation sector within the United States.

Advocating the utilization of public transit systems emerges as a compelling emission reduction strategy, primarily due to the typically lower per-passenger emissions compared to private vehicles. Additionally, this approach offers dual benefits: enhancing energy efficiency and alleviating traffic congestion. Nevertheless, private automobiles are still a very common mode of transportation in North Carolina. Statistics show that 85.9% of people commute by private car (Gomes, 2022). The Triangle's transportation infrastructure does offer alternatives to private vehicle commuting, with services provided by various agencies like GoTriangle, which is responsible for regional bus services, and plans for future transit expansions. Each county within the Triangle is gradually expanding their respective public transportation services. For example, according to GoTriangle fiscal year 2023 Annual Report (2023), four Bus Rapid Transit corridors are being developed by Wake County's Transit Plan. This can help Wake County better streamline traffic and shorten people's time on the road. In addition, GoTriangle and GoDurham leveraged funds from the Durham County Transportation Plan provided people with faster and more convenient buses through 2023, especially projects like transit emphasis corridors, bus speed and reliability projects and more (GoTriangle, 2023). However, the challenges posed by growing population are ongoing. Together, Wake, Durham, Orange, and Johnston counties are projected to experience an annual increase of over 32,000 residents, leading to congested roads and continuously extending commuting times (GoForward, 2023). In order to manage the rapid population growth and urban expansion in the Triangle region of North Carolina, the entire region needs a robust regional transportation network to meet the challenges of high-speed growth to provide transportation options and ensure regional mobility (GoTriangle et al., 2022).

In North Carolina, intercity passenger rail has gained prominence as an important mode of public transportation, driven by factors such as urban expansion and population growth. Passenger rail service is available in 16 cities across North Carolina currently. Several major lines include the Piedmont and the Carolinian which run through major North Carolina cities such as Raleigh, Durham, Greensboro, High Point, Salisbury, and Charlotte (Mooneyham, 2023). Notably, the NC By Train service witnessed over

522,000 rail passengers in 2022, operating four daily round trips between Raleigh and Charlotte and extending services to the Northeast (NCDOT, 2023). This figure not only sets a new annual ridership record in the service's 32-year history but also underscores the escalating demand and need for increased investment in passenger rail infrastructure in North Carolina (NCDOT, 2023).

In addition to the intercity passenger rail services offered across the state, the proposed commuter rail services in the Triangle region could expand access to quick and reliable commuting options. Commuter rail offers residents greater opportunities to access nearby urban areas due to its numerous stops and higher operating frequency. As a result, compared to intercity passenger rail, it is more effective in reducing the pressures associated with population growth and peak-hour commuting. It can help NC residents stay away from the frustration of gridlock traffic and ensure a smoother way to work. Also, the commuter rail increases accessibility and opportunities by enhancing connectivity. It facilitates residents' travel and improves their quality of life (U.S. Government Accountability Office, 2022). In the end, the proposed commuter rail can serve as an economic development driver that improves investments in communities and attracts new businesses in the future. According to a survey conducted by GoTriangle in early 2023, business leaders and residents in the Triangle area have shown strong support for a proposed commuter rail project in the region (GoForward, 2023). With strategic planning, sufficient funding, and community support, the expansion of public transit systems can serve as a cornerstone in the transition towards a more sustainable and equitable transportation future, offering a practical solution to some of the most pressing environmental and urban challenges of our time.

1.2 Motivation

While the adoption of rail transport is often touted for its potential to lower transportation emissions on a per-passenger basis, this is not guaranteed. This complex interplay is underscored by a book titled *Accelerating Decarbonization of the U.S. Energy System* (2021), which emphasizes that in order for rail systems to net a positive impact on emission reductions, they must operate at ridership levels that effectively offset the energy and emissions associated with their infrastructure and operations. Furthermore, the extensive investment required for the development and expansion of rail infrastructure, along with its consequential impacts on land use, represent important considerations in the pursuit of maximizing public good through transportation projects. Therefore, a multifaceted approach is essential to navigate these challenges, and the integration of rail systems into broader urban planning and development strategies is crucial for leveraging these investments to facilitate not only improved mobility but also economic development, environmental conservation, and social equity.

1.3 Area of Interest

To support NC government officials, transportation agencies, and environmental regulatory bodies in strategic planning of commuter rail projects, we aimed to assess how varying operation and technology scenarios – such as different train schedule, ridership, and fuel types – can affect the energy and emission

impact of a commuter rail project. Our analysis focuses on the recently proposed **Greater Triangle Commuter Rail (GTCR)** project (GoForward, 2023). The GTCR is a commuter rail route that would run through Wake County and Durham County. The route would utilize the “H-Line” owned and managed by North Carolina Railroad Company (NCRR), sharing the corridor with the existing freight and intercity passenger rail service (i.e., two Amtrak trains, Piedmont and Carolinian) (RSG, 2022).

Starting 2019, a series of GTCR studies (e.g., stations, schedules, ridership, infrastructure) have been led by GoTriangle in partnership with NC Capital Area Metropolitan Planning Organization (CAMPO), Durham - Chapel Hill - Carrboro Metropolitan Planning Organization (DCHC), North Carolina Department of Transportation (NCDOT), NCRR, and Durham, Wake, and Johnston County (GoTriangle et al., 2022). To decide the implementation stages, in early 2023, GoTriangle has initiated a “public feedback campaign” featuring online surveys, community gatherings, and public forums in Durham, Cary, Raleigh, and Clayton – communities traversed by the proposed GTCR. The completion of the entire project is expected to span a timeframe of 10 to 12.5 years, i.e., starting in service during 2033 - 2035 (GoTriangle et al., 2022).

According to the latest Phase II feasibility study report published in December 2022, this 37-mile corridor (**Figure 1**) has 14 possible stations linking West Durham and Auburn, with potential expansion towards Clayton to the east, providing 12,000 to 18,000 daily trips (GoForward, 2023). The stations (from west to east) we considered in this study included West Durham (WDU), Downtown Durham (DDU), East Durham (EDU), Ellis Road (ELI), Research Triangle Park (RTP), Morrisville (MOR), Downtown Cary (DCA), Corporate Center Drive (CCD), Blue Ridge Road (BRR), North Carolina State University (NCS), Downtown Raleigh (RAL), Hammond Road (HAM), Garner (GAR), and Auburn (AUB). For scoping reasons, Clayton was not included in this analysis, but it could be considered in future analyses.

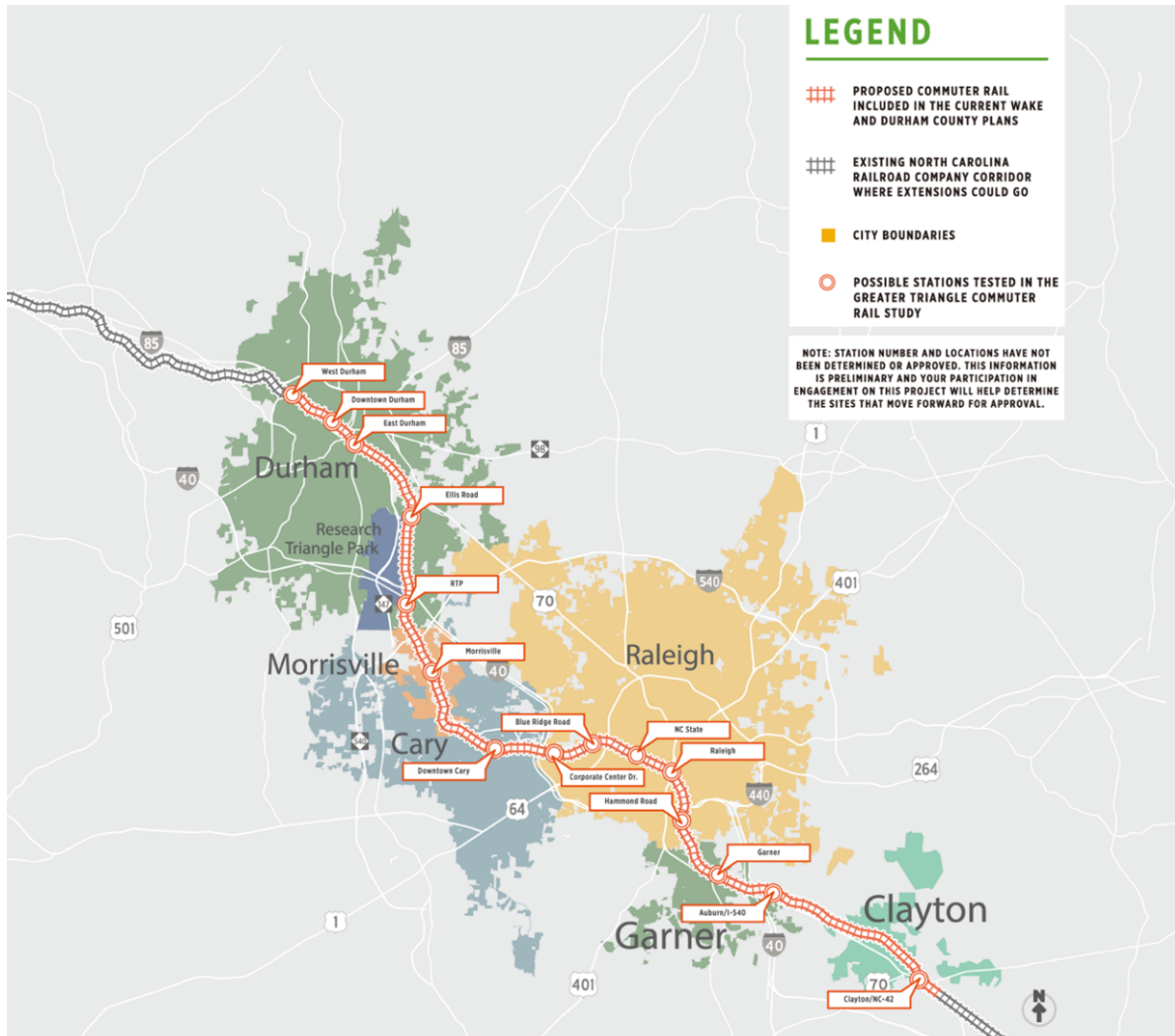


Figure 1. Route map of the proposed Greater Triangle Commuter Rail.

The proposed stations from west to east are West Durham (WDU), Downtown Durham (DDU), East Durham (EDU), Ellis Road (ELI), Research Triangle Park (RTP), Morrisville (MOR), Downtown Cary (DCA), Corporate Center Drive (CCD), Blue Ridge Road (BRR), North Carolina State University (NCS), Downtown Raleigh (RAL), Hammond Road (HAM), Garner (GAR), Auburn (AUB), and East Clayton (ECL) (Go Forward, 2023).

To provide the aforementioned stakeholders with precise and actionable insights, we have developed a detailed **spreadsheet model** to quantify the **energy consumption** and **GHG emissions** (carbon dioxide equivalent/CO₂-eq) of the **GTCR service under varying technology implementation scenarios**. This tool is designed to offer quick and accurate estimates of the environmental impacts associated with various operational configurations of the GTCR. It enables local government officials, state transportation authorities, and environmental regulatory agencies tasked with urban planning, transportation development, and emission control to make informed decisions. Although the current model only includes one commuter rail service, its framework and database are designed to be easily extensible to other routes or transportation

projects. This scalability and extensibility aids in the formulation of policies and strategies that balance the need for efficient transportation with environmental and sustainability objectives.

2. Methods

2.1 Literature Review

To identify critical technologies, services, and modeling approaches for estimating rail emissions, we examined the official reports from governmental and business sources, reputable news outlets, and the academic literature focusing on locomotive technologies, fuel types, route service strategies, and existing rail models. We also explored various methodologies for quantifying energy consumption and GHGs originating from rail operations.

