Journal of Public Transportation 26 (2024) 100097

Contents lists available at ScienceDirect

Journal of Public Transportation

journal homepage: www.journals.elsevier.com/journal-of-public-transportation

A comparison of time series methods for post-COVID transit ridership forecasting

Ashley Hightower [a], Abubakr Ziedan [a,b], Jing Guo [c], Xiaojuan Zhu [d], Candace Brakewood [a,*]

[a] Department of Civil and Environmental Engineering, University of Tennessee, 851 Neyland Drive, Knoxville, TN 37996, United States
[b] Transportation Planner, CDM Smith, Atlanta, GA 30328, United States
[c] School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha, China
[d] Office of Information Technology, University of Tennessee, 821 Volunteer Boulevard, Knoxville, TN 37996, United States

ARTICLE INFO

Keywords:
Ridership
Heavy rail
COVID-19
Forecasting

ABSTRACT

Transit agencies conduct system-level ridership forecasting for planning, budgeting, and other administrative purposes. However, the COVID-19 pandemic introduced substantial changes in transit ridership levels and seasonal patterns, which has impacted the performance of ridership forecasting. Although time series methods are commonly used for forecasting transportation demand, they have received limited use in practice for public transit ridership forecasting. This study compares the performance of seven time series forecasting methods for predicting system-wide, monthly transit ridership for heavy rail agencies in the continental United States. The forecasting methods are: ETS, ARIMA, STL with ETS, STL with ARIMA, TBATS, a neural network, and a hybrid model. Ridership was forecasted for pre- and post-COVID periods (pre- and post- March 2020), as well as for the full series (January 2002 to December 2023). The MAPE and MASE were used to compare forecast performance. Using the pre-COVID period, 43% of the models produced a MAPE below 5% and 82% produced a MAPE below 10%. Using the full-series and post-COVID periods, only about 10% of the models produced a MAPE below 5% and half produced a MAPE below 10%. The classical and hybrid methods outperformed the other models using the full series period, and the TBATS, neural network, and hybrid methods outperformed the other methods using the post-COVID period. The findings suggest that even a few years into the post-COVID era, patterns that were typical of heavy rail ridership before the pandemic have not returned at most agencies in the United States, posing challenges to forecasting post-COVID ridership.

1. Introduction

Transit agencies need system-level ridership forecasts for various planning and administrative functions, such as creating their operating budgets, prioritizing long-term investments, and planning service frequencies (Boyle, 2006; Plotch, 2022) . Transit agencies historically have forecasted ridership using various methods such as professional judgement, rules of thumb, direct ridership models, and the four step travel demand model (Boyle, 2006). However, these forecasting methods might not be suitable in the wake of the COVID-19 pandemic due to its continued impacts on travel behavior. Therefore, transit agencies may want to consider different methods to forecast ridership. This study aims to explore the performance of one potential method: time series forecasting.

Time series forecasting is used in many disciplines including transportation. Prior studies have applied time series analysis to forecast demand in different transportation sectors including roadways, airlines, and freight (Babcock et al., 1999; Jayanthi and Jothilakshmi, 2021; Nihan and Holmesland, 1980; Washington et al., 2020; Zheng and Wei, 2020; Wickham, 1995). Time series analyses typically depend on understanding the data generating mechanism to predict future values. This feature gives time series methods an advantage over other methods in the wake of the COVID-19 pandemic, since the assumptions used for other methods (e.g. ridership elasticities that depend on the relationship between ridership and other variables) may not be consistent during or after COVID-19. Therefore, this study seeks to compare the performance of several classical and some newer, more complex time series methods to forecast ridership for heavy rail in the continental United States (US) in the wake of the COVID-19 pandemic. This study can help inform transit agencies as they explore various forecasting approaches that

* Corresponding author.
E-mail address: cbrakewo@utk.edu (C. Brakewood).

https://doi.org/10.1016/j.jpubtr.2024.100097
Received 30 November 2022; Received in revised form 23 May 2024; Accepted 11 June 2024
Available online 18 June 2024
1077-291X/

could handle uncertainty in the post-COVID era.

