Data Science for Transportation (2026) 8:2
https://doi.org/10.1007/s42421-025-00145-1

RESEARCH

Robust Forecasting Framework for Energy-Informed Planning in Urban Rail Transit Systems

Zhuo Han[1] · Eleni Christofa[1] · Eric J. Gonzales[1] · Sean Donaghy[2] · Jimi Oke[1]

Received: 25 August 2025 / Revised: 18 November 2025 / Accepted: 5 December 2025

Abstract
Urban Rail Transit (URT) systems play a critical role in modern cities but consume significant amounts of energy. Understanding and predicting energy consumption patterns in these systems is vital for sustainable urban planning and cost-effective decision-making. This study develops a comprehensive framework for forecasting energy consumption based on planning metrics, ridership data, and historical energy and weather, using the Massachusetts Bay Transportation Authority (MBTA)'s URT system as a case study. The fitted eXtreme Gradient Boosting (XGBoost) model demonstrated high predictive accuracy, with a mean absolute percentage error (MAPE) of 3.3%, while ridge regression models reliably forecasted ridership and temperature to support future planning. We further evaluated eight operational plans under nine ridership–temperature scenarios. Each plan represents a set of operational adjustments that may be pursued to regulate service conditions or respond to disruptions. These analyses highlight the model's capability to provide insightful forecasts that inform decision-making processes and optimize energy usage, thereby supporting sustainable and efficient urban rail transit operations.

Keywords Urban rail transit · Energy forecasting · XGBoost · Operations planning · Decision-making

Introduction

The long-term decline in urban transit ridership coupled with greenhouse gas emissions reduction targets requires strategic planning and sustainable decision-making from transit agencies. Particularly vulnerable to these disruptions are urban rail transit (URT) systems, which are often indispensable to the regions they serve and provide efficient transportation to millions of people daily (APTA 2021). However, the operations of these extensive transportation networks incur significant energy demands, accounting for a substantial portion of total energy consumption in many areas, and costing millions of dollars annually (Shang and Zhang 2013). In 2019, operating rail transit systems in the United States consumed a substantial amount of energy, totaling 4953 GWh. At the same time, these systems facilitated a considerable volume of travel, with annual passenger miles reaching about 20 billion (Davis and Boundy 2022). Moreover, the patterns of energy consumption significantly vary among different rail systems. For instance, electricity consumption in Beijing surged from 650 GWh in 2006 to 1400 GWh in 2017 (Chenchen et al. 2022). This growing trend highlights the pressing need for sustainable energy planning, which is further amplified by the rapidly evolving demographic trends and urbanization patterns witnessed in many cities. An effective energy forecasting model thus serves not only as a tool for proactively managing demand in a sustainable and economical manner but also as a critical asset for planning and maintaining resilience under disruption.

To this end, this paper investigates the planning process of URT operations, focusing on forecasting energy

✉ Zhuo Han
zhuohan@umass.edu

✉ Eleni Christofa
echristofa@umass.edu

✉ Eric J. Gonzales
gonzales@umass.edu

✉ Sean Donaghy
sdonaghy@mbta.com

✉ Jimi Oke
jimi@umass.edu

1 Department of Civil and Environmental Engineering, University of Massachusetts Amherst, 130 Natural Resources Road, Amherst, MA 01003, USA

2 Massachusetts Bay Transportation Authority, 10 Park Plaza 5720, Boston, MA 02116, USA

Published online: 17 December 2025

consumption and analyzing strategies that impact energy usage. Using the Massachusetts Bay Transportation Authority (MBTA) as a case study, we develop a robust energy forecasting model that uses an eXtreme Gradient Boosting (XGBoost) to incorporate various planning metrics. Further, we evaluate different operational strategies to provide insights that support decision-making processes for sustainable and energy-efficient URT operations.

The primary contribution of our work is the development of a machine learning framework that forecasts daily energy consumption in a transit system while allowing planners to generate predictions based on line-specific decision variables, such as operating distance and speed. Moreover, our model enables using actual train trajectories to accurately plan energy. By integrating both time-dependent and time-independent variables, the framework supports short-term energy forecasting and provides a tool for decision-making.

Literature Review

Recent research has focused on decision-making, forecasting, and environmental impact assessment in URT systems. A tool based on fuzzy logic and algorithms was developed to mitigate greenhouse gas emissions from transportation, underlining the significance of stakeholder engagement and transparency in decision-making for sustainable transportation (Barcellos de Paula and Marins 2018). To et al. (2020) assessed the Hong Kong URT system using the triple bottom line framework and highlighted the environmental, economic, and social performance of URT systems. Sustainability indicators have been demonstrated to play a crucial role in guiding transportation planning and policy-making, particularly when there are well-established monitoring criteria (Santos and Ribeiro 2013). The complexity of urban traffic systems and the multifaceted impacts of URT systems on traffic, economy, society, and environment can be practically analyzed using a system dynamics model (Yang et al. 2014). Diverse approaches have been used to forecast the energy usage of URT systems in different contexts. For instance, multivariate adaptive regression spline (MARS) and support vector regression (SVR) models were applied to short-term electricity demand forecasting for Queensland, Australia. The MARS model was found to perform well for 0.5 and 1.0-hour predictions, whereas the SVR model excelled in day-ahead forecasting, especially when considering external variables such as weather data and economic factors (Al-Musaylh et al. 2018). Furthermore, binary nonlinear fitting regression and SVR models were also used to forecast energy consumption for URT systems. It was observed that the SVR model accurately predicted traction and total electricity consumption. However, both models faced challenges with HVAC system predictions due to complex influencing factors (Tang et al. 2021). Additionally, Wang et al. (2024) developed a deep reinforcement learning approach to optimize speed profiles for energy-efficient train operations in urban rail transit systems.

Beyond energy forecasting, operators are increasingly focused on using advanced models to aid decision-making in URT systems. A planning framework developed by Krishnan et al. (2015) minimizes system-wide costs by co-optimizing energy and transportation infrastructure. This framework assesses the impacts of high-speed rail investments in the U.S., demonstrating substantial long-term benefits, including cost savings, reduced fuel consumption, and mitigation of greenhouse gas emissions. Complementing this, Abbasabadi et al. (2019) introduced a data-driven framework for urban energy use modeling that employs machine learning to predict building and transportation energy use, highlighting its efficacy in aiding urban planners and policymakers. Similarly, Manzolli et al. (2021) utilized a scenario-based multi-criteria decision analysis with the preference ranking organization method for enrichment evaluations method to evaluate rapid transit systems, revealing the nuanced benefits across various economic and environmental scenarios. Additionally, Gu et al. (2014) proposed an energy-efficient train operation model using real-time traffic information, which showed significant potential in reducing energy consumption and maintaining punctuality in urban rail transit systems. Together, these studies highlight the critical role of sophisticated modeling tools in optimizing infrastructure investments and operational strategies, thereby promoting sustainable urban development.

Further expanding the scope, Huang and Liao (2023) developed a novel two-stage approach for energy-efficient timetabling in urban rail transit networks, addressing the complex interactions between passenger path choices and operational schedules to minimize energy consumption while accommodating passenger travel time constraints. Moreover, Hamurcu and Eren (2020) presented a strategic planning framework using analytical hierarchy process and fuzzy technique for order preference by similarity to ideal solution to prioritize sustainable urban transportation projects in a Turkish city, ultimately recommending the implementation of electric municipality buses for their environmental and operational benefits. Finally, Zhao and Deng (2013) developed a fuzzy multiobjective decision support model for urban rail transit projects in China, using multilevel comprehensive evaluation to consider influential objectives such as traveler attraction, environmental protection, project feasibility, and operation.

Despite the various energy modeling efforts for urban rail transit systems, there remains a notable gap in their ability to effectively assist planners in evaluating how energy consumption could change in response to operational alterations. Current models do not specifically cater to the analysis of disruption-response strategies and their potential energy outcomes, highlighting a crucial need to address the current gap in research regarding energy planning. In this paper, we estimate a system-wide daily energy forecasting model using planning variables, such as operating distance, average daily speed, and daily trips for URT systems. The URT system in Boston served as the study area. By analyzing various operating plans, we demonstrate how this model can support regular planning and decision-making in response to disruptions, not only promoting sustainability and energy efficiency but also enabling more accurate energy forecasting, which can lead to cost savings.

Data

Study Area

The case study area for the forecasting model was the Massachusetts Bay Transportation Authority's (MBTA) urban rail transit system (Fig. 1). The fourth largest transit agency in the U.S., the MBTA URT served 1.7 billion passenger miles in the area (Han et al. 2022). Comprising a light rail line (Green) and three heavy rail lines (Red, Orange, and Blue), this network accounted for an annual average electricity consumption of 422 GWh at a cost of approximately $38 million between 2009 and 2020 (Han et al. 2023).