One of the foundational elements of our review was an analysis of the existing emission estimation models tailored to rail transportation. The summary table of existing models we examined is in Appendix 1. These models varied substantially across several dimensions, including their geographical focus, rail type coverage (e.g., high-speed, intercity, commuter), fuel types considered, the life cycle of fuel (well-to-tank/WTT, and tank-to-wheels/TTW), the flexibility for user-defined inputs versus reliance on default values, and the types of emission quantified (e.g., CO₂, CH₄, N₂O, CO, NO_x, PM). This comparative analysis of existing models allowed us to understand the diverse parameters and assumptions underpinning energy and emission estimations in rail systems, providing insights about how our model can fit into the larger ecosystem of rail modeling.

In addition, we closely examined a review paper by Arne Heinold (2020), “Comparing emission estimation models for rail freight transportation.” From this review, we identified the model framework to adopt in our model: the microscopic model known as Assessment and Reliability of Transport Emission Models and Inventory Systems (ARTEMIS), which allows the most detailed data granularity and guarantees a more robust estimation. Given that we possess more granular data than what ARTEMIS typically requires, we have further simplified the calculation steps. The step-by-step equations derived from this process are illustrated in section 2.2.

2.2 Overview of the Microscopic Model for the Emission Estimation

Table 1. Notation of Trip and Train Parameters

Parameter	Description	Unit
m_{fuel}	Fuel weight	kg
u	Energy density	MJ/kg
k	Emission coefficient	kg·CO ₂ e/MMBtu
E	The anticipated energy necessary for propelling the vehicle along a given distance	J
F	The force needed to propel the train. The underscore represents the type of resistances.	Newton
d	Distance	m or km
h	Altitude difference	m
m_{loc}	Locomotive weight	kg or ton
m_{car}	Car weight	kg or ton
m_t	Total weight	kg or ton
v	Speed	m/s or km/h
c^{air}	Train specific aerodynamic coefficient	-
c^{roll}	Train specific rolling coefficient	-
c^{grade}	Downhill energy recovery coefficient	-
c^{accl}	Brake energy recovery coefficient	-
ρ	Air density (assumed 1.25 kg/m ³)	kg/m ³
A	Locomotive frontal area	m ²
g	Gravitational acceleration (assumed 9.8 m/s ²)	m/s ²
c_{loc}^{air}	Locomotive specific air resistance coefficient	-
c_{roll}^{air}	Car specific air resistance coefficient	-
c_{car}^{roll}	Locomotive specific rolling resistance coefficient	-
c_{car}^{roll}	Car specific rolling resistance coefficient	-
n_{car}	Number of cars	-
ϵ	Locomotive efficiency rate	-
R_f	Rate of fuel flow	kg/s
LHV	The lower heating value of the diesel (assumed 42700 kJ/kg)	kJ/kg

To initiate the development of our microscopic models, we adopt a consumption-based approach. We denote the amount of fossil fuel consumed by a vehicle during a transportation process as m_{fuel} . The emissions of GHG associated with this consumption are determined by multiplying m_{fuel} by the corresponding energy density u and emission coefficient k . This coefficient represents the amount of CO₂ equivalent emitted from combusting the fuel that produce a given unit of energy. MMBtu (million British thermal units) is one of the most prevalent units used to measure the emission factor in the literature. Therefore, we convert the energy unit using the constant 947.82. To standardize the unit measurement of

emissions across diesel and hydrogen, we use weight (kg) rather than volume (liter) to measure the amount of fuel. GHG emissions are subsequently calculated by the following equation:

$$GHG = m_{fuel} \cdot u \cdot 947.82 \cdot k \quad (1)$$

The fuel consumed is unknown since we do not possess the empirical data, and fuel consumption of any given trip may vary based on a wide array of factors. Thus, we estimate the fuel consumption by estimating the energy consumption:

$$E^{TTW} = m_{fuel} \cdot u \quad (2)$$

$$E^{TTW} = E^{kinetic} \cdot \frac{1}{\epsilon} \quad (3)$$

Here, E^{TTW} represents the anticipated tank-to-wheel chemical energy necessary for propelling the train along the track and $E^{kinetic}$ is the kinetic energy only taking physical principle into account. To account for the energy loss in the process of converting a fuel's chemical energy to meet a vehicle's kinetic energy needs, we apply the coefficient ϵ as the combined efficiency of the locomotive. The efficiency rate differs by locomotives, taking into account various factors such as the transmission, the engine, and the gearing, respectively. In this study, we use the value simulated by Lindgreen and Sorenson (2005), who validated their result with existing empirical diesel train data in Denmark. The rate for each component of the locomotive could be calculated with

$$\epsilon = \frac{Power}{R_f \cdot LHV} \quad (4)$$

where R_f is the rate of fuel flow and LHV denotes the lower heating value of the diesel.

As demonstrated later in this section, the estimation of $E^{kinetic}$ can be derived from the vehicle's attributes, route characteristics, and anticipated driving speed. Combining the equations above, the GHG emissions can be calculated by

$$GHG = \frac{E^{kinetic}}{\epsilon} \cdot 947.82 \cdot k \quad (5)$$

Given that emission coefficient (k) and locomotive efficiency rate (ϵ) are implemented as constants associated with the locomotive and its fuel type, the primary challenge associated with applying Equation (5) revolves around determining $E^{kinetic}$. We computed the energy consumption by breaking it down into the four primary resistances encountered during operations — rolling, aerodynamic, grade, and acceleration resistances — and multiplying them with the distance as shown in Equation (6). In our calculations of these energy demands, we largely adhere to the methodologies outlined by Lindgreen and Sorenson (2005).

$$E = \int \left(F_{roll} + F_{air} + F_{grade} + F_{accl} \right) \Delta d \quad (6)$$

To overcome air resistance, the required power F_{air} is calculated by

$$F_{air} = \frac{1}{2} \cdot c^{air} \cdot \rho \cdot A \cdot v^2 \quad (7)$$

where c^{air} denotes the drag coefficient (calculation is shown below), ρ is the air density and A is the frontal area of the locomotive, and v is the speed. The power F_{roll} that is required to overcome rolling resistance is calculated by

$$F_{roll} = c^{roll} \cdot m_t \cdot g \quad (8)$$

where c^{roll} is the rolling resistance coefficient (calculation is shown below), g is the gravitational acceleration, and m_t is the vehicle's total weight. To run uphill, power F_{grade} is computed by

$$F_{grade} = m_t \cdot g \cdot \frac{h}{d} \quad (9)$$

where h is the elevation difference between the given horizontal distance, d , making the last term the slope of the segment. Applying Newton's 2nd law, the last resistance, F_{accl} is the acceleration force that is required for increasing the train speed in a given time, which is computed by Equation (10). The a is the acceleration measured in speed change by a second (m/s^2).

$$F_{accl} = m_t \cdot a \quad (10)$$

Aside from the main component above, there are two coefficients that would change by the configuration and speed of the train. One is the air coefficient, the other is the rolling coefficient, both of which grow with increasing train length (Lindgreen & Sorenson, 2005). Thus, the c^{air} and the c^{roll} for a train are calculated by:

$$c^{air} = c_{loc}^{air} + c_{car}^{air} \cdot n_{car} \quad (11)$$

$$c^{roll} = c_{loc}^{roll} \cdot \frac{m_{loc}}{m_t} + c_{car}^{roll} \cdot \frac{m_{car} \cdot n_{car}}{10 \cdot m_t \cdot g} + c_1 \cdot \frac{\left[\frac{km}{h} \right]}{100} + c_2 \cdot \left(\frac{\left[\frac{km}{h} \right]}{100} \right)^2 \quad (12)$$

where n_{car} is the number of cars appended, whereas the m_{loc} and m_{car} denote the weight of locomotive and cars. The c_{loc}^{air} , c_{car}^{air} , c_{loc}^{roll} , c_{car}^{roll} , c_1 , c_2 are denoting the locomotive- and car-specific rolling and air resistance coefficient. They are typically treated as constants in a specific type of rail service (ibid. p.35). The typical values are given in the Appendix 2.

To address the energy recovered when a train run down the slope or brake, we add two other coefficients, c^{grade} and c^{accl} , to control the percent of recovery that only apply to negative gradient resistance and acceleration resistance values. The F_{grade-} and F_{accl-} denote the force values that are negative in gradient and acceleration resistances.

$$F_{grade}^{[N]} = F_{grade+} + c^{grade} \cdot F_{grade-} \quad (13)$$

$$F_{accl}^{[N]} = F_{accl+} + c^{accl} \cdot F_{accl-} \quad (14)$$

After obtaining all the resistances from Equation (7) to (14), we can simulate the energy required for every single segment by

$$E_{segment_i}^{kinetic} = (F_{air}^{[N]} + F_{roll}^{[N]} + F_{grade}^{[N]} + F_{accl}^{[N]})(x_2^{[m]} - x_1^{[m]}) \quad (15)$$

This is feasible due to the detailed route characteristics available in the scope of our study, which we analyzed at a spatial granularity of 50 feet (15.24 meters). The energy calculated by Equation (15) is then aggregated as described in Equation (4). With $E^{kinetic}$, we are able to calculate the GHG emissions using Equations (1), (2), and (3).

2.3 Data Collection and Compilation

As mentioned above, we broke down the route to **line segments** separated by a fixed distance of 50 feet. The speed and elevation change were given to each node and line at this granularity. The weight of the train, on the other hand, is the function of ridership and fuel. Thus, we created another dataset recording the ridership estimation for every single **trip** between the adjacent station. A **trip** is defined as each **interval** during which a train operates within a day. For example, a trip from Durham to Cary in the morning differs from the same journey in the afternoon due to variations in passenger numbers. By combining the line data (containing speed and slope information) with the trip data (containing weight information), ridership characteristics of the rail service and combining the coefficients and emission factors, we model daily emissions in the "Model_Line" sheet at the segment level. These emissions are then aggregated upwards to calculate emissions for each trip in the "Model_Trip" sheet. Each sheet represents the estimation for a full day's schedule. Additional constants, emission factors, and coefficients are organized in various lookup (LT) tables. The relationships between these components are illustrated in **Figure 2**, and a comprehensive list of variables for each sheet can be found in Appendix 3.

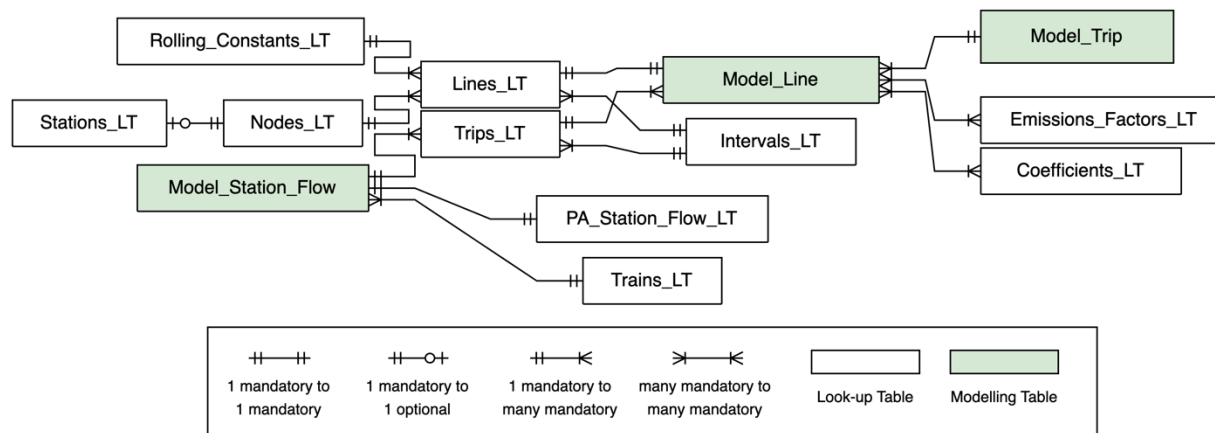


Figure 2. Entity Relationship Diagram of the model.

Plotted by the authors.