This paper proceeds as follows. First, a literature review that focuses specifically on the use of time series analysis for forecasting post-COVID transit ridership is presented. Next, the research questions are articulated. The data and method section is then introduced. Then, the results of the pre-COVID (before March 2020), full series (from January 2002 to December 2023), and post-COVID (after March 2020) analyses are presented. Last, conclusions and areas for future research are discussed.

2. Literature review

This section briefly discusses the findings of prior studies that used time series methods to explore transit ridership, particularly since the start of the COVID-19 pandemic. It should be noted that before COVID-19, numerous prior studies used time series forecasting methods such as autoregressive moving average (ARIMA) models (Kyte et al., 1988), exponential smoothing (Ge et al., 2013), and machine learning methods to forecast transit ridership (Azimian and Jiao, 2021; Chen et al., 2020; Ding et al., 2018; Egu and Bonnel, 2021; Ge et al., 2013; Kyte et al., 1988; Moghimi et al., 2022; Wang et al., 2018). Many pre-COVID studies have used a variety of other forecasting methods such as time series multiple regression models (Doi and Allen, 1986), ordinary least squares regression models (Mucci and Erhardt, 2018), and mode choice models (Chen and Naylor, 2011). Some pre-COVID studies have also explored the accuracy of transit forecasting and have generally found a tendency towards optimism bias, or forecasts that overestimate ridership (Hoque et al., 2023; Perry, 2017). While these and other prior studies forecasted transit ridership, they will not be further discussed since the focus of this section is on studies of ridership in the COVID era.

The first relevant study sought to forecast daily ridership from January 1, 2021 to May 31, 2021 at "L" train stations in downtown Chicago using a random forest model. The results showed that the daily ridership forecasts were accurate for the first month, but the performance of the forecasts decreased as the forecast horizon increased. The researchers also found that the average temperature had an impact on ridership, more so than the stay-at-home executive order (Azimian and Jiao, 2021).

The second relevant study is a 2022 publication that examined forecasting COVID-era transit ridership using daily turnstile data from January 2019 to July 2020 for the New York City subway by producing a non-stationary time series model. The model used several independent ARIMA models to bridge between break points. In the first part of the method, an ARIMA model was applied to the entire dataset, ignoring the break points. In the second part, the ridership data were split into multiple datasets at the breakpoints, and ARIMA models were applied to each dataset. The piecewise ARIMA model used in the second part of the paper was found to have performed substantially better than the single ARIMA model (Moghimi et al., 2022).

A third relevant study predicted the impacts of the COVID-19 pandemic on US transit ridership in the five most populous cities in the US (New York City, Los Angeles, Chicago, Houston, and Philadelphia) using monthly transit ridership data and epidemic data from March 2020 to April 2022. The researchers used the ARIMA and Prophet models to forecast ridership given COVID infections. The forecasting accuracy of the ARIMA model was found to be between 6% and 10%, and the forecasting accuracy of the Prophet model was found to be between 8% and 12%. The researchers stated that the insufficient stationarity, periodicity, and seasonality of the time series contributed to the difficulty in refining the Prophet model (Gao et al., 2023).

A fourth relevant study developed a neural architecture and feature search approach to automatically fine-tune the architecture and features of a deep neural network to predict any given route-level public transit ridership. The approach utilized a randomized local hyper-parameter search that searched for a set of parameters that would minimize both the prediction errors and the model complexity. The approach was tested on automatic passenger count data from five routes of the Chattanooga Area Regional Transportation Authority, from January 2019 to October 2021. The study found that optimizing neural networks for each route significantly improved their performance compared to generic neural networks (Ayman et al., 2022).