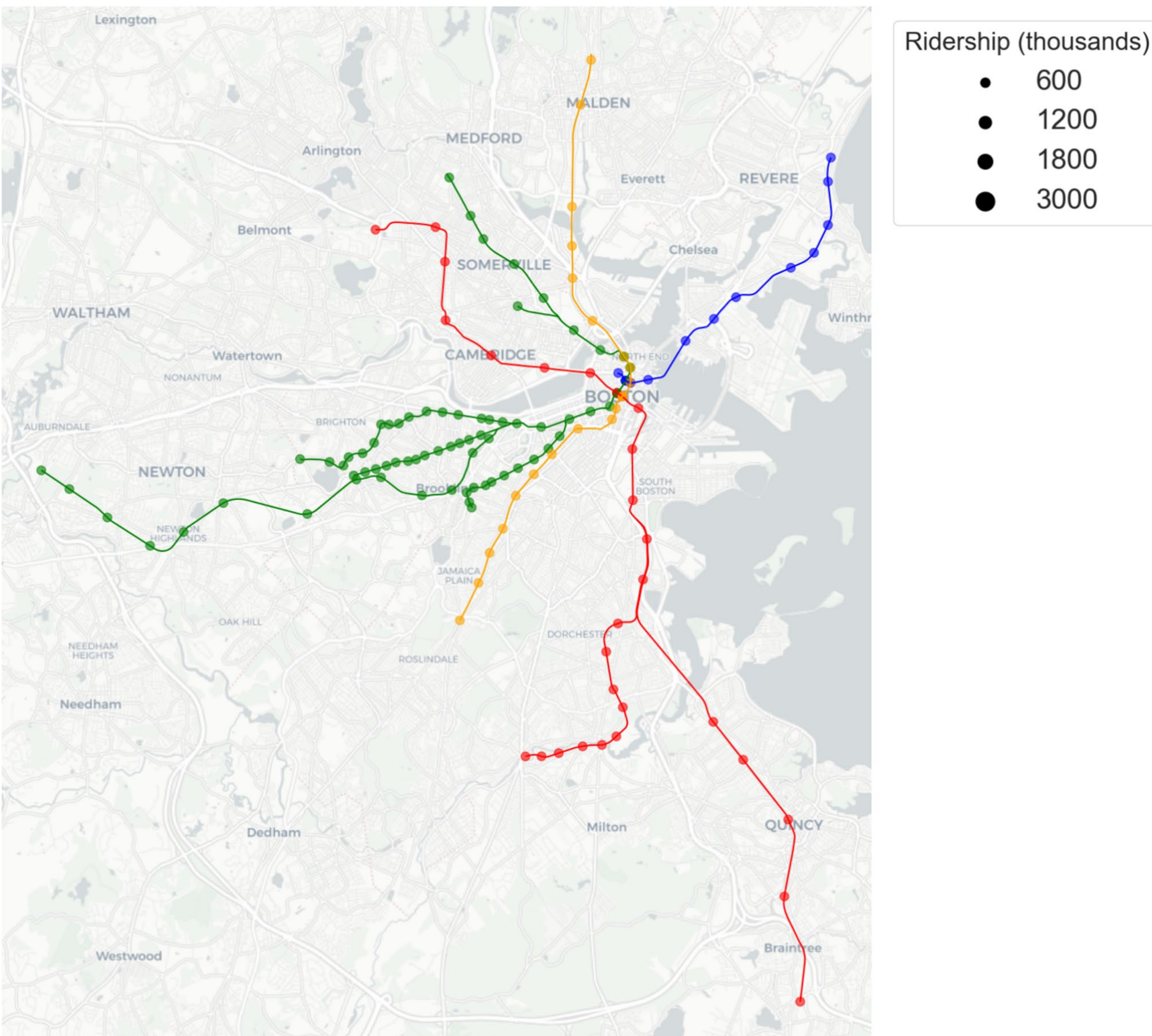

Fig. 1 The map of MBTA urban rail transit system with marker sizes indicating the range of annual ridership in 2021 across stations

For this study, we used daily data from 2019–2021, including system energy and five planning variables-ridership, operating speed, operating distance, operating time, and average speed-as predictors of energy consumption. Energy data were obtained from the MBTA, and planning variables were derived from the MIT-MBTA research database (Han et al. 2022, 2023). The train location timestamps (longitude and latitude) were processed to compute operating time and distance, which were aggregated to the daily level and used to derive the average speed for each line.

Energy

We collected and analyzed system energy data from the MBTA urban rail transit rail network over the period 2019 to 2021. The data were obtained from energy meters installed throughout the system and captured at an hourly resolution. For the purpose of our analysis, however, we aggregated this data into daily totals as shown in Fig. 2.

Temperature

Between 2019 and 2021, the URT system experienced a consistent decrease in energy consumption, a trend attributed to the impacts of COVID-19. We obtained average daily temperatures in Boston from the National Oceanic and Atmospheric Administration (NOAA) database for the years 2019 through 2021 [21] as shown inFig. 2.

Ridership

System-wide ridership from January 2019 to December 2021 shows an initial period of high and relatively stable levels until early 2020, followed by a dramatic drop with the onset of the COVID-19 pandemic in March 2020. This sharp decline reflects the impact of lockdowns and travel restrictions, significantly disrupting train operations and keeping ridership low through much of 2020. A slow recovery began around mid-2020 as restrictions eased and vaccines were rolled out as shown in Fig. 3.

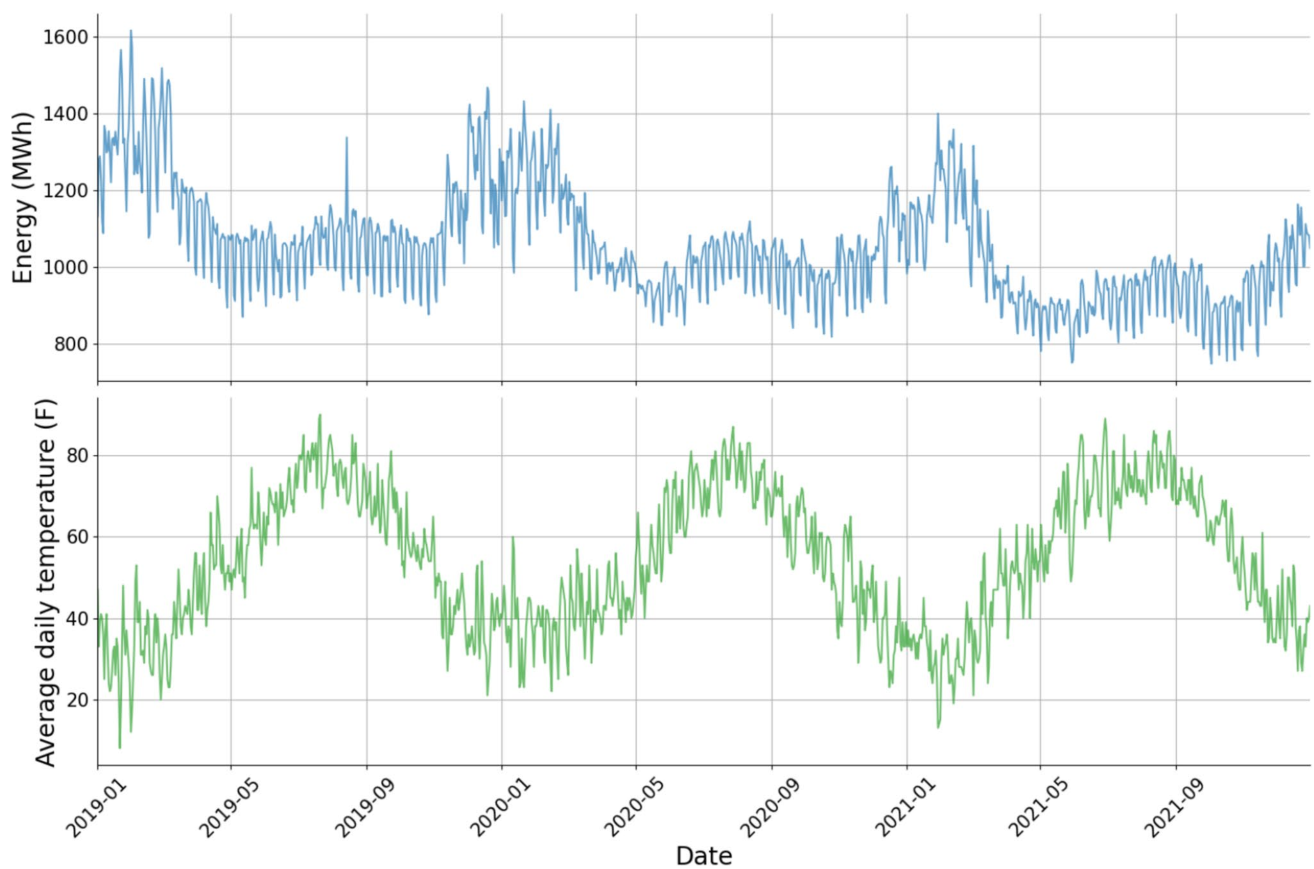

Fig. 2 The daily time series of energy consumption and temperature from 2019 to 2021

Fig. 3 The time series of the planning metrics extracted from GTFS data and system-wide ridership

Planning Variables

In this study, we focus on key planning metrics derived from the General Transit Feed Specification (GTFS) data as shown in Fig. 3. Specifically, we used the line-specific number of trips, operating distance, and average speed. These metrics were identified as key components of planning by agency and provide insights into daily planning.

Average Daily Speed

From January 2019 to December 2021, the Red, Blue, Mattapan, and Orange Lines consistently maintain average daily speeds ranging between 16 to 20 mph. Despite minor fluctuations (Fig. 3), these lines exhibit steady performance throughout the period. In contrast, the Green Line demonstrates significantly lower speeds, predominantly between 8 and 12 mph, which can likely be attributed to its status as the only light rail in the system, inherently operating at lower speeds.

Operating Distance

The Green Line shows the most variable operating distances, ranging between 6000 and 9000 miles over the observed period, with frequent fluctuations (Fig. 3). In contrast, both the Red and Orange Lines maintain more stable operating distances, with the Red Line covering 6000 to 8000 miles and the Orange Line spanning 4,000 to 6,000 miles. Although these lines exhibit periodic adjustments, they follow consistent service patterns. All lines experienced a significant reduction in operating distance around mid-2020 as a result of the COVID-19 pandemic, followed by a gradual recovery.

Number of Trips

The Green Line consistently records the highest number of daily trips, ranging between 1200 and 1800 (Fig. 3), indicating its heavy utilization. The Red Line experienced an increase in trips around mid-2019 before stabilizing at 300 to 400 trips toward the end of the period. In contrast, the Orange and Mattapan Lines exhibit greater fluctuations, particularly around early 2020, while the Blue Line remains more stable with fewer trips than the other lines.