2.3.1 Train Data

Our analysis of commuter rail infrastructure was primarily informed by data from the "Greater Triangle Commuter Rail Feasibility Study: Phase II Summary Report" (GoTriangle et al., 2022). This study indicated that the proposed service would likely have nine trainsets, each equipped with four coach cars, totaling thirty-six coaches across the system. Both diesel locomotives and Diesel Multiple Units (DMU) were considered by the study authors. Different from diesel locomotives, DMUs do not require a separate locomotive. A DMU is a train set with two to four cars, powered by diesel engines located in the first or last cars, with the middle cars unpowered (*Diesel Multiple Unit*, 2024). Within the GTCR plan, each trainset is designed to offer a seating capacity of 600, and each trainset includes one locomotive and four coach cars. Cumulatively, this configuration results in a system-wide seating capacity of 5,400 seats (GoTriangle et al., 2022).

While no trainset and locomotive model have been decided upon yet, GoTriangle indicated that the new Siemens Charger could be a relevant locomotive to model. Consequently, for the purposes of this analysis, we modeled the Siemens ALC-42 Charger as our diesel-electric option. This model has a mass of 127 metric tons, with dimensions measuring 21.79 meters (71.5 feet) in length, 3.05 meters (10 feet) in width, and 4.5 meters (14.7 feet) in height (Siemens Mobility, Inc., 2023).

Additional data relevant to our study were sourced from external research and comparative analysis. According to the "Simulation of Energy Consumption and Emissions from Rail Traffic," the efficiency rate for diesel locomotives and electrical locomotives are 0.35 and 0.65 respectively (Lindgreen & Sorenson, 2005). To address gaps in our dataset for the GTCR, we drew parallels with the Coaster commuter rail service in San Diego County, California. The Coaster's route has a very similar length to the GTCR, with both services covering shorter commuter distances. Bilevel commuter rail cars are currently used on this line, recognized as the lightest (50 metric tons) and most economically beneficial option in North America

(Sklar, 2020). This rail car has great potential to become a suitable and efficient model. Therefore, we assumed a bilevel car type for our analysis.

2.3.2 Ridership Data

We used production-attraction flows corresponding to weekday ridership from GTCR Appendix H: Phase 2 Ridership Analysis Technical Memorandum (Table 80) to inform our ridership assumptions (RSG, 2022). This production-attraction (PA) matrix was transformed into the table in “PA_Station_Flow_LT” sheet within the model. Assuming that there will be 20 (8-2-8-2)¹ daily round trips between West Durham and Auburn, and the commuter rail fare is equal to the regional bus fare, the GTCR Report team estimated a weekday total flow of 12,183 in 2040 using the Simplified Trips-on-Project Software (STOPS) developed by Federal Transit Administration’s (FTA) with the 2018 ridership as the baseline (RSG, 2022). The 2018 data came from two sources: (1) three relatively recent transit rider surveys covering most of the corridor to be served by the GTCR project, and (2) Census Transportation Planning Products (CTPP) Journey-to-Work (JTW) flow data from the 2006-2010 Census American Community Survey (ACS) (**Figure 3**).

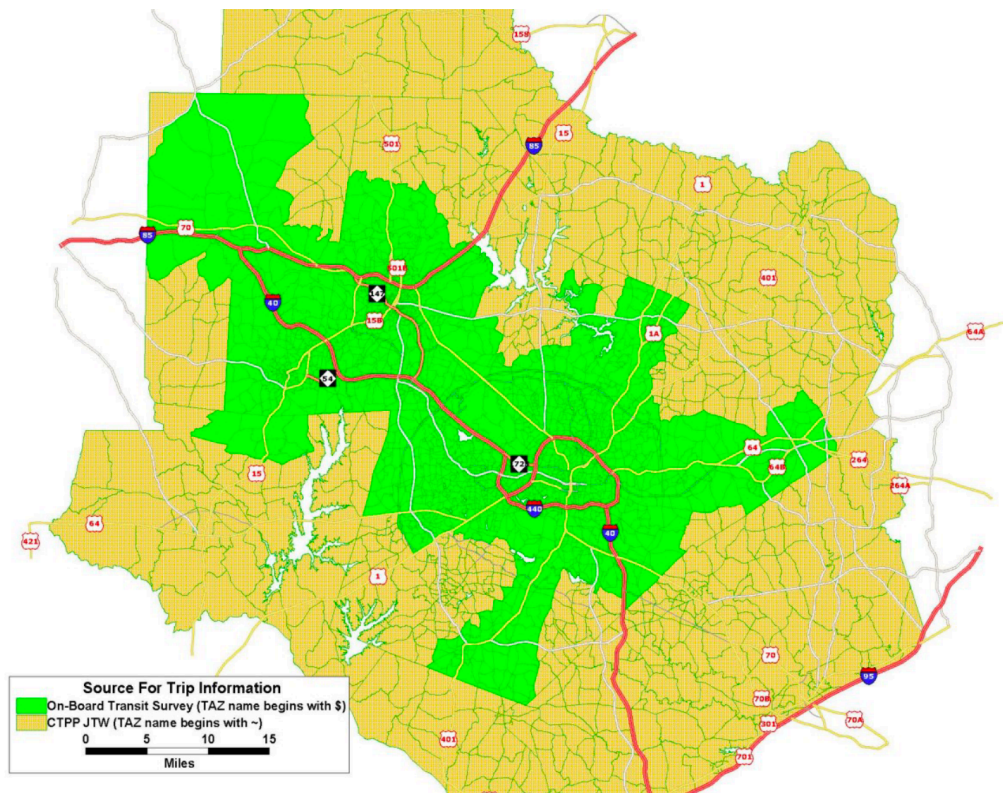


Figure 3. Modeling area with trip information from on-board transit surveys and CTPP JTW data.

Source: (RSG, 2022)

¹ 8-2-8-2 sequence entails 8 round trips during the morning peak period, 2 round trips in the midday period, 8 round trips during the evening peak period, and 2 round trips in the evening period.

To capture the station flow pattern, we first calculated the proportion of each station flow from the default daily total flow along the entire route, then the daily total station flow was obtained by multiplying the proportion with the user input flow value. The daily station flow of morning (PA) and evening (AP) trips during peak (PK) and off-peak (OP) periods were then calculated from the daily total station flow. Specifically, we split the total station flow into four components: (1) all morning trips during peak period (i.e., **morning peak trips, PA-PK**), (2) all morning trips during off-peak period (i.e., **midday trips PA-OP**), (3) all **evening peak trips (AP-PK)**, and (4) all **evening off-peak trips (AP-OP)**. Assuming all the passengers who travel from station A to station B in the morning would travel back from B to A in the evening, the four trip types were calculated using the equation:

$$Flow_{triptype} = Flow_{total} \cdot Proportion_{PK/OP} \cdot Proportion_{PA/AP} \quad (16)$$

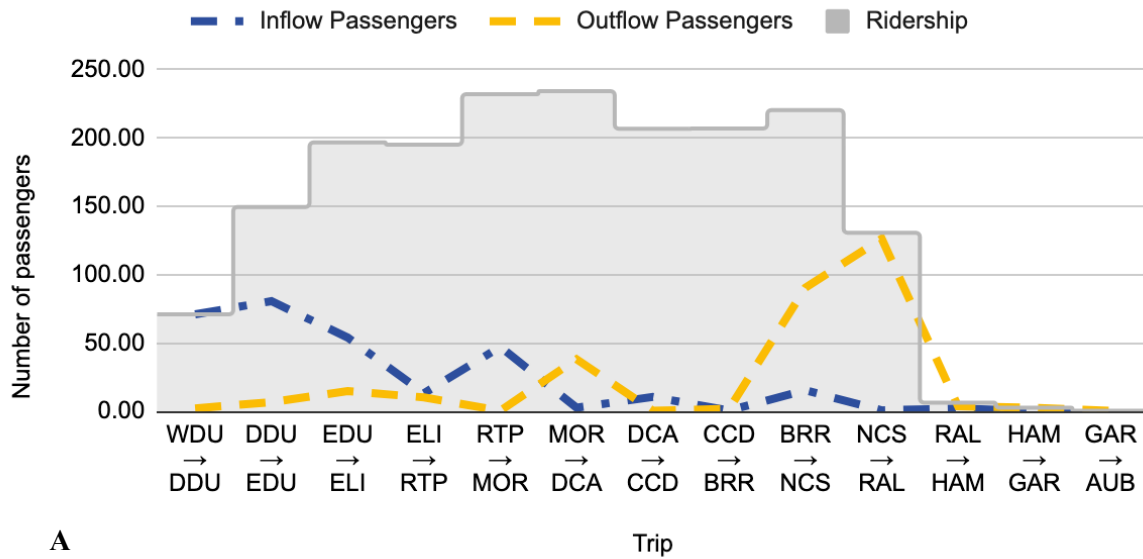
where $Proportion_{PA/AP} = 0.5$

The “Peak/Off-peak Proportion” is the proportion of ridership during peak hours and off-peak hours during a day which can be adjusted by users in the model dashboard (“Peak / Off-Peak Ratio”). The **daily station flows per train** were then obtained by dividing the total station flow during each period by **the number of trips** during each period.

Next, we derived the **inflow passengers at origin stations** and the **outflow passengers at destination stations** for every two adjacent stations (both directions, east and west bound) from the daily station flow per train. To obtain the inflow passengers for east (west) bound trips, we summed morning (PA)/evening (AP) flow for all stations east (west) of the current origin station using the peak (PK)/off-peak (OP) column. While for the outflow passengers for east (west) bound trips, we summed PA/AP flow for all stations west (east) of the current destination station using PK/OP column.

Finally, the **ridership** of GTCR between every two adjacent stations was calculated by summing the ridership of the previous trip and the inflow passengers at the current origin station, and then subtracting the outflow passengers at current destination station (**Figure 4**). To estimate the total passenger weight, we assumed the **average passenger weight** to be 79 kg (175 pounds), which aligns with the assumption adopted by most transit agencies and track component suppliers (Lin et al., 2016).

Eastbound Morning Peak Passenger Flow and Ridership (8-2-8-2)



Westbound Evening Peak Passenger Flow and Ridership (8-2-8-2)

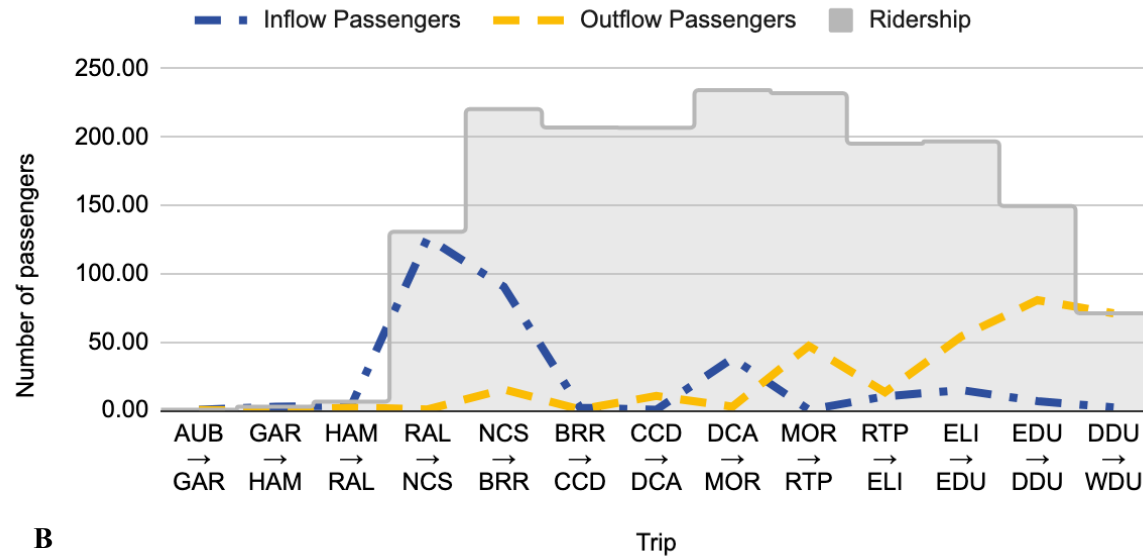


Figure 4. Weekday passenger flow and ridership for peak-hour trips within the 8-2-8-2 schedule. (A) Eastbound trips during morning peak hours. (B) Westbound trips during evening peak hours. Plotted by the authors.