One last relevant study produced real-time, one-day ahead forecasts at 147 bus rapid transit stations in Bogota, Columbia using data from August 2015 to May 2021, during which time there existed highly dynamic and uncertain conditions (COVID-19 and a month-long protest). The researchers developed an open-source infrastructure with five forecasting methods: ARIMA, seasonal ARIMA, multilayer perceptron (MLP), convolutional neural networks (CNN), and long-short-term memory (LSTM). The models were applied using two learning strategies: static (estimated using only the test data) and online (model was re-trained as new data became available). Under stable conditions, all models had similar performance, although the static CNN model consistently underperformed, and the online LSTM outperformed the other models. For COVID-19 conditions, the online LSTM outperformed the other models. All models performed substantially worse under the dynamic conditions compared to the stable conditions. The researchers also found that it took the machine learning models one and a half months to learn the new demand patterns (Caicedo et al., 2023).

This brief review shows that time series methods have been used in a limited number of studies to forecast transit ridership in the wake of the COVID-19 pandemic. However, these prior studies did not compare the forecasting performance of time series methods for all of the heavy rail agencies in the continental US, which could provide insight into the general suitability of time series forecasting for heavy rail, as well as the post-COVID ridership patterns on heavy rail. This study aims to compare the performance of several different time series forecasting methods for pre-COVID and post-COVID transit ridership for all heavy rail agencies in the continental US. The results of this study could inform transit agencies on which forecasting methods may perform best with minimal complexity, especially for system-level, post-COVID, heavy rail forecasts.

3. Research questions

The research questions are as follows:

1. Which of the seven time series forecasting methods (ETS, ARIMA, STL-ETS, STL-ARIMA, TBATS, neural network, or hybrid), if any, produce acceptable forecasts for monthly system-level heavy rail ridership data, pre- and post-COVID?
2. Does the period of time used to train the models impact the performance of post-COVID heavy rail ridership forecasts?
3. Are there still substantial differences in pre- and post-COVID heavy rail ridership patterns?

4. Data and method

In this section, the data source and methods used in this study are discussed in four parts. First, the ridership data from the different transit agencies used in this study are explained, and the historical context of recent ridership changes is briefly discussed. Next, information regarding the software and code used to develop the models is provided. Then, the time series methods used in this study are described. Last, the measures used to evaluate the forecasting performance are discussed.

4.1. Data and period of analysis

Monthly ridership data in terms of unlinked passenger trips (UPT) were downloaded from the National Transit Database website (Federal Transit Administration, 2023). The dataset used in this study included ridership for heavy rail in the continental US. The fourteen agencies used in this study are shown in Table 1. In Table 1, the minimum, maximum, and average monthly unlinked passenger trips on heavy rail are shown

Table 1
Heavy rail agencies in the continental US showing minimum, maximum, and average unlinked passenger trips (UPT) pre- and post-COVID.

Agency Name	Pre-COVID (before March 2020)			Post-COVID (after March 2020)			Recovery in Ridership from Dec 2019 to Dec 2023		
	Mean	Min	Max	Mean	Min	Max	Trips in Dec 2019	Trips in Dec 2023	2023 / 2019
Chicago Transit Authority	17,568,046	13,265,040	22,663,990	7,386,012	2,235,078	11,509,844	15,923,640	8,766,560	55.1%
County of Miami-Dade	1,532,757	1,079,749	1,996,221	902,467	287,803	1,239,336	1,433,928	1,177,272	82.1%
Los Angeles County Metropolitan Transportation Authority	3,599,656	616,954	4,629,231	1,934,341	1,132,356	2,528,557	3,401,148	1,724,702	50.7%
Maryland Transit Administration	1,058,714	124,877	1,367,600	185,900	84,936	352,425	579,668	293,169	50.6%
Massachusetts Bay Transportation Authority	12,876,889	9,359,700	16,679,922	5,686,246	1,182,140	7,882,855	11,545,872	6,397,500	55.4%
Metropolitan Atlanta Rapid Transit Authority	6,027,105	4,401,906	7,982,627	2,095,629	993,578	2,906,653	4,428,131	2,410,636	54.4%
MTA New York City Transit	205,182,028	148,267,939	253,609,943	125,865,343	20,254,269	183,903,625	236,357,677	173,256,395	73.3%
Port Authority Trans-Hudson Corporation	6,625,755	2,687,713	8,620,158	3,242,174	414,576	5,236,791	7,080,358	4,561,550	64.4%
Port Authority Transit Corporation	839,383	663,580	1,044,748	346,943	83,152	500,393	895,747	451,425	50.4%
San Francisco Bay Area Rapid Transit District	9,674,381	7,203,696	12,556,439	2,936,377	656,653	4,845,494	9,309,469	3,919,728	42.1%
Southeastern Pennsylvania Transportation Authority	7,774,345	5,432,593	9,821,204	3,742,342	1,263,591	5,220,404	7,717,104	4,010,239	52.0%
Staten Island Rapid Transit Operating Authority	611,521	368,986	828,475	284,901	35,979	506,733	604,767	412,352	68.2%
The Greater Cleveland Regional Transit Authority	505,912	223,497	851,608	229,188	159,213	328,061	460,306	305,495	66.4%
Washington Metropolitan Area Transit Authority	21,884,752	16,023,750	27,860,355	7,077,317	1,182,655	13,872,488	17,783,221	10,814,219	60.8%