Methods

Figure 4 displays a reseach framwork to forecast future energy. Planning metrics are derived from GTFS data, while actual operating metrics are computed from raw train trajectories. We compared these two sets of metrics to estimate the distribution of discrepancies, which reveals deviations between planned and actual operations and allows us to generate and add noise directly to the original planning metrics. These generated metrics, together with historical energy consumption, ridership, and temperature serve as inputs

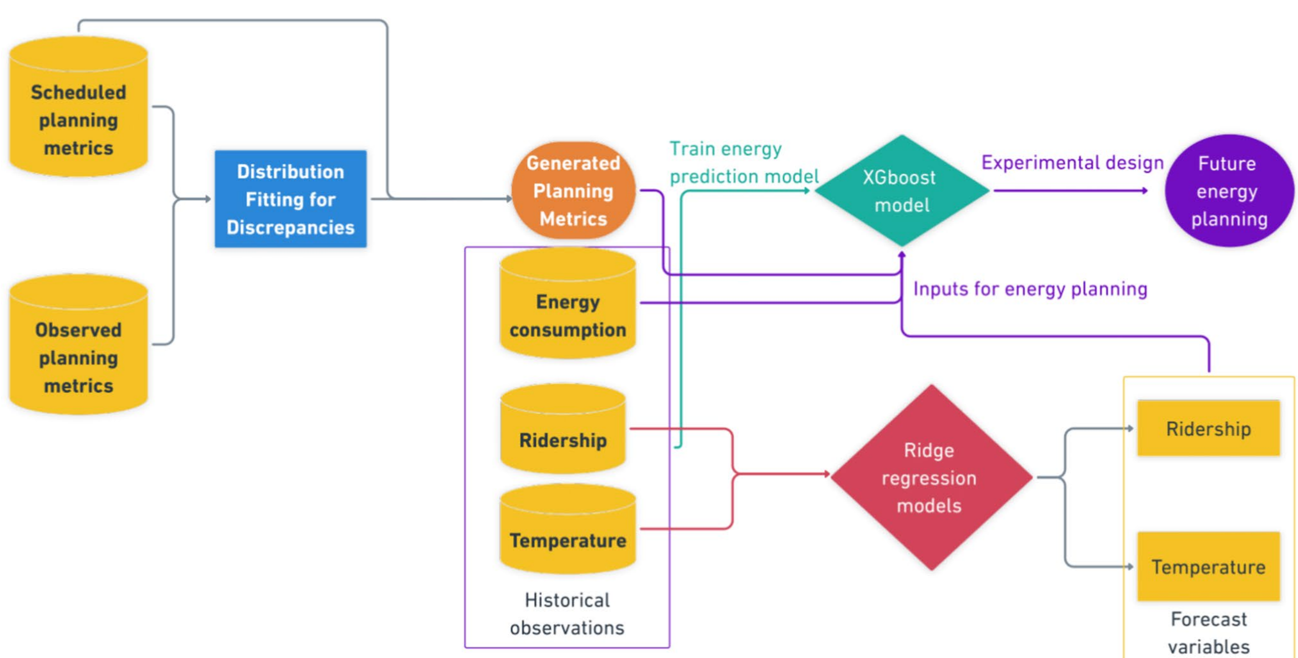

Fig. 4 Pipeline for scenario planning of energy consumption using both planning metrics and actual operating metrics

to a energy forecasting model. In parallel, ridge regression models are trained to forecast temperature and ridership, which are then used as forecasting inputs for future energy planning. Additionally, we conduct scenario analyses to evaluate how different operational strategies impact future energy outcomes.

Scenario Input Generation

We observed discrepancies between actual and planned trajectories, which arise from the variability in real-world operations compared to scheduled times. There was also a noticeable difference in the distributions of the discrepancies by weekend and weekday (see Fig. 5). Thus, we fitted models to these discrepancies in order to generate realistic planning metric inputs (speed, distance, number of trips) for future scenario evaluations. In addition, we modeled the discrepancies specific to weekdays ($d = 1$) and weekends ($d = 0$).

Our approach was as follows. We define the expected value of a planning metric j on day i for line ℓ as $\overline{x}_{ij\ell}$, and its actual or observed value as $x_{ij\ell}$, then the observed discrepancy $\epsilon_{ij\ell}$ is given by:

$$\epsilon_{ij\ell} = x_{ij\ell} - \overline{x}_{ij\ell} \tag{1}$$

We assume that each discrepancy is governed by weekday- and weekend-specific probability density functions (PDFs) f_d with parameters $\boldsymbol{\theta}$. Thus:

$$\epsilon_{ij\ell} \sim f_d(\epsilon_{ij\ell};\boldsymbol{\theta}) \tag{2}$$

In order to generate realistic inputs for future scenarios, we fitted various probability density functions to each discrepancy distribution via maximum likelihood estimation and selected the best-fitting one using the Akaike and Bayesian information criteria (AIC and BIC) as fitness metrics. The results are summarized in Table 1, while specifications of the optimal probability density functions (PDFs) are provided in the appendix.

We then sampled the discrepancy terms $\hat{\epsilon}_{ij\ell}$ from their corresponding fitted distributions (with optimal parameters $\hat{\boldsymbol{\theta}}$ and substitute them into Eq. 3, allowing us to generate realistic perturbed values of the planning metrics as scenario inputs, while preserving the distinct variability patterns observed between weekdays and weekends. The final generated planning metrics $\hat{x}_{ij\ell}$ are given by:

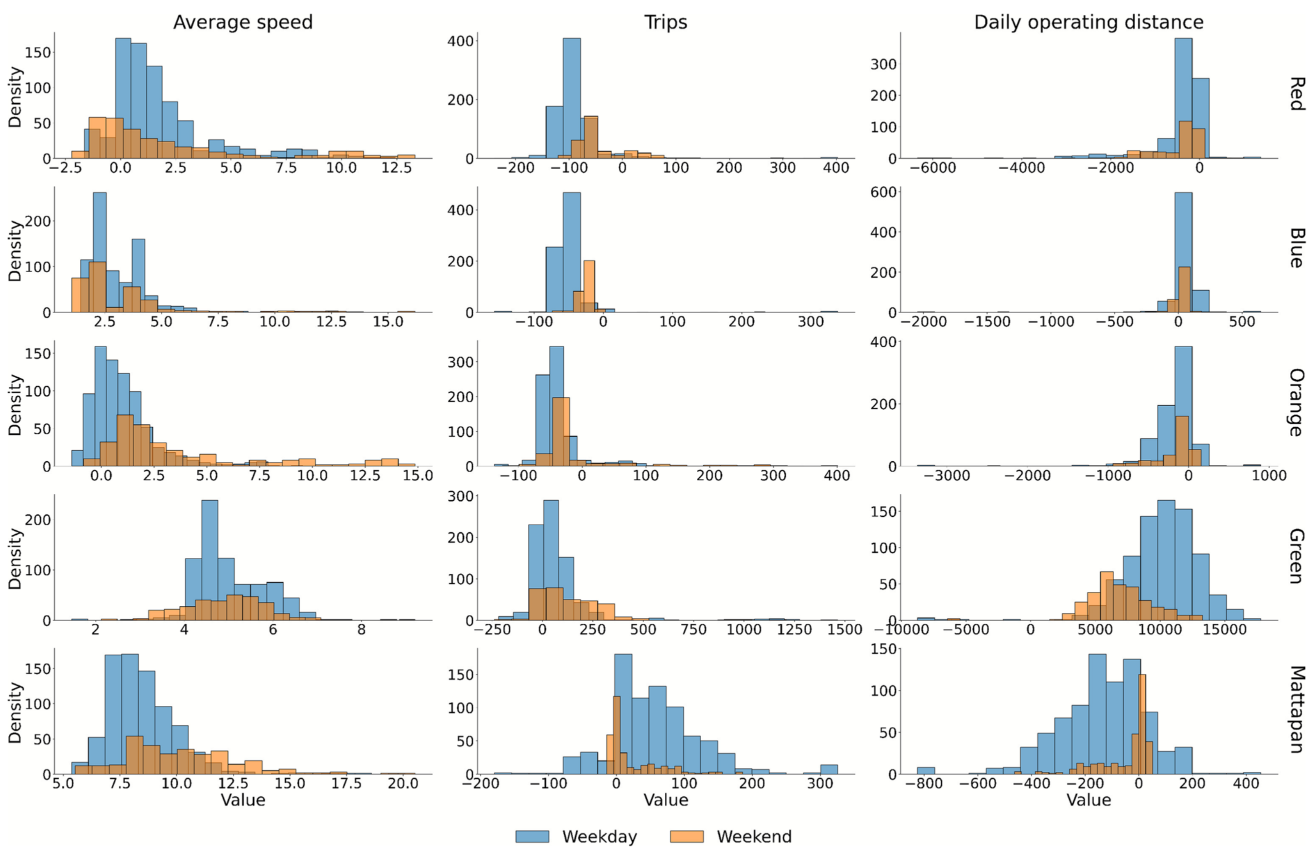

Fig. 5 The distributions of the discrepancy between scheduled data and actual operations

Table 1 Fitted discrepancy distributions and parameters for speed, distance, and trip count by line and day type (1 ≡ weekday; 0 ≡ weekend)