2.3.3 Speed data

2.3.3.1 Data Preparation

To forecast the rail-running speed for the forthcoming commuter rail service between Auburn and Durham, we developed a machine-learning model using empirical data from the intercity rail service operating on the same corridor (Amtrak, n.d.). Through interviews with the client, GoTriangle staff, and NCDOT staff, we identified the primary factors associated with changes in train speed. These factors were narrowed down to five types: continuous characteristics such as curvature and gradient, as well as discrete features such as stations, crossings, and switches.

We utilized data observations from a broader geographic footprint - the entirety of the H-line - to train the model. This dataset encompasses the Piedmont/Carolinian and the Crescent routes operated jointly NCDOT's Rail Division and Amtrak. Our dependent variable, speed, was scraped from empirical intercity rail time-location data from an unofficial site that utilizes Amtrak's Track-A-Train service (Amtrak, n.d.). We collected data from January 24th to February 7th and from March 3rd to 9th in 2024. The data were structured in GTFS-RT² format, with updates occurring approximately every 45 seconds on average, with a maximum gap of 3 minutes between updates. To ensure data reliability, we validated the data by cross-referencing station locations and speed limits recorded in the NCDOT's schematic of the rail roads³(2017). The maximum empirical speed observed was approximately 80 mile per hour, which matched the maximum speed limit in the schematic. **Figure 5** shows the instantaneous speed of the 14,340 datapoints we collected.

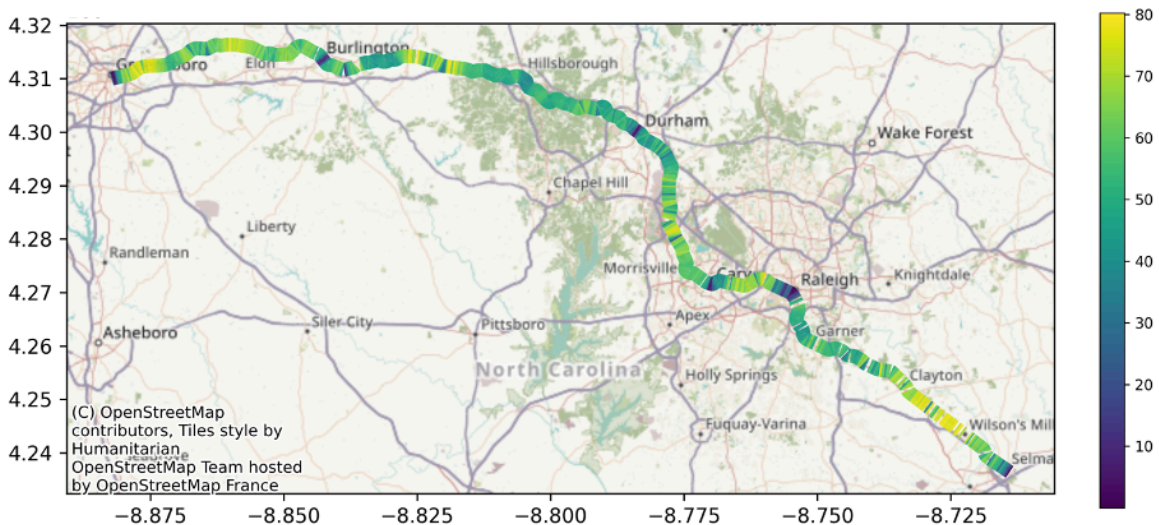


Figure 5. Average speed calculated from empirical data for each 0.1 mile on the Amtrak service using the H-line. Plotted by the authors. The dark segments are consecutively Greensboro, Burlington, Durham, Cary, Raleigh, and Selma-Smithfield, corresponding to the existing Amtrak intercity passenger rail stations.

² GTFS-RT (General Transit Feed Service – Real Time) is a widely-used format that allows public transportation agencies to provide real-time updates about their fleet to application developers.

³ This schematic provides information on speed limits, curvature, superelevation, and crossing details along the entire rail line, indexed by milepost. The schematic is sourced from NCDOT, courtesy of an email from Jon Dees, Rail Planning Consultant within NCDOT.

To characterize key explanatory variables along the route, we calculated the variables at 50-foot intervals along the entire track length. For each node's curve and slope features, we estimated the arc angle and elevation change between its preceding and succeeding nodes. To ensure a smooth profile and accommodate any sudden changes in the route, such as when crossing a river, we conducted a zonal analysis by averaging the elevations of the adjacent six points for each node. Additionally, to consider the timing of the operator's braking decision, we created variables that identify the maximum angles and slopes ahead within specified ranges at 2,500-foot intervals—approximately ½-mile - extending up to 10,000 feet (See the full list of variables fed in to the model in Appendix 4 and the processing code in Appendix 5).

For the crossings, we utilized the crossing inventory provided by the Office of Railroad Safety under the Federal Railroad Administration (2024) and the station list offered by the Bureau of Transportation Statistics (2023). We found the switches by identifying the milepost where there are two tracks on the H-Line existing track schematic provided by NCDOT (2017). After identifying which nodes correspond to discrete features that are or may be associated with changes in speed (e.g., crossings, stations, and switches), we calculated the distances to and from the following and preceding crossings, stations, and switches at each 50-foot interval node.

2.3.3.2 Random Forest Model Setting and Result

We used a random forest model instead of a linear regression model because random forest techniques tend to be robust to multicollinearity across the explanatory variables and do not have parametric assumptions about the distributions of model variables. Random Forest reduces overfitting and handles complex relationships in data better than traditional regression models, which can be limited by assumptions like linearity (Breiman, 2001). Additionally, Random Forest provides more robust predictions, especially in the presence of noisy data or outliers, making it a preferred choice for many predictive modeling tasks where accuracy and generalization are crucial (Liaw & Wiener, 2002). Moreover, it only includes the number of variables in the random subset at each node and the number of trees in the forest as a parameter, thus rendering it easier to use than other models.

We trained the model using a range of five hundred to four thousand trees and utilized nine variables in the subset at each node. The percentage of variance explained vary from 68.61% to 68.73% across the models. We thus use the 500-tree model to save the processing time for future generalization. The final model explained 68.67% of the variance in the test dataset, with a mean square residual of 142.74. The most impactful variables were the distance to and from the station, which makes sense as these directly correspond with the train coming to a stop or starting from a stop. The other top variables all had similar mean minimal depths, further illustrating the importance of the distance to and from stations as a predictor of speed (**Figure 6**).

After we trained the model, we applied the model to a dataset that included the hypothetical stations associated with the station plans detailed within the GTCR project documentation. The simulation result is shown in **Figure 7**.

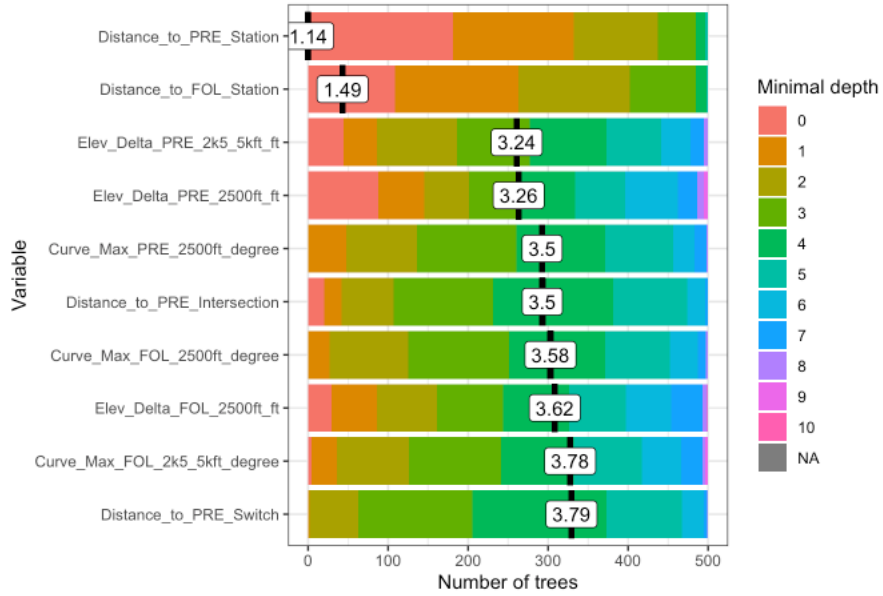


Figure 6. Distribution of minimal depth and its mean of each variable in the Random Forest model for the speed estimation.

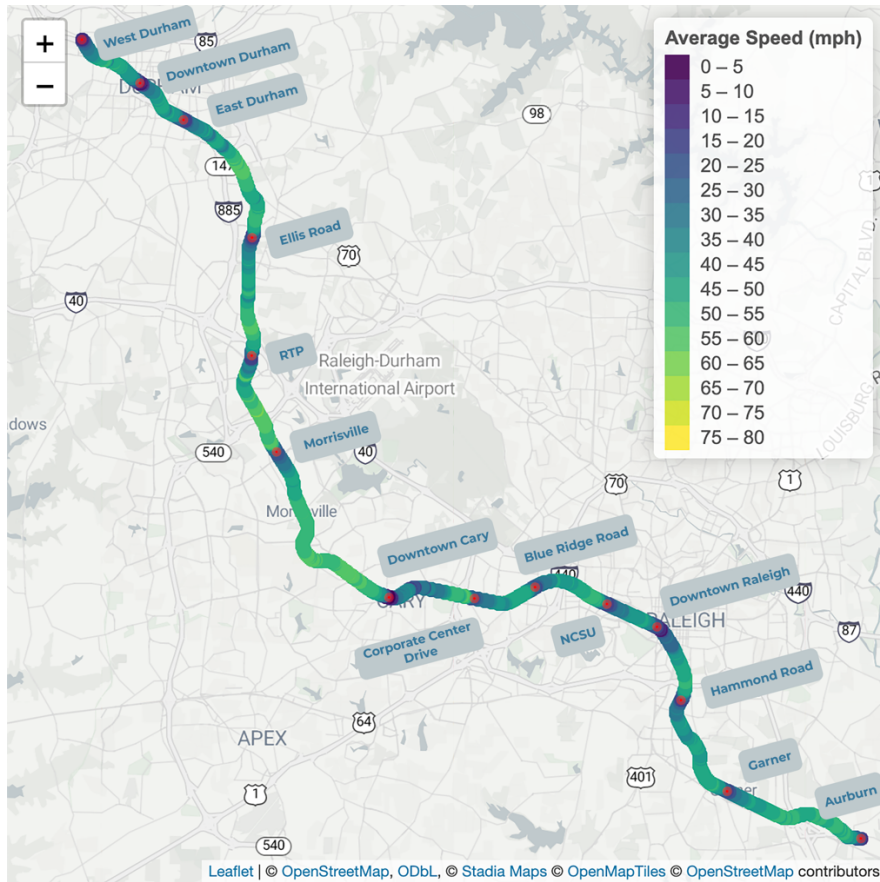


Figure 7. The estimated operating speed along the GTCR. Plotted by the authors. The red dots are where the proposed stations are.

2.3.4 Emission Factors

The model allows users to choose from two locomotive types: **hybrid diesel** and **hydrogen (H₂) fuel cell** locomotives. The model can return GHG emissions of the selected trip in the form of **CO₂-eq** which is calculated from the emission factors of three dominant GHGs – CO₂, CH₄ (methane), and N₂O (nitrous oxide) – and their global warming potential in IPCC Fifth Assessment Report (AR5) (Greenhouse Gas Protocol, 2014). The emission factors for diesel fuels and H₂ were retrieved from the latest (2023) **GREET 1 model** (Argonne National Laboratory, 2023), and the fuel life cycle considered was well-to-wheel (WTW), i.e., the emissions associated with fuel extraction, distribution, and combustion. Detailed information related to fuels with links to corresponding data sources can be found in the “Emissions_Factors_LT” sheet of our model.

Diesel locomotives, which have historically dominated the rail system, rely on internal combustion engines. The primary emissions from diesel locomotives include CO₂, NO_x, particulate matter (PM), and sulfur dioxide (SO₂). The emission factor for diesel locomotives varies based on the engine type, age, operational practices, and maintenance but is generally high due to the combustion of fossil fuels (EPA, 2009). Current passenger train services in North Carolina employ **diesel-electric (or hybrid diesel) locomotives** which use a diesel engine to generate electricity that powers traction motors (NCDOT, 2020). While they share similar profiles with traditional diesel engines, improvements in engine efficiency and emissions control technologies have led to reduced GHG emissions. However, they still emit substantial amounts of NO_x and PM.