Data Source: National Transit Database, December 2023 Adjusted Database (Federal Transit Administration, 2023)

for pre- and post-COVID periods. The percent recovery in ridership as of the last observation available at the time of analysis (December 2023) was also found; ridership recoveries were found to range from about 40–80%.

Prior to the pandemic, ridership at many agencies in the US was already declining, but the onset of the pandemic led to dramatic

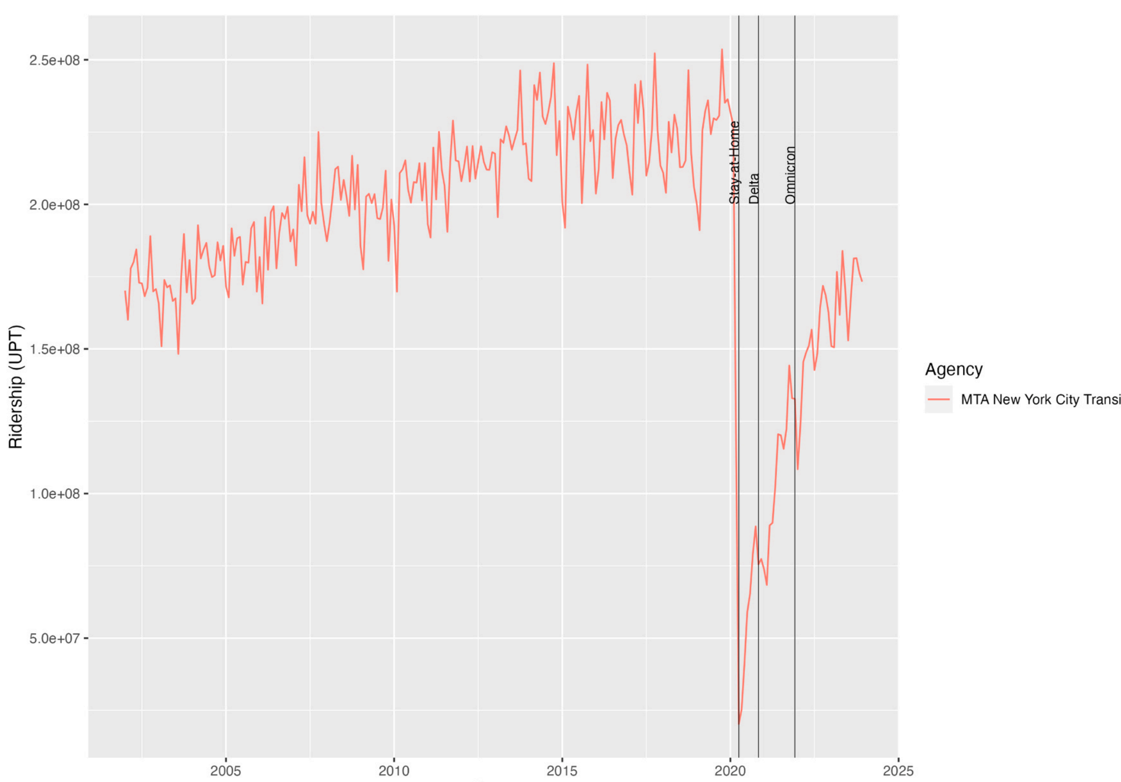

Fig. 1. Ridership at MTA New York City Transit from January 2002 to December 2023.