Discrepancy metric	Line	Day type	Distribution	Parameters	Estimates
Speed (mph)	Blue	1	Exponential Power	(α, x_0, λ)	(0.87, 1.43, 3.59)
	Blue	0	Rayleigh	(x_0, σ)	(−0.09, 2.86)
	Green	1	Exponential Power	(α, x_0, λ)	(0.44, 4.46, 2.82)
	Green	0	Cauchy	(x_0, γ)	(6.62, 1.08)
	Mattapan	1	Exponential Power	(α, x_0, λ)	(0.39, 5.41, 2.21)
	Mattapan	0	Exponential Power	(α, x_0, λ)	(0.42, 5.09, 2.21)
	Orange	1	Exponential Power	(α, x_0, λ)	(0.39, 2.59, 1.34)
	Orange	0	Cauchy	(x_0, γ)	(5.90, 0.75)
	Red	1	Exponential Power	(α, x_0, λ)	(0.31, 5.34, 1.50)
	Red	0	Exponential Power	(α, x_0, λ)	(0.24, 5.91, 1.32)
Operating distance (miles)	Blue	1	Cauchy	(x_0, γ)	(58.83, 23.77)
	Blue	0	Power Law	(α, x_0, λ)	(0.96, −262.83, 877.95)
	Green	1	Cauchy	(x_0, γ)	(104.42, 19.63)
	Green	0	Power Law	(α, x_0, λ)	(0.99, −464.85, 1245.85)
	Mattapan	1	Power Law	(α, x_0, λ)	(1.33, −302.33, 897.57)
	Mattapan	0	Cauchy	(x_0, γ)	(125.20, 35.52)
	Orange	1	Power Law	(α, x_0, λ)	(0.91, −259.28, 801.05)
	Orange	0	Power Law	(α, x_0, λ)	(0.81, −267.67, 761.52)
	Red	1	Cauchy	(x_0, γ)	(73.07, 25.88)
	Red	0	Cauchy	(x_0, γ)	(93.63, 28.78)
Trip count	Blue	1	Cauchy	(x_0, γ)	(−54.70, 5.17)
	Blue	0	Power Law	(α, x_0, λ)	(0.51, −74.00, 308.00)
	Green	1	Power Law	(α, x_0, λ)	(0.56, −218.16, 149.26)
	Green	0	Power Law	(α, x_0, λ)	(0.51, −225.80, 137.51)
	Mattapan	1	Power Law	(α, x_0, λ)	(1.24, −179.23, 505.24)
	Mattapan	0	Cauchy	(x_0, γ)	(10.27, 5.79)
	Orange	1	Power Law	(α, x_0, λ)	(0.42, −201.68, 112.39)
	Orange	0	Power Law	(α, x_0, λ)	(0.47, −149.34, 104.20)
	Red	1	Cauchy	(x_0, γ)	(−39.11, 4.17)
	Red	0	Cauchy	(x_0, γ)	(−29.10, 4.86)

$$\hat{x}_{ij\ell} = \overline{x}_{ij\ell} + \hat{\epsilon}_{ij\ell}, \quad \hat{\epsilon}_{ij\ell} \sim f_d(\epsilon_{ij\ell};\hat{\boldsymbol{\theta}}) \tag{3}$$

Energy Prediction Model

We compared three machine learning models to evaluate their capacity for energy planning. Each model was fine-tuned for comparing the performance. The candidate models include:

- **XGBoost:** An efficient gradient boosting framework that builds additive decision trees, well-suited for tabular data and capable of capturing complex nonlinear relationships.
- **Random Forests:** An ensemble method that constructs multiple decision trees and averages their outputs, offering robustness to overfitting and good performance on diverse datasets.
- **FCNN:** A Fully Connected Neural Network that models nonlinear dependencies through multiple dense layers, suitable for capturing complex patterns in structured input data.

Forecasting Models for Temperature and Ridership

Reliable forecasts of ridership and temperature are essential for assisting future energy planning. We forecast daily average temperature and system-wide ridership and evaluate the relative performance of three forecasting models. To provide reliable forecasting, we further expanded the data used for model estimations. For ridership, daily records were extracted from January 2014 to May 2025, with data from January 2014 to December 2024 used for training and forecasts evaluated on the 2025 period. For temperature, daily averages were collected from January 2020 to July 15, 2025, with the most recent 30 days reserved for testing. The forecasts of ridership and temperature generated by these models are then used as direct inputs for the energy prediction model.

Experimental Design

To evaluate the impact of external conditions on system energy consumption, we first designed a set of nine scenarios representing all possible combinations of three ridership levels and three temperature levels. Ridership levels were scaled as low (LR = 0.5), normal (NR = 1.0), and high (HR = 1.5), reflecting reduced (by 50%), baseline, and elevated (by 50%) demand, respectively. Temperature levels were defined as low (LT = 0.75), normal (NT = 1.0), and high (HT = 1.25), corresponding to 25% below normal, baseline, and 25% above normal conditions, respectively. Each scenario corresponds to a unique pairing of a ridership level and a temperature level, as summarized in Table 2.

Under these scenarios, we also developed eight operational plans to capture a spectrum of service strategies that the MBTA might implement under different conditions presented in Table 3. P1 (Winter Resilience) simulates harsh weather with reduced speeds (–10%) and fewer trips (–20%), while P2 (Speed Boost) represents infrastructure-driven improvements that increase speed by 20%. P3 (Service Frequency Boost) models high-frequency service with 50% more trips, and P4 (Late-Night Expansion) extends service hours by increasing distance by 5% and trips by 10%. P5 (Core Capacity Push) reflects significant capacity upgrades, combining faster speeds (+15%), longer distance (+10%), and more trips (+20%). P6 (Emergency Snow) captures weather-driven cutbacks with slower speeds (–5%) and fewer trips (–10%), whereas P7 (Access Expansion) emphasizes equity with a modest 5% increase in trips. Finally, P8

Table 2 Composition of each scenario in terms of ridership (R) and temperature (T) levels. LR/NR/HR correspond to −50%, 0%, +50% change in ridership, and LT/NT/HT correspond to −25%, 0%, +25% change in temperature. A checkmark (✓) indicates the scenario component included in each plan

Scenario	Ridership level			Temperature level		
	LR (-50%)	NR (0%)	HR (+50%)	LT (-25%)	NT (0%)	HT (+25%)
LR + LT	✓			✓		
LR + NT	✓				✓	
LR + HT	✓					✓
NR + LT		✓		✓		
NR + NT		✓			✓	
NR + HT		✓				✓
HR + LT			✓	✓		
HR + NT			✓		✓	
HR + HT			✓			✓

Table 3 Operation plans for the MBTA urban rail transit system. Values represent percentage changes relative to the baseline operation

Line	Metric	P1	P2	P3	P4	P5	P6	P7	P8
Red	Avg Speed	– 10%	+20%	0%	0%	+15%	– 5%	0%	0%
	Distance	0%	0%	0%	+5%	+10%	0%	0%	– 10%
	# Trips	– 20%	0%	+50%	+10%	+20%	– 10%	+5%	– 5%
Blue	Avg Speed	– 10%	+20%	0%	0%	+15%	– 5%	0%	0%
	Distance	0%	0%	0%	+5%	+10%	0%	0%	– 10%
	# Trips	– 20%	0%	+50%	+10%	+20%	– 10%	+5%	– 5%
Orange	Avg Speed	– 10%	+20%	0%	0%	+15%	– 5%	0%	0%
	Distance	0%	0%	0%	+5%	+10%	0%	0%	– 10%
	# Trips	– 20%	0%	+50%	+10%	+20%	– 10%	+5%	– 5%
Green	Avg Speed	– 10%	+20%	0%	0%	+15%	– 5%	0%	0%
	Distance	0%	0%	0%	+5%	+10%	0%	0%	– 10%
	# Trips	– 20%	0%	+50%	+10%	+20%	– 10%	+5%	– 5%
Mattapan	Avg Speed	– 10%	+20%	0%	0%	+15%	– 5%	0%	0%
	Distance	0%	0%	0%	+5%	+10%	0%	0%	– 10%
	# Trips	– 20%	0%	+50%	+10%	+20%	– 10%	+5%	– 5%

(On-Demand Shuttle Integration) integrates shuttle substitution by reducing distance (–10%) and trips (–5%).

Results and Discussion

Energy Prediction

Based on the model performance presented in Table 4, XGBoost shows the best performance shown in Fig. 6, achieving the lowest training error and high test accuracy (RMSE: 48.6 MWh, MAPE: 3.3%), indicating good predictive capability. The Random Forest model exhibits similar predictive accuracy on the test set (RMSE: 48.1 MWh, MAPE: 3.3%) but demonstrates higher training error (RMSE: 52.3 MWh). The FCNN model achieves moderate results with training and test RMSEs of 34.4 MWh and 51.6 MWh, respectively, along with slightly higher MAPE (3.6%), reflecting good predictive ability but less robustness than XGBoost. Overall, XGBoost emerges as the preferable model for predicting energy consumption.

Table 4 Comparison of model performance

Model	Training error		Test error	
	RMSE (MWh)	MAPE (%)	RMSE (MWh)	MAPE (%)
XGBoost	28.5	2.1	48.6	3.3
Random Forest	21.3	1.4	52.6	3.2
FCNN	34.4	2.3	51.6	3.6

Forecasting performance for ridership and temperature

To forecast daily ridership (Fig. 7) and average temperature (Fig. 8), we evaluated three forecasting models-Ridge regression, XGBoost, and ARIMA-each trained and tuned after hyperparameter optimization. Their performance is summarized in Table 5. Ridge regression was ultimately selected because it achieved the best overall predictive accuracy while retaining clear interpretability. The models include several lagged and smoothed features to

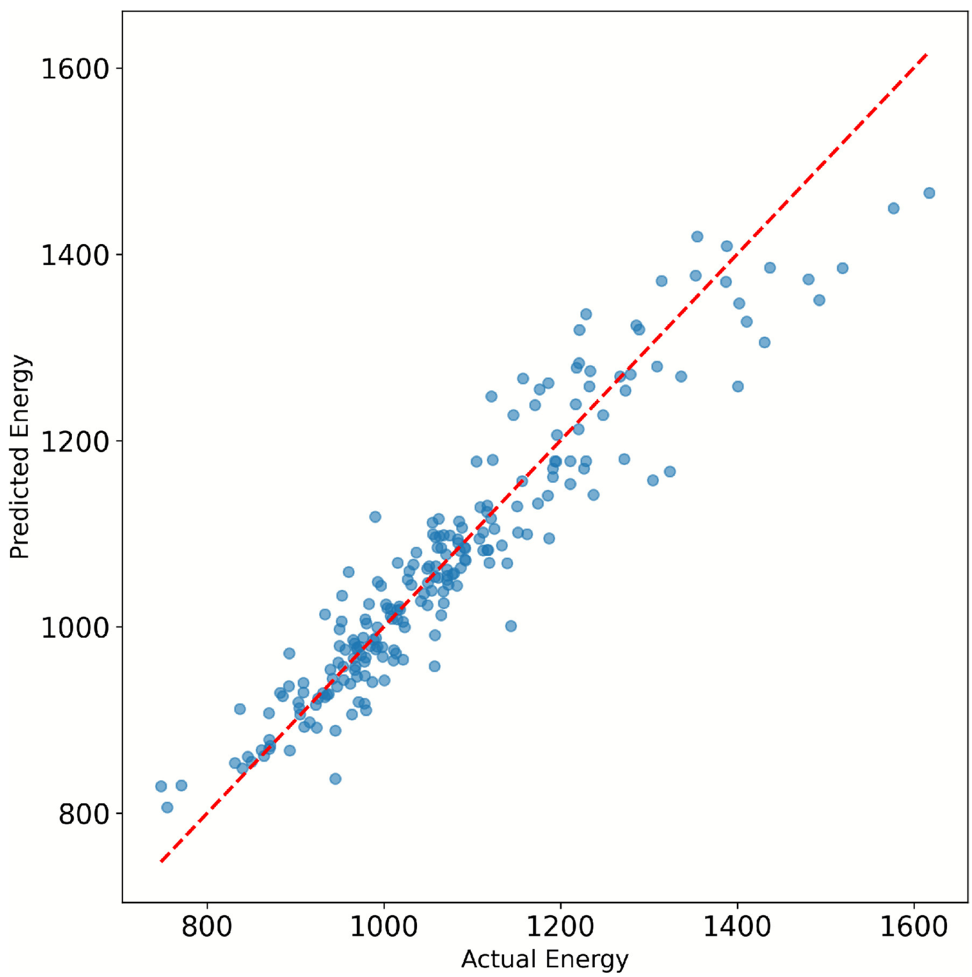

Fig. 6 The forecasting performance of the XGBoost model for daily energy consumption