On the other hand, **H₂ fuel cell** technology offers a promising alternative for commuter trains since it produces zero tailpipe emission. This makes it an attractive option for reducing the environmental impact of transportation. However, the deployment of hydrogen-powered trains entails the development of new infrastructure, from production and distribution to refueling stations (El-Shafie et al., 2019). More importantly, the environmental benefits of hydrogen fuel depend substantially on the production methods used (Pahwa & Pahwa, 2014). Currently, the majority of hydrogen is generated through steam reforming of methane (grey H₂), a process that, while effective, produces substantial carbon emissions (Hydrogen Council, 2021). Yellow H₂ is produced from water electrolysis powered by a mix of renewable and fossil fuel energy sources. The emissions factor of yellow H₂ depends on the emissions intensity of the grid. This emissions intensity currently remains high in many energy service territories but is expected to decrease substantially over the coming decades (Enerdata, 2023). By contrast, electrolysis can achieve much lower emissions when powered exclusively by renewable energy sources (green H₂) or nuclear energy (pink H₂) (Hydrogen Council, 2021). Therefore, a clean, sustainable transition to hydrogen fuel cell technology in the rail sector not only requires the development of new fueling infrastructure but also hinges on the adoption of sustainable hydrogen production methods (EFI Foundation, 2023).

2.4 Scenario Modeling Development

The current [spreadsheet model](#)⁴ is presented in a Google Sheet format, with its comprehensive structure depicted in the diagrams found in Appendix 3. In brief, the model features two primary tabs: "Dashboard" and "Outputs." These tabs allow users to create scenarios and view results respectively. Subsequent tabs contain internal calculations and lookup tables essential for the model's functionality. Below, the principal adjustable parameters in the Dashboard are described in detail.

2.4.1 Train Parameters

Users have the option to select between two types of locomotives: hybrid diesel and hydrogen fuel cells. The hybrid diesel locomotive can be fueled with either low-sulfur diesel⁵ or biodiesel (B20)⁶. In the case of the hydrogen fuel cell locomotive, users can evaluate the emissions of four hydrogen types as detailed in section 2.3.4: grey, yellow, pink, and green hydrogen. Additionally, users can modify the number of locomotives and railcars to assess how train weight influences energy consumption and emissions.

2.4.2 Trip Parameters

2.4.2.1 Weekday Total Flow

In the model, the default total weekday GTCR flow from West Durham to Auburn is set at 12,183 as explained in section 2.3.2. However, projected ridership is predicted to fluctuate from 10,000 to 18,000 depending on the fare scenario (RSG, 2022). Consequently, we have made this a user-configurable variable, labeled "Weekday Total," on the model dashboard.

2.4.2.2 Service Schedule and Peak/Off-peak Ratio

Users can choose one of the two service scenarios: 8P-2O-8P-2O (8-2-8-2) and 3P-3P (3-3). The 8-2-8-2 scenario consists of 8 trips during the morning peak, 2 trips during midday, followed by 8 trips during the evening peak and 2 additional trips in the evening. The 3-3 scenario comprises three peak trips in both the morning and evening. In addition, the ratio of ridership during peak hours and off-peak hours is adjustable, while the ratios of some existing commuter rail services in the U.S. range from 0.66:0.34 to 0.81:0.19 (Division of Strategic Planning & Performance, 2022; Humphrey, 2012). If the 3P-3P scenario is chosen, the peak and off-peak ratio should be adjusted to 1, as this scenario does not include any off-peak trips.

⁴ The [spreadsheet model](#) is maintained by the Applied Data Research Institute.

⁵ Diesel fuel currently available for off-highway (or non-road) use in the United States is ultra-low sulfur diesel (ULSD), containing 15 parts per million or less of sulfur (EIA, 2023).

⁶ Biodiesel is available in several different mix ratios, including B100 (100% biodiesel), B20 (a mix of 20% biodiesel and 80% petroleum diesel), B5, and B2, with B20 being the most frequently used blend in the United States (DOE Vehicle Technologies Program, 2011).

2.4.3 Station

Besides returning the energy consumption and GHG emissions of an entire GTCR roundtrip between West Durham and Auburn (14 stations), the model also allows users to investigate the energy and emissions of travel from specific origin and destination stations.

3. Results and Discussions

We have developed an analytical framework that integrates the data collected above, drawing connections between the dependent variables (i.e., energy consumption and emissions), the intermediate variables (e.g., ridership and system efficiency), and user-specified input parameters.

With the framework, our spreadsheet model allows users to build their own scenarios by adjusting parameters related to fuel technologies and rail service strategies. With different combinations of parameters, users may assess the sensitivity of energy consumption and emissions to changes in factors such as fuel type, train frequency, and route coverage. For the road emission comparison, we also allow users to specify the fuel economy (mile per gallon, MPG) of light-duty-vehicle (LDV) in North Carolina in 2040 and the distance adjustment factor. This distance adjustment factor is a number between 0-1 that allows the model to accommodate the typically shorter commuting distances associated with road transportation.

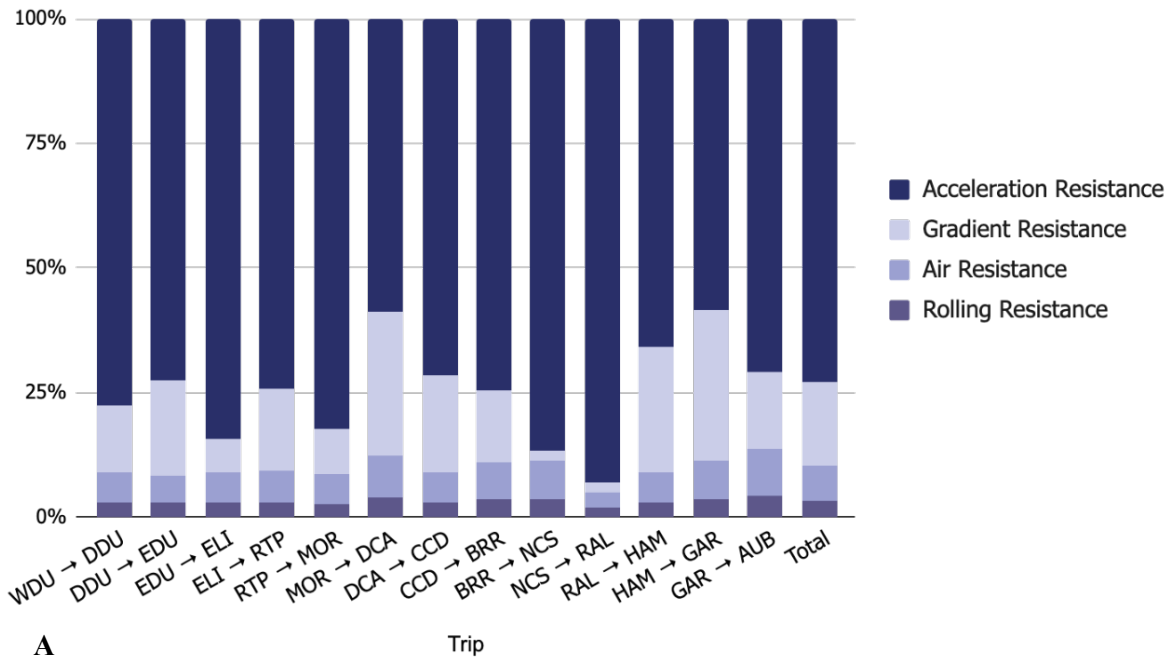
In the following results, we implement an 8-2-8-2 schedule, with the trainset consisting of 1 locomotive and 4 railcars. The peak and off-peak passenger ratio is 0.85:0.15, prioritizing more passengers during peak times. Both the slope coefficient and deceleration coefficient are 0.001 to take the energy recovery from downhill coasting and brake operation into account.

The model enables users to determine: 1) the amount of emission reduction achieved by using a hydrogen locomotive compared to a diesel locomotive, and 2) the number of passengers required to guarantee emission reduction via commuter rail compared to single-occupancy vehicles with the same number of commuters. In essence, the model is designed to facilitate rapid reparameterization for users to compare emissions across different fuel types and transportation modes.

3.1 Compositions of Driving Resistance

Using a physical approach to estimate energy consumption renders it fixed for a specific route and operational timetable. The proportion of each resistance is shown in **Figure 8**. When analyzing the resistances encountered during eastbound trips, it is observed that both air and rolling resistance are relatively minor contributors. The predominant factors affecting the train's performance is acceleration resistance. For westbound trips, while the acceleration resistance is still the primary contributor to the overall resistance, the gradient resistance plays a more dominant role compared to eastbound trips since the westbound trips consistently involve an uphill journey.

Eastbound Trips Resistance Composition



Westbound Trips Resistance Composition

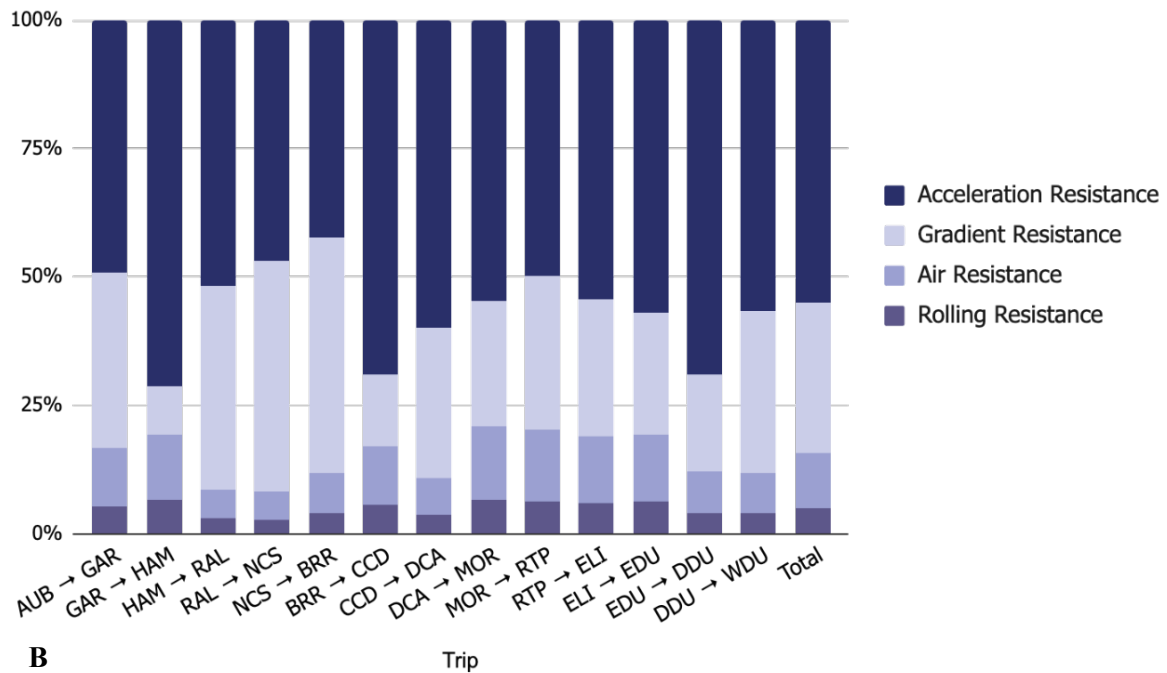


Figure 8. The composition of driving resistance for every two adjacent stations.

(A) Eastbound trips from West Durham to Auburn. (B) Westbound trips from Auburn to West Durham. Plotted by the authors.