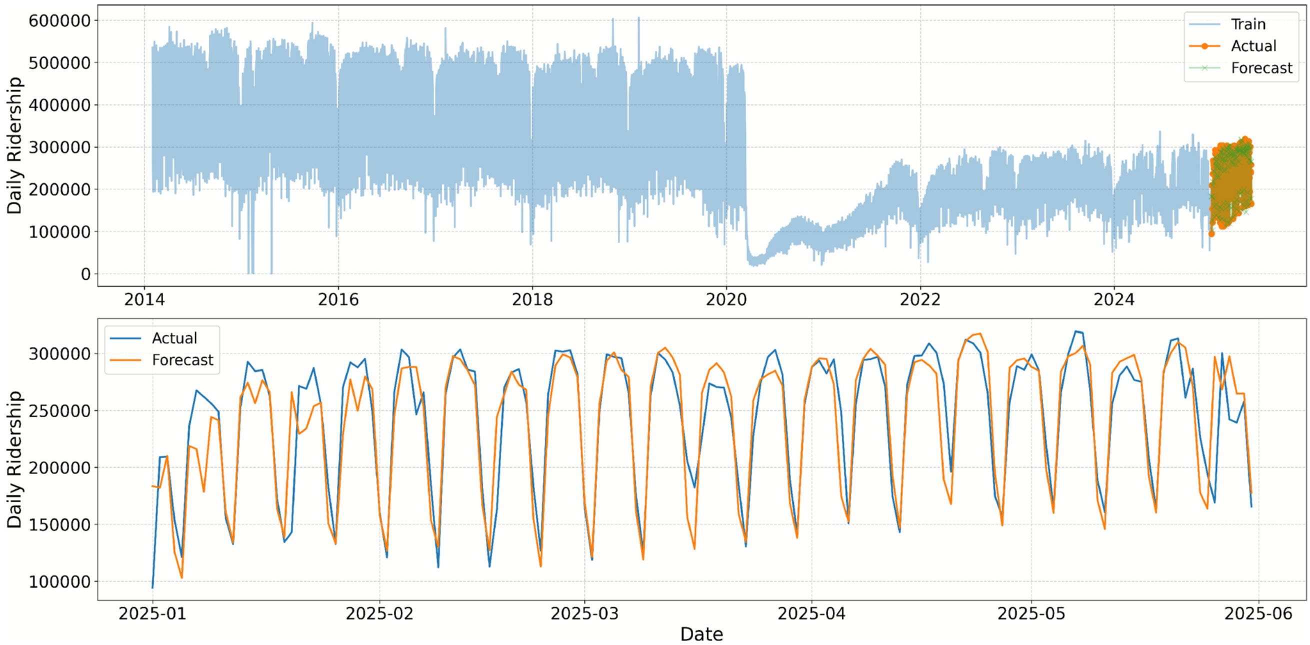

Fig. 7 Daily ridership forecasting results. The top figure shows the full historical series along with the ridge regression forecasts. The bottom figure zooms into the test window, comparing observed ridership against ridge regression forecasts

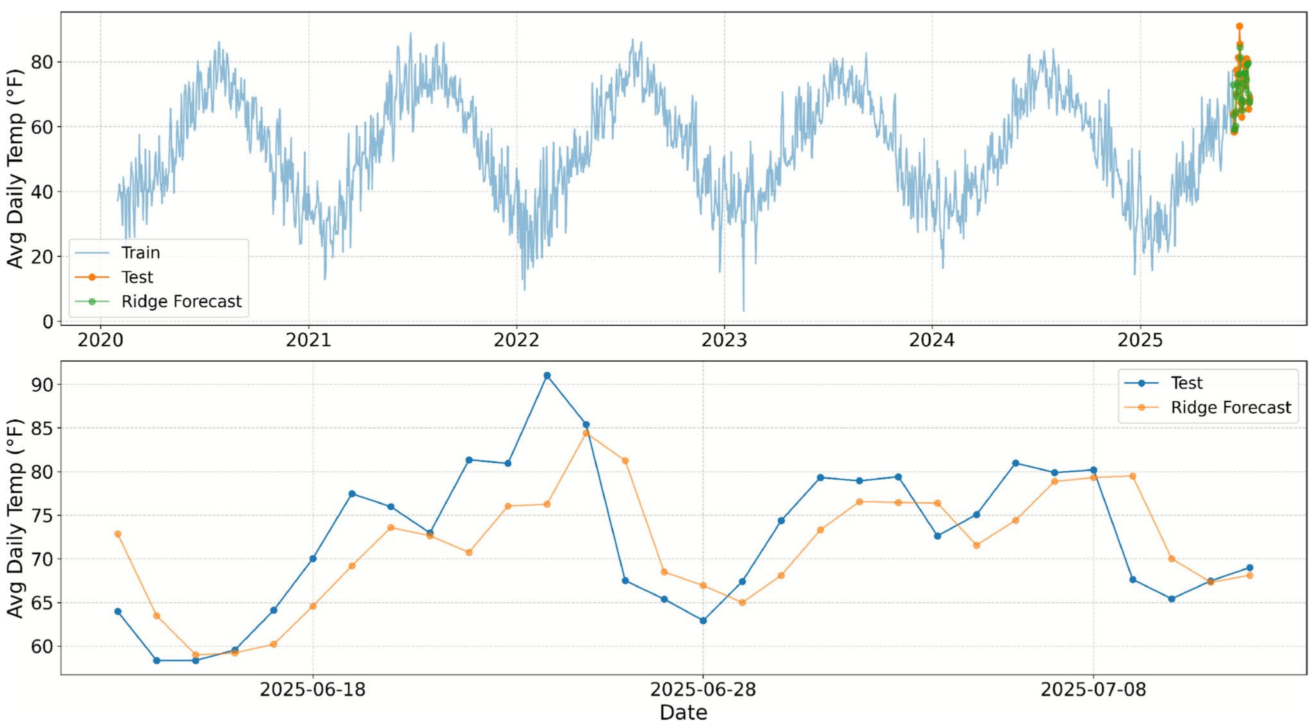

Fig. 8 Daily average temperature forecasting results. The top figure shows the full historical series along with the ridge regression forecasts. The bottom figure zooms into the test window, comparing observed temperatures against ridge regression forecasts

Table 5 Forecasting performance metrics for daily ridership and average daily temperature

	Ridership		Temperature	
Model	**RMSE**	**MAPE (%)**	**RMSE (°F)**	**MAPE (%)**
Ridge regression	7.98×10^8	8.90	37.49	6.443
XGBoost	7.57×10^8	9.00	33.76	5.745
ARIMA	6.07×10^9	28.50	77.55	10.169

capture both short-term dependencies and medium-term trends, and their forecasting performance is summarized in Table 5.

Ridership

On the other hand, among all predictors of ridership shown in Table 6, the 7-day rolling mean shows the strongest positive effect, indicating that short-term ridership patterns over the past week strongly influence today's demand. In contrast, the 30-day rolling mean has a substantial negative effect, suggesting an inverse relationship between long-term ridership trends and current ridership levels. The one-day lag of ridership further confirms the strong autocorrelation in daily ridership, while additional lag features at the weekly and monthly scale also exhibit positive effects, revealing periodic fluctuations in travel demand.

Day-of-week indicators consistently show negative effects, implying that Mondays generally have the highest ridership, while other days tend to be lower by comparison. The declines are most pronounced on weekends, reflecting the typical reduction in commuter transit usage, whereas weekdays show smaller reductions.

Average Daily Temperature

In terms of the temperature forecasting model as shown in Table 7, the one-day lag of temperature represents the largest positive coefficient, which highlights a strong relation with daily temperature. The 30-day rolling mean of the temperature show a positive coefficient, reflecting a form of monthly dependency. The 7-day rolling mean has a smaller but still positive coefficient, and captures weekly patterns. Notably, the coefficient for the 1-month lag of temperature is negative, implying an inverse relationship between temperature levels today and those from exactly one month prior. Lastly, the 1-week lag of the temperature yields the smallest positive coefficient, indicating a modest contribution of weekly recurrence.

Table 6 Coefficients and interpretations of features in the ridership ridge regression model

Feature	Coefficient	Interpretation
7-day rolling mean	6.23×10^4	Average daily ridership over the past 7 days
30-day rolling mean	-5.57×10^4	Average daily ridership over the past 30 days
One-day lag of ridership	4.23×10^4	Ridership on the previous day
28-day lag of ridership	4.07×10^4	Ridership 28 days prior
1-week lag of ridership	3.58×10^4	Ridership 7 days prior
14-day lag of ridership	3.29×10^4	Ridership 14 days prior
Saturday	-3.04×10^4	Day-of-week dummy variable
Sunday	-2.35×10^4	Day-of-week dummy variable
Friday	-1.39×10^4	Day-of-week dummy variable
Thursday	-1.38×10^4	Day-of-week dummy variable
Wednesday	-1.31×10^4	Day-of-week dummy variable
Tuesday	-1.19×10^4	Day-of-week dummy variable

Table 7 Coefficients and interpretations of features in the average daily temperature ridge regression model

Feature	Coefficient	Interpretation
One-day lag of temperature	10.48	Temperature on the previous day
30-day rolling mean	3.07	Average daily temperature over the past 30 days
7-day rolling mean	1.43	Average daily temperature over the past 7 days
14-day lag of temperature	1.17	Temperature in 14 days prior
1-month lag of temperature	−0.97	Temperature in 30 days prior
1-week lag of temperature	0.29	Temperature in 7 days prior

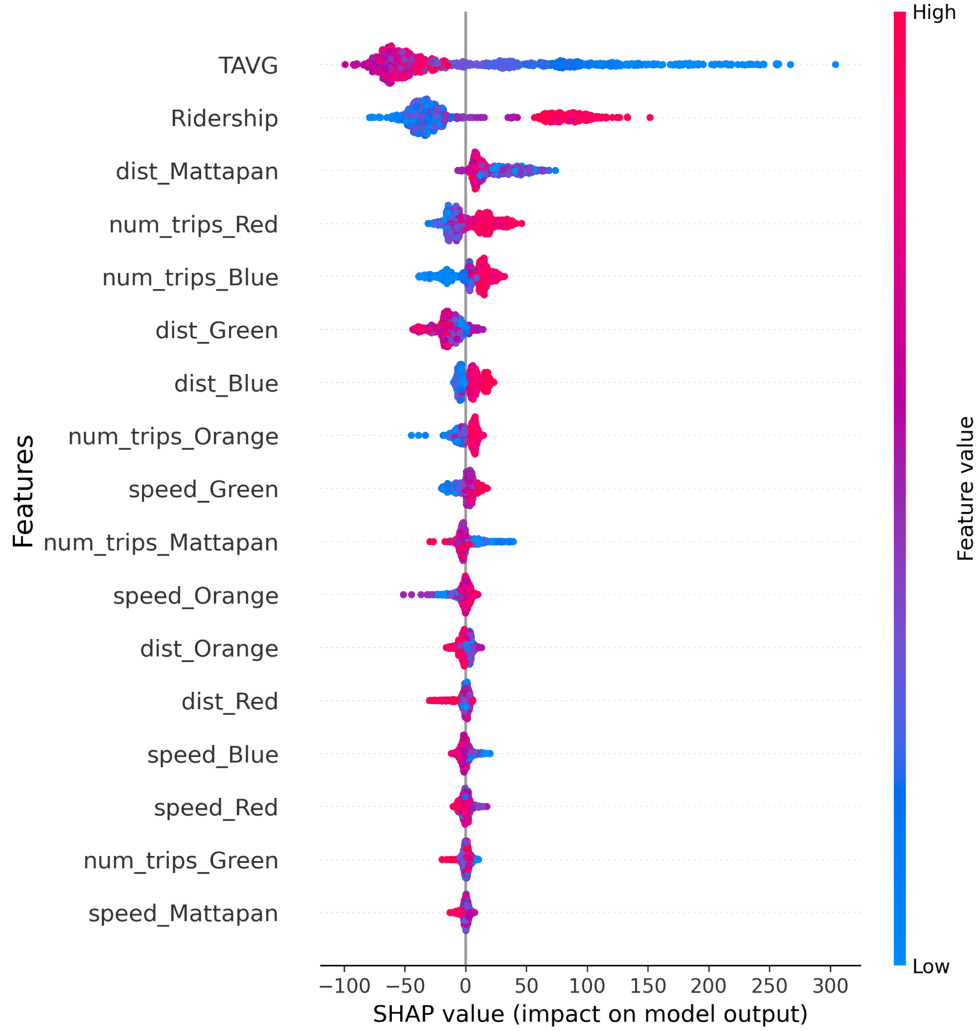

Fig. 9 SHAP plot of the XGBoost model

Sensitivity Analysis

Figure 9 identifies temperature and ridership as the two most influential predictors. Higher temperatures generally correlate with increased predicted energy consumption, suggesting temperature has a substantial positive impact on energy demand. Conversely, Ridership with both high and low values significantly influence model output in opposite directions.

The planning variables include distances and number of trips, indicating that operational variables related to trip frequency and distance contribute significantly to prediction variability. In contrast, variables such as speed across various lines show minimal impact, clustering near zero SHAP values, implying limited influence on energy predictions.

Planning Variables

The sensitivity of predicted energy consumption to variations in three key planning variables—average speed, service distance, and number of trips—across the five MBTA lines is shown in Fig. 10. For all lines, increasing the service and distance results in a sharp increase in energy consumption, particularly on the Red and Blue Lines. Changes in average speed exhibit more complex effects. For light rail lines—Mattapan and Green—decreasing speed significantly reduces energy consumption, likely due to lower acceleration demands and reduced auxiliary loads. In contrast, on heavy rail lines such as the Red Line, decreasing speed still reduces energy use but to a smaller extent. The Orange and Blue Lines show relatively flat or slightly increasing energy trends as speed increases, indicating different operating conditions or demand profiles.

The impact of number of trips varies across lines. Increasing trip frequency leads to increased energy use, especially for heavy rail lines (Red and Blue), consistent with higher total train movements. For light rail lines, particularly Mattapan, the sensitivity is much weaker, suggesting lower incremental energy costs per trip.

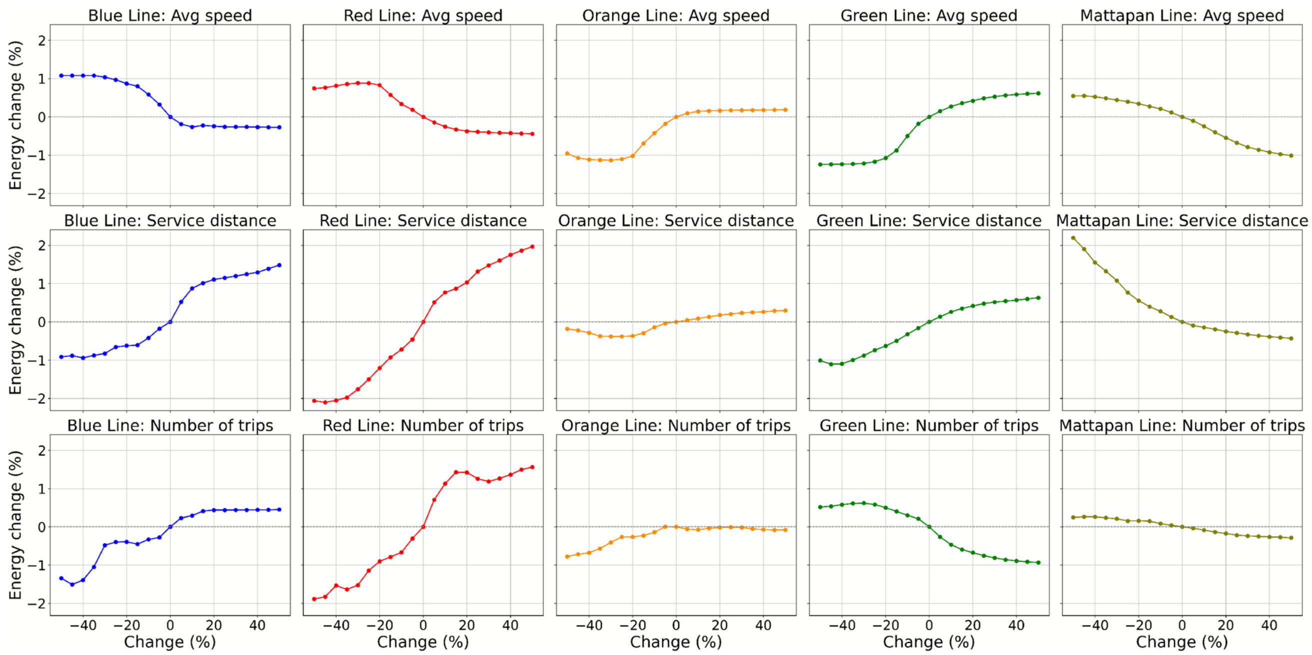

Fig. 10 Sensitivity of energy consumption to planning variables

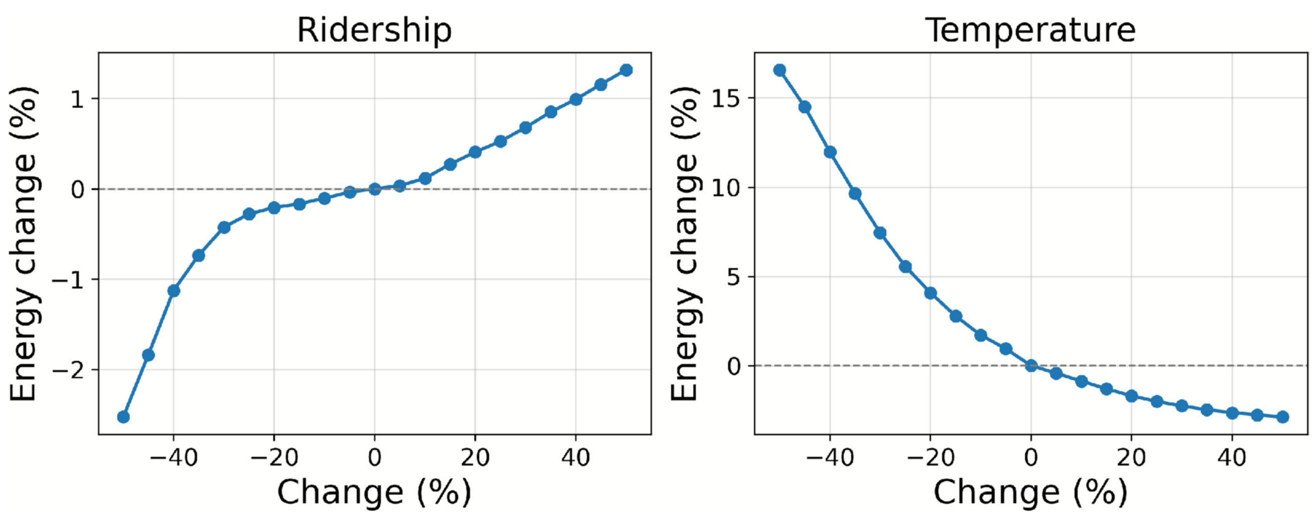

Fig. 11 Sensitivity of energy consumption to ridership and temperature

Ridership and Temperture

In terms of the sensitivity of predicted energy consumption to changes in ridership and average daily temperature as shown in Fig. 11. Cold temperature drives large increases in energy use, while warming yields only modest savings. A 40% reduction in temperature corresponds to roughly a 15–17% rise in energy, but even a 40–50% increase in temperature trims energy by only 2–3%. In contrast, ridership has a relatively small but positive influence on energy consumption. Increasing ridership slightly raises energy demand, while decreases in ridership produce only minor reductions.