3.2 Emissions from Different Fuel Types

Figure 9 illustrates the daily GHG emissions for the GTCR service utilizing various fuels within the 8-2-8-2 schedule (20 roundtrips). Our analysis indicates that the least environmentally friendly hydrogen option (grey H₂) offers a 3% reduction in emissions compared to the most eco-conscious diesel option (B20 biodiesel) (Panel A). This is attributed to the fact that, despite grey H₂ having a higher emission factor, H₂-powered locomotives achieve double the efficiency of their diesel counterparts. Notably, when contrasting more eco-friendly yellow H₂ against low-sulfur diesel, emissions are cut by 54% (Panel B).

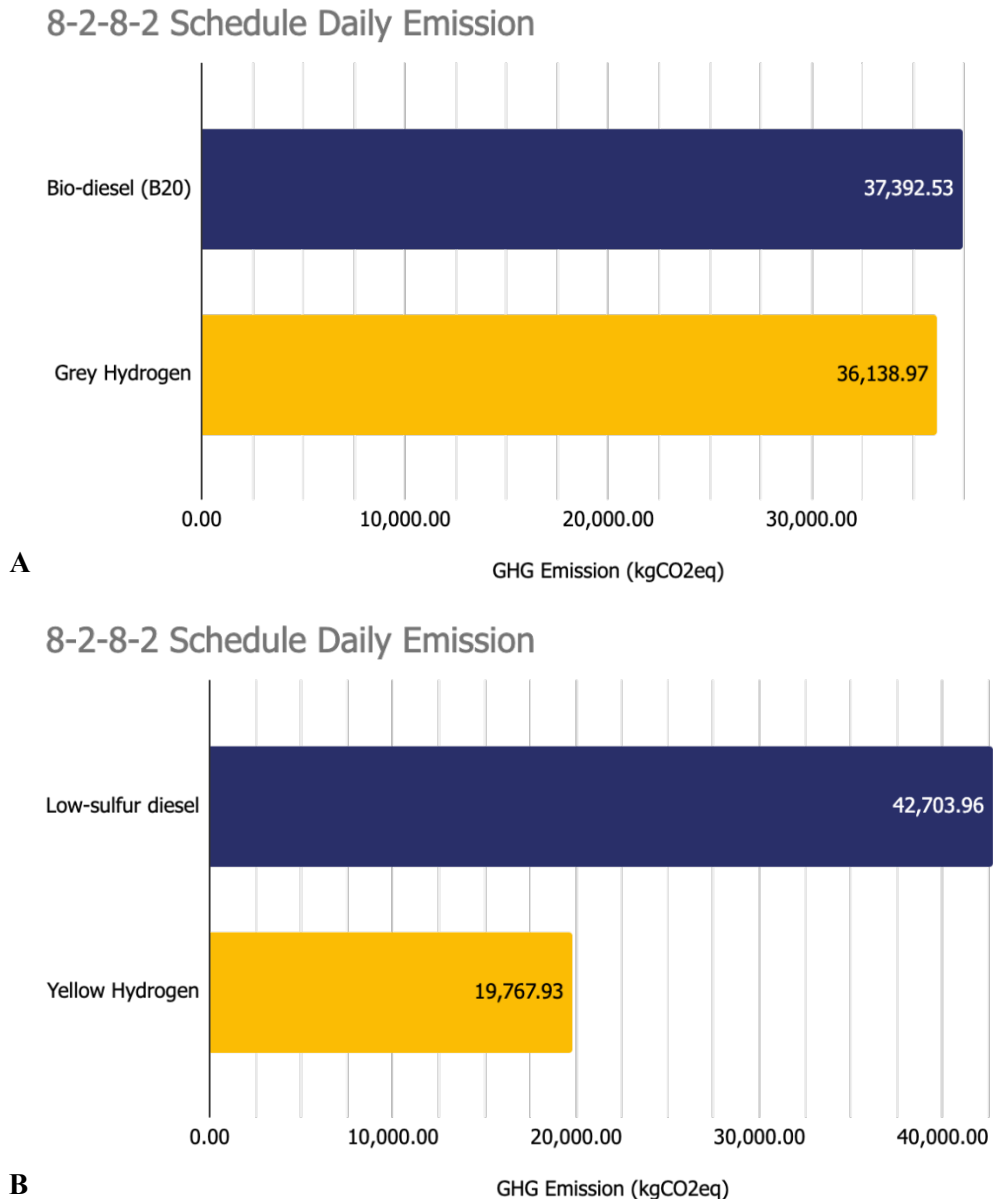


Figure 9. The daily GHG emissions of GTCR service utilizing several fuel types within the 8-2-8-2 schedule. (A) The comparison between the biodiesel (B20) and the grey hydrogen. (B) The comparison between the low-sulfur diesel and the yellow hydrogen. Plotted by the authors.

In short, the transition from diesel to hydrogen fuel cell locomotives can result in a substantial reduction in GHG emissions. While there remain technical and economic challenges related to hydrogen fuel adoption, the long-term environmental benefits of hydrogen fuel cells make them an attractive option for rail service providers interested in delivering passenger rail services with a reduced GHG footprint. Strategic investments in hydrogen infrastructure and technology development, supported by policy incentives, will be critical to realizing the potential of hydrogen fuel cell locomotives.

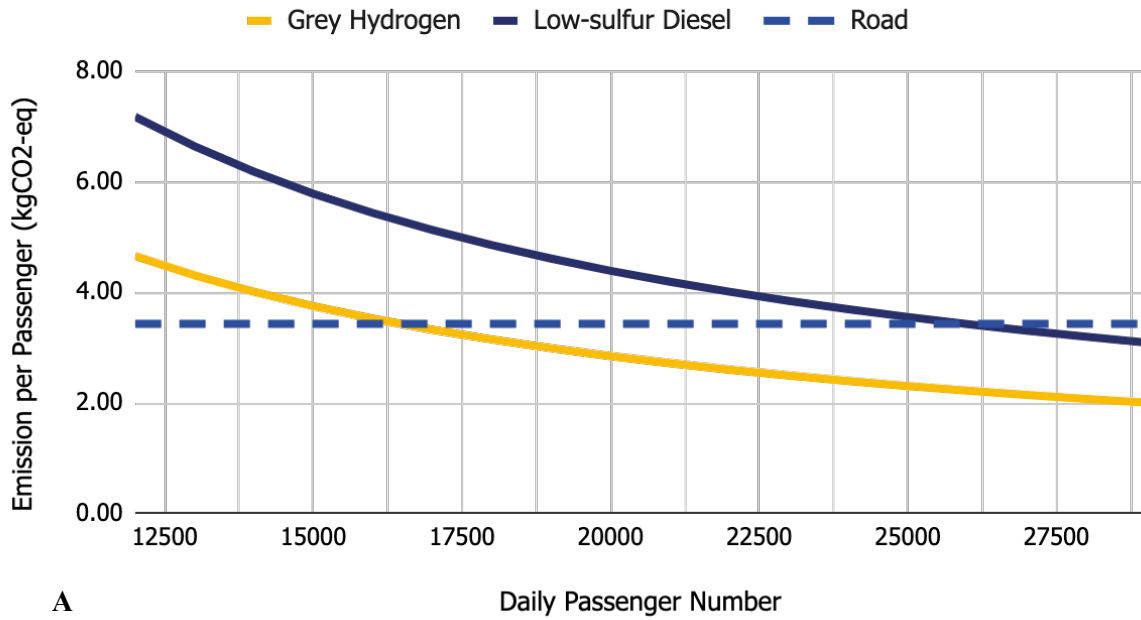
3.3 Emissions Comparison with Road Transportation

To assess the environmental benefits of commuter rail relative to single-occupancy vehicles, we considered the number of vehicle miles travelled offset by rail passengers. We assume the rail passengers are all single-occupancy vehicle drivers. We estimate the emissions per passenger of road commute by dividing the discounted passenger-mile with an assumed average fuel economy for vehicles in North Carolina, which stands at 25.6 miles per gallon⁷. The inherent energy consumption of operating a train means that a higher passenger count is required to distribute the emissions more efficiently on a per capita basis. In contrast, emissions per passenger for a single-occupancy vehicle remain consistent, as they do not benefit from shared energy consumption.

Figure 10 (Panel A) indicates that commuter rail emissions per passenger decrease as ridership increases, reaching a point of environmental advantage over single-occupancy passenger vehicles at a ridership level of approximately 16,500 passengers if the locomotive is powered by grey H₂. The point of environmental advantage shifts to 26,000 passengers when utilizing low-sulfur diesel fuel. Figure 11 (Panel B) further illustrates this point, indicating that without sufficient ridership, commuter rail may not always present a lower-emission alternative to road transportation by single-occupancy vehicles.

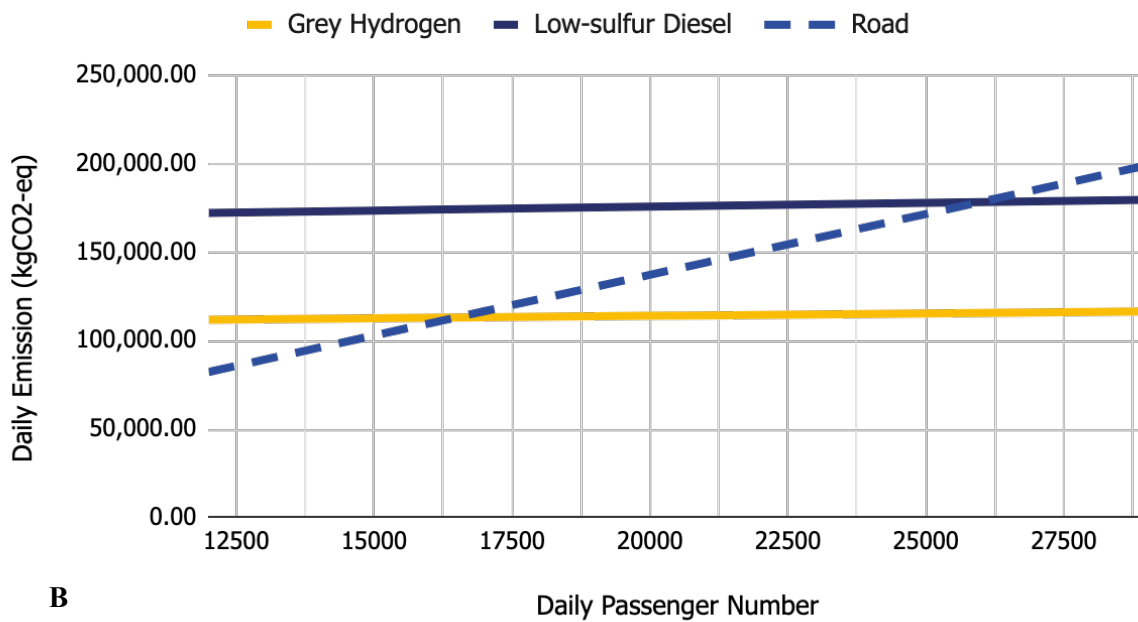
⁷ We extrapolate the average mpg number by applying the estimated growth in fuel economy in National scale (U.S. Energy Information Administration, 2024) to the average mpg in North Carolina in 2018 (Vehicle Technologies Office, 2021).

Emission Per Passenger of Rail and Road Transportation



A

Daily Emission of Rail and Road Transportation



B

Figure 10. The comparison of GHG emissions from GTCR service and road transportation (single-occupancy passenger vehicle) by passenger number.

(A) Emission per passenger. (B) Total emission per day. Plotted by the authors.

Overall, these analyses underscore the critical role of strategic planning in the expansion of commuter rail systems. In summary, the modeling results emphasize the essential conditions for commuter rail systems to meet in reducing energy consumption and emissions, underscoring the need for thorough analysis to inform decision-making and policy development. The model's existence offers a framework for decision-makers to identify the primary drivers of emissions in a proposed commuter rail system. The figures derived from this robust model could empower government transportation sectors and planning organizations not only to justify public investments in significant transportation projects but also establish evidence-based policy objectives to pave the way for a decarbonized transportation future.

4. Limitation & Future Work

We encountered three main limitations in the study. The first concerns the uncertainty inherent in estimating single-point values for future projections, such as the emission factor of hydrogen or the fuel economy of LDVs in 2040. Given that commuter rail projects typically span more than a decade, overcoming this challenge is inherently difficult. Our approach involves providing users with flexibility to explore a range of values, allowing for an understanding of the magnitude of uncertainty.