Energy Evaluation of Service Plans

The base plan shows that cold days (LT) push energy well above the overall baseline, and the effect is amplified under higher ridership, with HR+LT corresponding to the darkest red cells in Fig. 13, indicating the highest increases. By contrast, hot days (HT) generally bring energy below baseline, with the reduction smallest when ridership is high, as seen under HR+HT where the cells are only light blue compared with the darker blue cells under LR+HT. At normal temperatures (NT), energy scales with demand: LR+NT is below baseline, NR+NT is near zero, and HR+NT is clearly above, which is visible in Fig. 13 as a gradual shift from blue to light gray to red across the middle rows. In short, temperature is the primary driver, with ridership acting as a secondary amplifier.

Across all ridership–temperature scenarios, Fig. 12 and Fig. 13 show that changing the number of trips and the operated distance produces the largest energy shifts; average speed has a smaller, second-order effect in this model. This is reflected in Fig. 13, where plans P3 and P5 consistently display the darkest red shades across most rows, indicating the strongest energy increases. In contrast, speed-focused plans, such as P2 (+20% speed), raise energy modestly, which appears in Fig. 13 as mostly light red cells, reinforcing that speed is a secondary lever compared to service volume.

Energy savings are clearest in plans that cut train-miles or cycles. P8 (–10% distance, –5% trips) shows blue cells across nearly all rows in Fig. 13, confirming broad energy reductions. P6 (–5% speed, –10% trips) also shows mostly blue but with lighter intensities, with the trip reduction again being the main driver. Figure 13 shows that trip cuts reliably translate into negative energy impacts, whereas speed reductions alone have a weaker effect.

Scenario comparisons further highlight the asymmetric influence of external conditions. Under LR+HT, all plans appear dark blue with reductions approaching –10%, while under HR+LT, all plans appear dark red with increases exceeding +16% (peaking near +18% in P8). Even under normal ridership and temperature (NR+NT), plan-level

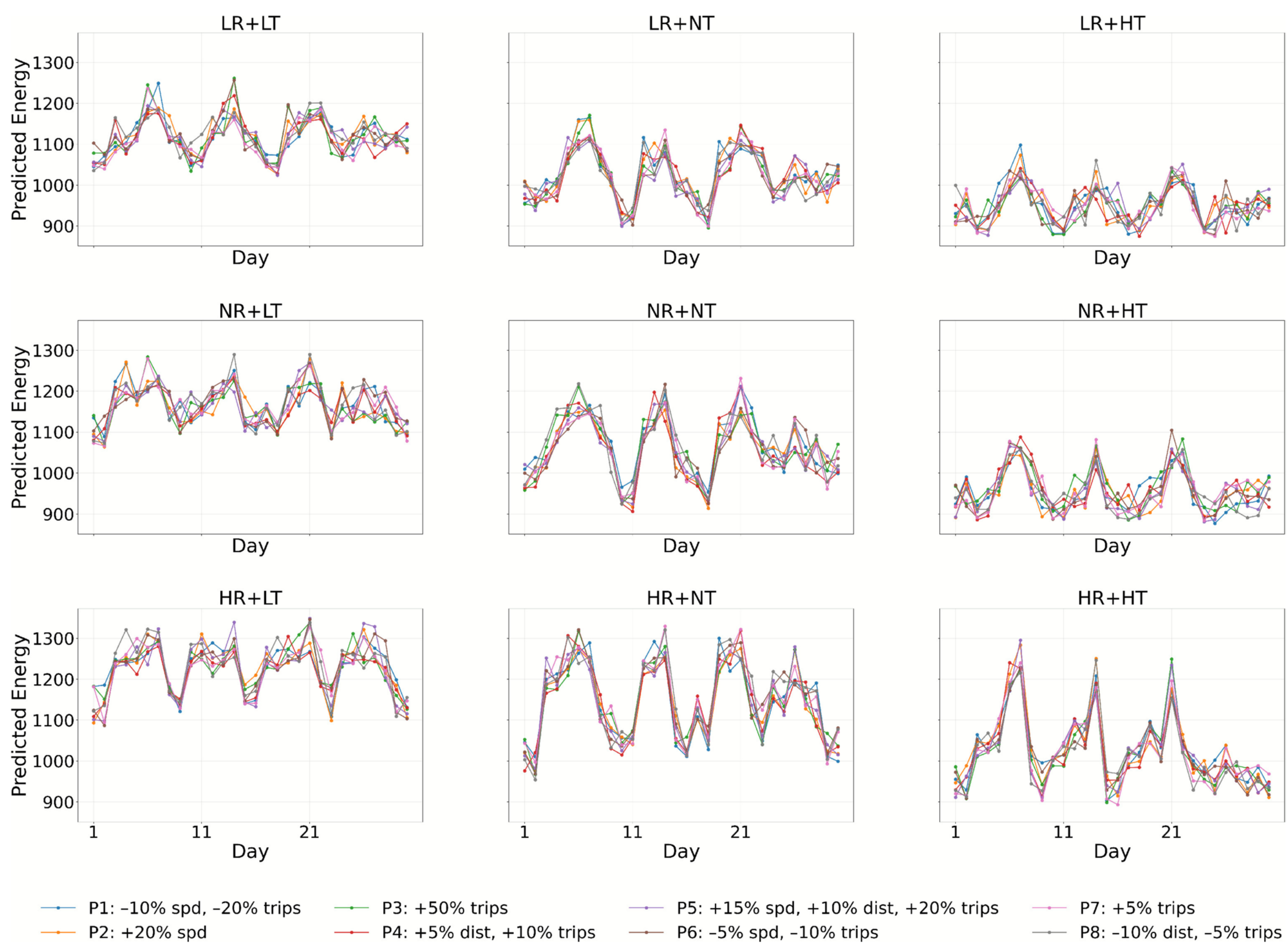

Fig. 12 The forecasting performance of predicted daily energy consumption across ridership–temperature scenarios. Each subplot displays 30-day forecasts under the eight operational plans (P1–P8)

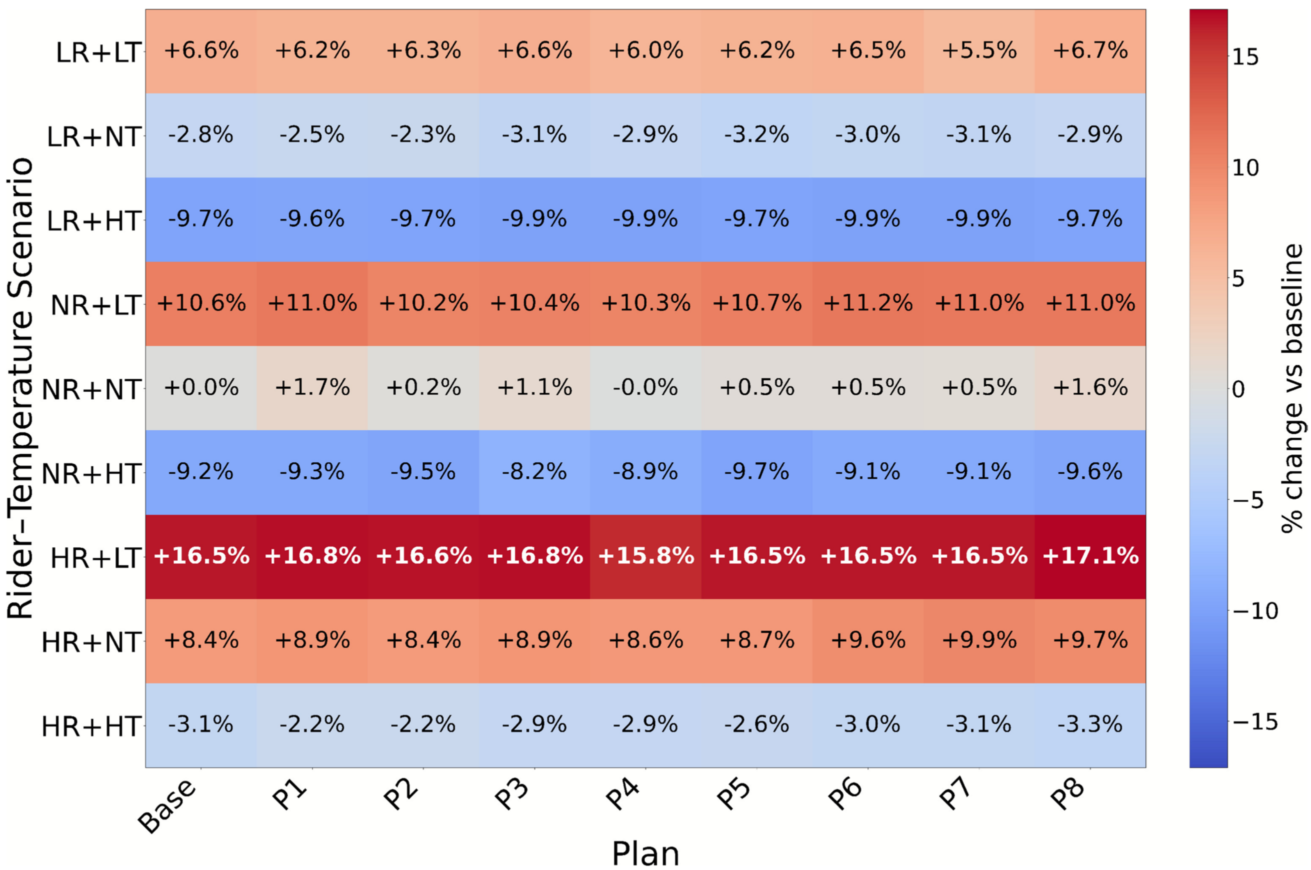

Fig. 13 A heatmap showing how the energy changes compared with the base plan

shifts remain small are shown as mostly gray cells in Fig. 13, confirming that planning interventions mainly shift the energy level rather than alter day-to-day patterns.