The second limitation relates to the generalizability of our model, which is contingent upon data availability in other regions. The microscopic approach we employed relies on two key conditions: (1) the existence of geographically relevant speed-location data that can be used to estimate vehicle behavior for hypothetical service routes; and (2) the availability of ridership estimates of inflow and outflow at each station. In the absence of these conditions being met, it may be difficult to apply the model to a given service route.

Third, the current model does not account for changing fuel weight as a function of energy consumed. This is because incorporating the changing weight of fuel based on energy consumption and fuel efficiency was determined to be beyond the temporal scope of the project. With a full diesel tank, fuel can weigh more than the aggregate mass of passengers. Therefore, including variable fuel weight represents an important future improvement to the model.

Furthermore, the present calculations are limited to GHG emissions. Future research could expand to include additional air pollutants recognized under Criteria Air Pollutants, such as carbon monoxide and sulfur dioxide (US EPA, 2014). This expansion would allow for a comprehensive study of the emissions of other types of pollutants and their potential impacts on the health of local residents. Besides that, our study primarily focuses on the calculation of a train's energy consumption and emissions, but specific emission reduction strategies across various scenarios can be further analyzed in the future by integrating additional parameters into the model.

Lastly, incorporating economic cost-benefit calculations into the model could provide significant value to policymakers since the implementation of alternative fuel technologies would come with its own unique costs. For example, if implementing hydrogen fuel trains, a passenger rail service provider would need to consider the capital costs of building one or more fueling stations, as well as the mechanism by which the

rail service provider would acquire hydrogen fuel (e.g., by generating it on site or by investing in pipeline infrastructure to have it delivered).

Overall, these enhancements significantly contribute to fulfilling North Carolina's commuter rail transportation planning requirements and promote sustainable environmental development. This model, therefore, stands as a crucial tool not only for local rail transportation planning but also for broader environmental management initiatives. To encapsulate the work accomplished, its value, and its potential impact on the region, a brief conclusion is proposed. Such a synthesis underscores the strategic importance of our efforts and aligns with the long-term vision for transportation and environmental sustainability in North Carolina.

Acknowledgement

We extend our deepest gratitude to a wide array of individuals and organizations whose invaluable contributions have been instrumental in the development and success of this project. Our acknowledgment begins with recognizing the United States for its ongoing efforts to address and mitigate greenhouse gas (GHG) emissions within the transportation sector, an area identified as a significant contributor to the nation's environmental footprint.

We are particularly thankful to NCDOT, GoTriangle, and other regional transit authorities in North Carolina for their data provision and for their long-term planning around sustainable transportation solutions. The demographics projections and support from GoTriangle have underscored the critical need for innovative transit strategies to accommodate the Triangle region's anticipated growth. We appreciate the strong support from business leaders and residents in the Triangle area, as evidenced by the report conducted by GoTriangle (Go Forward), which reflects the community's commitment to sustainable and efficient transportation.

Our efforts were significantly bolstered by the Applied Data Research Institute (ADRI), whose expertise in data analytics and systems modeling has been pivotal in navigating the complexities of this project. ADRI's proactive engagement with key transportation authorities and their guidance on Geographic Information Systems (GIS) work have been invaluable in ensuring the accuracy and completeness of our analyses. Special recognition is extended to Alexander Yoshizumi for this exceptional contribution in rail modeling guidelines that have been crucial in refining our analytical processes. His expertise and meticulous approach have significantly enhanced the quality and reliability of our model outputs.

We are also profoundly thankful to advisor Dr. Timothy Johnson for his guidance and wisdom throughout the project. His expertise in the field and his role as a steadfast advisor have been indispensable in navigating the project's challenges and achieving our goals.

Lastly, we acknowledge that collaborative spirit and shared vision of all stakeholders and policymakers involved in this initiative. By focusing on different scenarios pertinent to the region's needs and employing evidence-based assumptions, we hope to catalyze meaningful change towards sustainable transportation solutions in North Carolina. This work highlights the critical role that infrastructure projects, particularly those focused on sustainable transportation, can play in achieving broader environmental and social goals. Through continued collaboration, innovation, and commitment to evidence-based practices, we can realize the vision of a more sustainable and prosperous North Carolina for current and future generations.

Appendices

Appendix 1: Existing Emission Estimation Models

The [table](#) compares six models for transportation emission estimation by parameters such as fuel type, environmental impacts, user-defined inputs, and default values/optional user inputs. The table is maintained by the Applied Data Research Institute.

Appendix 2: Typical Value of Rolling and Aerodynamic Resistances Coefficients

Equipment Type	Constants Type	Constants	Source
4-axle locomotive	c_{roll}^{loc}	0.0025-0.0035	Wende and Gralla, 1997
6-axle locomotive	c_{roll}^{loc}	0.0035-0.0045	Wende and Gralla, 1997
Passenger Train Car	c_{roll}^{car}	0.0006	Lindgreen and Sorenson, 2005, p. 29
Passenger Train Car	c_1	0.0025	Lindgreen and Sorenson, 2005, p. 29
Passenger Train Car	c_2	0.005	Lindgreen and Sorenson, 2005, p. 29
Locomotive	c_{air}^{loc}	1.1	Lindgreen and Sorenson, 2005, p. 30
Passenger Train Car	c_{air}^{car}	0.218	Lindgreen and Sorenson, 2005, p. 30

Appendix 3: Model Sheet Structure

[Schedule] could be either 8-2-8-2 or 3-3 in this model

Model_[Schedule]_Trip	Model_[Schedule]_Line	Model_Line_Peak	Model_Station_Flow
Trip_ID	Trip_ID	Line_ID	Productions
Locomotive	Line_Consist	Average_Speed_mps	Attractions
Cars	Resistance_rolling	Resistance_rolling	PA
Base_Weight_kg	Rolling_Constant	Rolling_Constant	AP
Origin_Node	Total_Weight_kg	Total_Weight_kg	PA_Direction
Destination_Node	Resistance_air	Resistance_air	AP_Direction
Origin	Aerodynamic_Constant	Aerodynamic_Constant	Total_Flow
Destination	Average_Speed_mps	Resistance_gradient	PA_PK_Flow
OD_ID	Resistance_gradient	Slope	PA_OP_Flow
Round_Trip_ID	Slope	Resistance_acceleration	AP_PK_Flow
Interval_ID	Resistance_acceleration	Acceleration_mpss	AP_OP_Flow
Service	Acceleration_mpss	Trip_ID	PA_PK_RpT
Type	Kinetic_Energy_J	Passenger	PA_OP_RpT
Direction	TTW_Energy_Diesel_MJ	Kinetic_Energy_J	AP_PK_RpT
Inflow_Origin_Passengers	Emission_Diesel_kgCO2e	TTW_Energy_Diesel_MJ	AP_OP_RpT
Outflow_Destination_Pass	TTW_Energy_Hydrogen_kgCC	Emission_Diesel_kgCO2e	
Ridership_Passengers	Emission_Hydrogen_kgCC		
Passenger_Weight_kg	Interval		
Total_Weight_kg			
Kinetic_Energy_J			
Emission_Diesel_kgCO2e			
Emission_Hydrogen_kgCC			
Passenger_Mile			
Emission_Per_Pass_Mile_Diesel_kgCO2e			
Emission_Per_Pass_Mile_Hydrogen_kgCO2e			
Energy_Road_Btu			
Emission_Road_kgCO2e			

Figure A1.11 The modeling spreadsheets and their attributes for the microscopic emission model. Plotted by the authors.

Coefficients_LT	Stations_LT	Trains_LT	Rolling_Constants_LT	Nodes_LT
Constant	Station_ID	Type	Trip_ID	Node_ID
Unit	Staion_Key	Fuel_Type	Line_Consist	Curve_CUR_degree
Value	Station_Name	Model	Total_Weight_kg	Curve_Max_FOL_2500ft
References	Node_ID	Manufacturer	Average_Speed_kmph	Curve_Max_FOL_2k5_5
	East_Line_ID	Amount_Train	Rolling_Constant	Curve_Max_FOL_5k_7k
	West_Line_ID	Weight_kg		Curve_Max_FOL_7k5_1
	x_coordinate	Length_m		Curve_Max_PRE_2500ft
	y_coordinate	Width_m		Curve_Max_PRE_2k5_5
		Height_m		Curve_Max_PRE_5k_7k
		Front_area_m2		Curve_Max_PRE_7k5_1
		Capacity_car		Elev_Delta_CUR_ft
		Fuel_Capacity		Elev_Delta_FOL_2500ft
		Range_km		Elev_Delta_FOL_2k5_5f
		References		Elev_Delta_FOL_5k_7kf
				Elev_Delta_FOL_7k5_1f
				Elev_Delta_PRE_2500ft
				Elev_Delta_PRE_2k5_5f
				Elev_Delta_PRE_5k_7kf
				Elev_Delta_PRE_7k5_1f
				Distance_to_FOL_Inters
				Distance_to_FOL_Statio
				Distance_to_FOL_Switch
				Distance_to_PRE_Inters
				Distance_to_PRE_Static
				Distance_to_PRE_Switch
				Distance_to_PRE_Switch
				Loco
				Coach
				CCU
				x_coordinate
				y_coordinate
				Speed_mph
				Direction

Emissions_Factors_LT	PA_Station_Flow_LT	Intervals_LT	Trips_LT	Lines_LT
Fuel_Type	Productions	Interval_ID	Trip_ID_8-2-8-2	Line_ID
Subtype1	Attractions	Start_Station	Line_Consist	Nodes
Subtype2	PA	End_Station	Total_Weight	Start_Node
Energy_Content_low	AP	Start_Station_Name	Line_Amount	End_Node
Energy_Content_high	Station_Flow	End_Station_Name	Round_Trip_ID	Line_Length_ft
Fuel_Density_kg_gal	PA_to_AP_Ratio	Start_Node_ID	Departure_Time	Averagae_Speed_mph
CO2_EF_kg	Proportion_of_Total	End_Node_ID		Averagae_Speed_mps
CH4_EF_g	PA_Direction	Start_Line		Slowest_second
N2O_EF_g		End_Line		Fastest_second
CO2e_kg_MMBtu		Trip_Model_3-3		Slope
CO2e_kg_kg_low		Trip_Model_3-3_Amount		Acceleration_mpss
CO2e_kg_kg_high		Trip_Model_8-2-8-2		Interval
References		Trip_Model_8-2-8-2_Amount		
		Line		
		Line_Model_8-2-8-2_Amount		

Figure A2.12 The lookup tables and their attributes for the microscopic emission model.

Plotted by the authors.