Limitations

This framework is developed based partly on MBTA operations during a period heavily influenced by COVID-19, when ridership and service patterns were atypical. As a result, the patterns of the estimated sensitivities may not fully generalize to post-pandemic conditions or to transit systems with different network layouts. Generalizing this framework to other systems would therefore require retraining the model using local operational and environmental data.

Conclusion

This paper presented a data-driven framework for forecasting and planning energy consumption in urban rail transit systems. Using the MBTA as a case study, we integrated planning metrics with observed operational data through a discrepancy–generation model and coupled these with short-term forecasts of temperature and ridership. Among the candidate prediction models, XGBoost achieved the best out-of-sample accuracy for daily energy consumption, while Ridge regression provided reliable forecasts of temperature and ridership. Together, these components enable multi-week scenario evaluation that links operational choices and external conditions to energy outcomes.

Sensitivity analysis and plan experiments highlight three central insights. First, temperature emerges as the dominant driver of variation: cold conditions push energy demand sharply above baseline, while equivalent warming yields only modest reductions. Second, service volume is the most effective operational control: increasing trips and operated distance raises energy use substantially, while trimming them reduces consumption across scenarios. Third, speed functions as a secondary lever for modest adjustments but less consequential than changes in running train numbers or traveling distance, with line-specific differences most evident on the lower-speed light-rails.

Overall, the framework offers a practical tool for evaluating operational strategies-such as late-night expansion, service frequency boosts, or capacity pushes-under varying ridership–temperature conditions. It also helps clarify trade-offs between access and energy efficiency. By bridging operational planning and energy forecasting, this approach advances the ability of transit agencies to manage demand sustainably, reduce costs, and improve resilience under future disruptions.

Appendix

We provide the specifications of the PDFs that were optimally fitted to the discrepancies $\epsilon_{ij\ell}$.

Cauchy distribution The probability density function (PDF) of the Cauchy distribution is given as:

$$f(x;\mu,\gamma) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x-\mu}{\gamma}\right)^2\right]} \tag{4}$$

where, μ represents the location parameter, which indicates the peak of the distribution, and γ is the scale parameter, representing the half-width at half-maximum.

Power Law distribution The PDF of the power law distribution is given by:

$$f(x;\alpha,x_0,\lambda) = \frac{\alpha\lambda^\alpha}{(x-x_0)^{\alpha+1}} \quad \text{for} \quad x > x_0 \tag{5}$$

where α is the shape parameter, x_0 is the location parameter indicating the minimum value for which the power-law behavior holds, and λ is the scale parameter.

Rayleigh distribution The PDF of the Rayleigh distribution is given by:

$$f(x;x_0,\sigma) = \frac{x-x_0}{\sigma^2} e^{-\frac{(x-x_0)^2}{2\sigma^2}} \tag{6}$$

where x_0 is the location parameter and σ is the scale parameter.

Exponential Power distribution The PDF of the Exponential Power distribution is given by:

$$f(x;\alpha,x_0,\lambda) = \frac{\alpha}{2\,\lambda\,\Gamma\left(\frac{1}{\alpha}\right)} \exp\left[-\left(\frac{|x-x_0|}{\lambda}\right)^\alpha\right], \tag{7}$$

where $\alpha > 0$ is the shape parameter, x_0 is the location parameter, and $\lambda > 0$ is the scale parameter.

Acknowledgements We would like to thank Dr. Hao Yin, Dr. Anil Gurcan and Michael Flanary, all of the Massachusetts Department of Transportation, for their inputs and support. Finally, we appreciate the administrative support of University of Massachusetts Transportation Center staff: Kimberley Foster, Matt Mann and Michelle Clark, during the course of the research.

Author Contributions The authors confirm their contribution to the paper as follows: study conception and design: JO, EC, EG, ZH; data collection: ZH; methodological implementation: ZH; analysis and interpretation of results: ZH, JO, EC, EG; draft manuscript preparation: ZH, JO. All authors reviewed the results and approved the final version of the manuscript.

Funding This study was undertaken as part of the Massachusetts Department of Transportation Research Program, funded by the Federal Highway Administration (FHWA) State Planning and Research (SPR) funds via the Massachusetts Department of Transportation. The authors are solely responsible for the facts, the accuracy of the data and analysis, and the views presented herein.

Data Availability All public data (GTFS, etc.) and code used in this study are available at https://github.com/narslab/mbta-energy-cast-dashboard.git. Energy and other data specific to MBTA are available upon reasonable request for the purpose of reproducing research.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This study did not require ethical approval as it did not involve any human or animal subjects.

References

Abbasabadi N, Ashayeri M, Azari R, Stephens B, Heidarinejad M (2019) An integrated data-driven framework for urban energy use modeling (UEUM). Appl Energy 253:113550. https://doi.org/10.1016/j.apenergy.2019.113550

Al-Musaylh MS, Deo RC, Adamowski JF, Li Y (2018) Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. Adv Eng Inform 35:1–16. https://doi.org/10.1016/j.aei.2017.11.002

APTA (2021) Ridership Report. Technical report, American Public Transportation Association, Washington DC

Barcellos de Paula L, Marins FAS (2018) Algorithms applied in decision-making for sustainable transport. J Clean Prod 176:1133–1143. https://doi.org/10.1016/j.jclepro.2017.11.216

Chenchen Z, Dongyin L, Xuemei X, Yanhui W (2022) Modeling and analysis of global energy consumption process of urban rail transit system based on Petri net. Journal of Rail Transport Planning & Management 21:100293. https://doi.org/10.1016/j.jrtpm.2021.100293

Davis SC, Boundy RG (2022) Transportation Energy Data Book Edition 40. https://tedb.ornl.gov/

Gu Q, Tang T, Cao F, Song Y-D (2014) Energy-Efficient Train Operation in Urban Rail Transit Using Real-Time Traffic Information. IEEE Trans Intell Transp Syst 15(3):1216–1233. https://doi.org/10.1109/TITS.2013.2296655

Hamurcu M, Eren T (2020) Strategic Planning Based on Sustainability for Urban Transportation: An Application to Decision-Making. Sustainability 12(9):3589. https://doi.org/10.3390/su12093589

Han Z, Gonzales E, Christofa E, Oke J (2022) Modeling System-Wide Urban Rail Transit Energy Consumption: A Case Study of Boston. Transportation Research Record, 03611981221096442. https://doi.org/10.1177/03611981221096442

Han Z, Gonzales E, Christofa E, Oke J (2023) Line-Specific Energy Modeling Framework for Urban Rail Transit Systems: A Case Study of Boston. Transportation Research Record,

03611981231170181. https://doi.org/10.1177/03611981231170181
Huang K, Liao F (2023) A novel two-stage approach for energy-efficient timetabling for an urban rail transit network. Transportation Research Part E Logistics and Transportation Review 176:103212. https://doi.org/10.1016/j.tre.2023.103212
Krishnan V, Kastrouni E, Pyrialakou VD, Gkritza K, McCalley JD (2015) An optimization model of energy and transportation systems: Assessing the high-speed rail impacts in the United States. Transportation Research Part C Emerging Technologies 54:131–156. https://doi.org/10.1016/j.trc.2015.03.007
Manzolli JA, rovão JP, Henggeler Antunes C (2021) Scenario-Based Multi-criteria decision analysis for rapid transit systems implementation in an urban context. eTransportation 7:100101. https://doi.org/10.1016/j.etran.2020.100101
National Oceanic and Atmospheric Administration. https://www.noaa.gov/
Santos AS, Ribeiro SK (2013) The use of sustainability indicators in urban passenger transport during the decision-making process: The case of Rio de Janeiro. Brazil Current Opinion in Environmental Sustainability 5(2):251–260. https://doi.org/10.1016/j.cosust.2013.04.010
Shang B, Zhang X (2013) Study of Urban Rail Transit Operation Costs. Procedia Soc Behav Sci 96:565–573. https://doi.org/10.1016/j.sbspro.2013.08.066
Tang Z, Yin H, Yang C, Yu J, Guo H (2021) Predicting the electricity consumption of urban rail transit based on binary nonlinear fitting regression and support vector regression. Sustain Cities Soc 66:102690. https://doi.org/10.1016/j.scs.2020.102690
To WM, Lee PKC, Yu BTW (2020) Sustainability assessment of an urban rail system - The case of Hong Kong. J Clean Prod 253:119961. https://doi.org/10.1016/j.jclepro.2020.119961
Wang D, Wu J, Wei Y, Chang X, Yin H (2024) Energy-saving operation in urban rail transit: A deep reinforcement learning approach with speed optimization. Travel Behaviour and Society 36:100796. https://doi.org/10.1016/j.tbs.2024.100796
Yang Y, Zhang P, Ni S (2014) Assessment of the Impacts of Urban Rail Transit on Metropolitan Regions Using System Dynamics Model. Transportation Research Procedia 4:521–534. https://doi.org/10.1016/j.trpro.2014.11.040
Zhao J, Deng W (2013) Fuzzy multiobjective decision support model for urban rail transit projects in China. Transport 28(3):224–235. https://doi.org/10.3846/16484142.2013.829119