Appendix 4: Variables for the Speed Machine-Learning Model

Variable	Description
Speed_mph	Instantaneous speed from the Amtrak real time data but removing zeroes that are not at a station to avoid capturing unusual stops.
Curve_CUR_degree	Current curve's angle with regard to the next and the previous point
Curve_Max_FOL_2500ft_degree	The maximum angle in the ten following observations, measured in angular degrees.
Curve_Max_FOL_2k5_5kft_degree	The maximum angle between the eleventh to the twentieth following observations, measured in angular degrees.
Curve_Max_FOL_5k_7k5ft_degree	The maximum angle between the twenty first to the thirtieth following observations, measured in angular degrees.
Curve_Max_FOL_7k5_10kft_degree	The maximum angle between the thirty-first to the fortieth following observations, measured in angular degrees.
Curve_Max_PRE_2500ft_degree	The maximum angle in the ten preceding observations, measured in angular degrees.
Curve_Max_PRE_2k5_5kft_degree	The maximum angle between the eleventh to the twentieth preceding observations, measured in angular degrees.
Curve_Max_PRE_5k_7k5ft_degree	The maximum angle between the twenty first to thirtieth preceding observations, measured in angular degrees.
Curve_Max_PRE_7k5_10kft_degree	The maximum angle between the thirty-first to the fortieth preceding observations, measured in angular degrees.
Elev_Delta_CUR_ft	The change in elevation from the preceding observation to the following observation. Using observations spaced 50 feet apart, this corresponds to the change in elevation over a 100-foot run.
Elev_Delta_FOL_2500ft_ft	The maximum change in elevation (Elev_Delta_CUR_ft) from the ten following observations.
Elev_Delta_FOL_2k5_5kft_ft	The maximum change in elevation (Elev_Delta_CUR_ft) between the eleventh and the twentieth following observations.
Elev_Delta_FOL_5k_7k5ft_ft	The maximum change in elevation (Elev_Delta_CUR_ft) between the twenty first and the thirtieth following observations.
Elev_Delta_FOL_7k5_10kft_ft	The maximum change in elevation (Elev_Delta_CUR_ft) between the thirty first and the fortieth following observations.
Elev_Delta_PRE_2500ft_ft	The maximum change in elevation (Elev_Delta_CUR_ft) from the ten preceding observations.
Elev_Delta_PRE_2k5_5kft_ft	The maximum change in elevation (Elev_Delta_CUR_ft) between the eleventh and the twentieth preceding observations.
Elev_Delta_PRE_5k_7k5ft_ft	The maximum change in elevation (Elev_Delta_CUR_ft) between the twenty first and the thirtieth preceding observations.
Elev_Delta_PRE_7k5_10kft_ft	The maximum change in elevation (Elev_Delta_CUR_ft) between the thirty first and the fortieth preceding observations.
Distance_to_FOL_Intersection	Given the direction of the train, distance along route to the intersection
Distance_to_FOL_Station	Given the direction of the train, distance along route to the station
Distance_to_FOL_Switch	Given the direction of the train, distance along route to the next switch
Distance_to_PRE_Intersection	Given the direction of the train, distance along route to the previous intersection

Distance_to_PRE_Station	Given the direction of the train, distance along route to the previous station
Distance_to_PRE_Switch	Given the direction of the train, distance along route to the previous switch
Loco	The count of locomotives in the trainset
Coach	The count of coach cars in the trainset
CCU	The count of de-motored locomotive cars in the trainset

Appendix 5: Scripts for Data Processing

All the scripts - including webpage scraping, processing, modeling, and ArcGIS tools - can be found in this GitHub [repository](#), which is licensed under Creative Commons Attribution-ShareAlike 4.0 International.

References

- Accelerating Decarbonization of the U.S. Energy System*. (2021). National Academies Press.
<https://doi.org/10.17226/25932>
- Amtrak. (n.d.). *Track a Train*. Track Your Train with Google Maps. Retrieved February 7, 2024, from
<https://www.amtrak.com/track-your-train-with-google-maps>
- Argonne National Laboratory. (2023, December 21). *Argonne GREET R&D Model*. Energy Systems and Infrastructure Analysis. <https://greet.anl.gov/>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Congressional Budget Office. (2024, January 18). *The Demographic Outlook: 2024 to 2054*.
<https://www.cbo.gov/publication/59697>
- Diesel multiple unit*. (2024, March 19). Locomotive Wiki.
https://locomotive.fandom.com/wiki/Diesel_multiple_unit
- Division of Strategic Planning & Performance. (2022). *Metra Ridership Trends 2021 Annual Report*.
<https://metra.com/sites/default/files/inline-files/2021%20Annual%20Ridership%20Report%20v6.1.pdf>
- DOE Vehicle Technologies Program. (2011). *Biodiesel Basics*.
[https://www.energy.gov/scep/articles/biodiesel-basics-fact-sheet-vehicle-technologies-program-vtp#:~:text=B100%20provides%20the%20greatest%20emissions,%25%20\(see%20Figure%201\).](https://www.energy.gov/scep/articles/biodiesel-basics-fact-sheet-vehicle-technologies-program-vtp#:~:text=B100%20provides%20the%20greatest%20emissions,%25%20(see%20Figure%201).)
- Duncan, C. (2023, January 11). *Will the Triangle ever get commuter rail? 5 things to know*. Spectrum News 1. <https://spectrumlocalnews.com/nc/charlotte/news/2023/01/11/will-the-triangle-ever-get-commuter-rail--5-things-to-know-about-the-decades-long-effort>
- EFI Foundation, D. T. (2023, February 28). *EFI Foundation unveils The U.S. Hydrogen Demand Action Plan*. EFI Foundation. <https://efifoundation.org/insights/efi-unveils-the-u-s-hydrogen-demand-action-plan/>

- EIA. (2023, December 22). *Diesel fuel explained*. U.S. Energy Information Administration (EIA).
<https://www.eia.gov/energyexplained/diesel-fuel/>
- El-Shafie, M., Kambara, S., & Hayakawa, Y. (2019). *Hydrogen Production Technologies Overview*.
<https://www.h2knowledgecentre.com/content/journal2618>
- Enerdata. (2023). *2050 Projections for CO2 Intensity of Electricity Generation*.
<https://eneroutlook.enerdata.net/forecast-world-co2-intensity-of-electricity-generation.html>
- EPA, N. S. C. for E. P. (2009). *Technical Highlights Emission Factors for Locomotives*.
<https://nepis.epa.gov/Exe/ZyNET.exe/P100500B.TXT?ZyActionD=ZyDocument&Client=EPA&Index=2006+Thru+2010&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5Czyfiles%5CIndex%20Data%5C06thru10%5CTxt%5C00000010%5CP100500B.txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C-&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=hpfr&DefSeekPage=x&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=1&ZyEntry=1&SeekPage=x&ZyPURL>
- European Commission. (2023). *GHG emissions of all world countries: 2023*. Publications Office.
<https://data.europa.eu/doi/10.2760/953322>
- GoForward. (2023). *GTCR Feasibility—Ready For Rail*. <https://www.readyforrailinc.com/feasibility/>
- Gomes, T. (2022, May 23). *Story Recipe: How to access commuting data*. Carolina Demography.
<https://carolinademography.cpc.unc.edu/2022/05/23/story-recipe-how-to-access-commuting-data/>
- GoTriangle. (2023). *GoTriangle Fiscal Year 2023 Annual Report*.
https://gotriangle.org/sites/default/files/publications/fy23_annual_report_web.pdf
- GoTriangle, NCDOT, DCHC, CAMPO, Durham County, Wake County, Johnson County, & NCR. (2022). *Greater Triangle Commuter Rail Feasibility Study: Phase II Summary Report*.

- https://www.readyforrailinc.com/wp-content/uploads/2022/12/GTCR_Feasibility-Study-Summary-Report.pdf
- Greenhouse Gas Protocol. (2014). *Global Warming Potential Values*.
https://ghgprotocol.org/sites/default/files/ghgp/Global-Warming-Potential-Values%20%28Feb%2016%202016%29_1.pdf
- Heinold, A. (2020). Comparing emission estimation models for rail freight transportation. *Transportation Research Part D: Transport and Environment*, 86, 102468.
<https://doi.org/10.1016/j.trd.2020.102468>
- Humphrey, T. J. (2012, December 21). *MBTA Commuter Rail Passenger Count Results*.
https://www.ctps.org/data/html/studies/transit/2012_MBTA_Commuter_Rail_Passenger_Counts/MBTA_Commuter_Rail_Passenger_Count_Results.html
- Hydrogen Council. (2021). *Hydrogen decarbonization pathways: A life-cycle assessment*.
https://hydrogencouncil.com/wp-content/uploads/2021/01/Hydrogen-Council-Report_Decarbonization-Pathways_Part-1-Lifecycle-Assessment.pdf
- Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest*. 2.
- Lin, X., Edwards, J., Dersch, M., & Ruppert, C. J. (2016). *Load Quantification for Light Rail , Heavy Rail , and Commuter Rail Transit Infrastructure*. <https://www.semanticscholar.org/paper/Load-Quantification-for-Light-Rail-%2C-Heavy-Rail-%2C-Lin-Edwards/7b1cb5f81af164e4a91a9bbbe07251122fc63c49>
- Lindgreen, E. B. G., & Sorenson, S. C. (2005). *Simulation of Energy Consumption and Emissions from Rail Traffic*.
- Mooneyham, S. (2023, December 20). *Passenger Rail Expansion Rolling Down the Tracks*. NCLM News Center. <https://www.southerncitymagazine.org/passenger-rail-expansion-rolling-down-the-tracks/>
- NC Department of Transportation, R. D. (2017). *H-Line Existing Track Schematic*.
- NC OSBM. (2022, December 30). *NC's Population to Reach 14.0 Million by 2050*.
<https://www.osbm.nc.gov/blog/2022/12/30/ncs-population-reach-140-million-2050>

NCDOT. (2023, January 18). *N.C. 's Passenger Rail Service Breaks Annual Record for Ridership*.

NCDOT. <https://www.ncdot.gov:443/news/press-releases/Pages/2023/2023-01-18-ncdot-ncbytrain-ridership.aspx>

NCDOT, R. D. (2020). *NCDOT Rail Equipment Overhaul*.

Pahwa, P. K., & Pahwa, G. K. (2014). *Hydrogen Economy*. The Energy and Resources Institute (TERI).

RSG. (2022, July). *GTCR_Appendix-H-Ridership-Analysis-Tech-Memo*. Ready For Rail.

<https://www.readyforrailinc.com/gtcr-appendices/>

Shirley, C. (2022, December 13). *Emissions of Carbon Dioxide in the Transportation Sector* |

Congressional Budget Office. <https://www.cbo.gov/publication/58861>

Siemens Mobility, Inc. (2023). *ALC-42 Charger Diesel-Electric Locomotive*.

https://assets.new.siemens.com/siemens/assets/api/uuid:b2e899ce-b43e-448b-91a2-dea1784f01b2/Amtrak-ALC-42-Data-Sheet_original.pdf

Sklar, D. L. (2020, July 16). *Bombardier Signs Contract with NCTD for the Supply of BiLevel Commuter*

Rail Cars. Times of San Diego. <http://timesofsandiego.com/business/2020/07/16/bombardier-signs-contract-with-nctd-for-the-supply-of-bilevel-commuter-rail-cars/>

U.S. Energy Information Administration. (2024). *Light-Duty Vehicle Miles per Gallon by Technology*

Type. <https://www.eia.gov/outlooks/aeo/data/browser/#/?id=50-AEO2019®ion=0-0&cases=ref2019&start=2017&end=2050&f=Q&linechart=ref2019-d111618a.4-50-AEO2019&sourcekey=0>

US EPA. (2023, October 31). *Fast Facts on Transportation Greenhouse Gas Emissions* [Overviews and

Factsheets]. <https://www.epa.gov/greenvehicles/fast-facts-transportation-greenhouse-gas-emissions>

US EPA, O. (2014, April 9). *Criteria Air Pollutants* [Other Policies and Guidance].

<https://www.epa.gov/criteria-air-pollutants>

US EPA, O. (2015, December 29). *Sources of Greenhouse Gas Emissions* [Overviews and Factsheets].

<https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions>

U.S. Government Accountability Office. (2022, January 25). *Commuter Rail: Information on Benefits and Funding Challenges for Service in Less Urbanized Communities* | U.S. GAO.

<https://www.gao.gov/products/gao-21-355r>

US Greenhouse Gas Emissions by Sector. (n.d.). Resources for the Future. Retrieved January 28, 2024, from <https://www.resources.org/archives/us-greenhouse-gas-emissions-by-sector-infographic/>

USDOT, B. of T. S. (2023, November 16). *Amtrak Stations*. [https://data-](https://data-usdot.opendata.arcgis.com/datasets/amtrak-stations)

[usdot.opendata.arcgis.com/datasets/amtrak-stations](https://data-usdot.opendata.arcgis.com/datasets/amtrak-stations)

USDOT, F. O. of R. S. (2024). *Crossing Inventory Data (Form 71)—Current* [dataset].

https://data.transportation.gov/Railroads/Crossing-Inventory-Data-Form-71-Current/m2f8-22s6/about_data

Vehicle Technologies Office. (2021, March 1). *FOTW# 1175, March 1, 2021: Vehicles Registered in the District of Columbia Averaged 22 Miles per Gallon in 2018* | Department of Energy.

<https://www.energy.gov/eere/vehicles/articles/fotw-1175-march-1-2021-vehicles-registered-district-columbia-averaged-22>