decreases in ridership. The effects of the pandemic on heavy rail ridership were first apparent in March of 2020, which was when many American states and cities initially implemented a mandatory or advisory stay-at-home order (Moreland et al., 2020). The pre-pandemic downward trend and the effect of the pandemic on transit ridership can be seen in Fig. 1 (for the largest heavy rail agency in the US), Fig. 2 (for nine mid-sized heavy rail agencies), and Fig. 3 (for the four smallest heavy rail agencies). The agencies are placed on separate plots according to the scale of their ridership so that the trends and ridership changes are clearly visible. At most agencies there is a clear decrease in ridership starting approximately around 2015, although the ridership decline started earlier at some agencies. An exception to the trend is the agencies in and around the New York City area; the ridership levels at these agencies appear to have had minimal decreases. However, all agencies had dramatic ridership decreases in March and April of 2020. From February to March of 2020, ridership declines at the heavy rail agencies in this study ranged from 32% to 55%, and from February to April ridership declines ranged from 55% to 94%. Therefore, February 2020 was considered to be the last pre-COVID month, and April 2020 was considered to be the first COVID-era month. The periods of analysis were selected based on these findings.

Models with multiple time periods were tested in this study in order to: 1) compare forecasting performance pre- and post-COVID, 2) investigate the impact of the number of observations on the forecasting performance, and 3) isolate the effects of the pandemic on post-COVID ridership trends and patterns. Fig. 4 shows a visualization of the three time periods selected for analysis. In each time period, the last 12 observations were held out to be used as testing data. The first time period used in this study was the pre-COVID period. The training data were from January 2002 to February 2019, and the test data were from March 2019 to February 2020. This time period was used to establish a baseline performance for the seven forecasting methods under "normal" conditions. The second time period was the full data series. The training data were from January 2002 through December 2022, and the testing data were from January 2023 to December 2023. This time period was used to forecast post-COVID ridership while incorporating pre-pandemic ridership trends. The third time period was the post-COVID period. The training data were from April 2020 through December 2022, and the testing data were from January 2023 to December 2023, which was the latest available data point at the time of analysis. This time period was used to evaluate the performance of the forecasting methods under the "new normal" conditions, without considering prior patterns in transit ridership. All forecasts were made for the following 12 months in order to compare the forecasted ridership to the actual ridership from the testing datasets.

4.2. Software

The software used in this paper initially was based on a web application called Route Trends and Info created by staff from Minneapolis Metro Transit (Minneapolis Metro Transit, N.D.). The application runs off R Software (R Core Team, N.D.). Metro Transit's application supports seven forecasting methods: 1) exponential smoothing (ETS); 2) autoregressive integrated moving average (ARIMA); 3) and 4) Seasonal and Trend Decomposition using LOESS, using ETS or ARIMA (STL-ETS and STL-ARIMA); 5) the exponential smoothing state space model with trigonometric seasonality, Box-Cox transform, ARMA errors, trend, and seasonal periods (TBATS); 6) a feed-forward autoregressive neural network with a single hidden layer (NNET); and 7) a hybrid model with equal weights of the ETS, STL, TBATS, and NNET methods(Minneapolis Metro Transit, N.D.; Hyndman and Athanasopoulos, 2018d).

The source code for the application is available to the public on GitHub (Lind et. al., N.D.). The source code was modified substantially by the authors of this paper and used directly in R (v4.3.1) in order to change the model parameters, output additional performance measures (e.g., mean absolute scared error), and output customized graphs (R Core Team, N.D.). For example, Minneapolis Metro's application uses the ETS method in their hybrid model, but in this study, the ARIMA

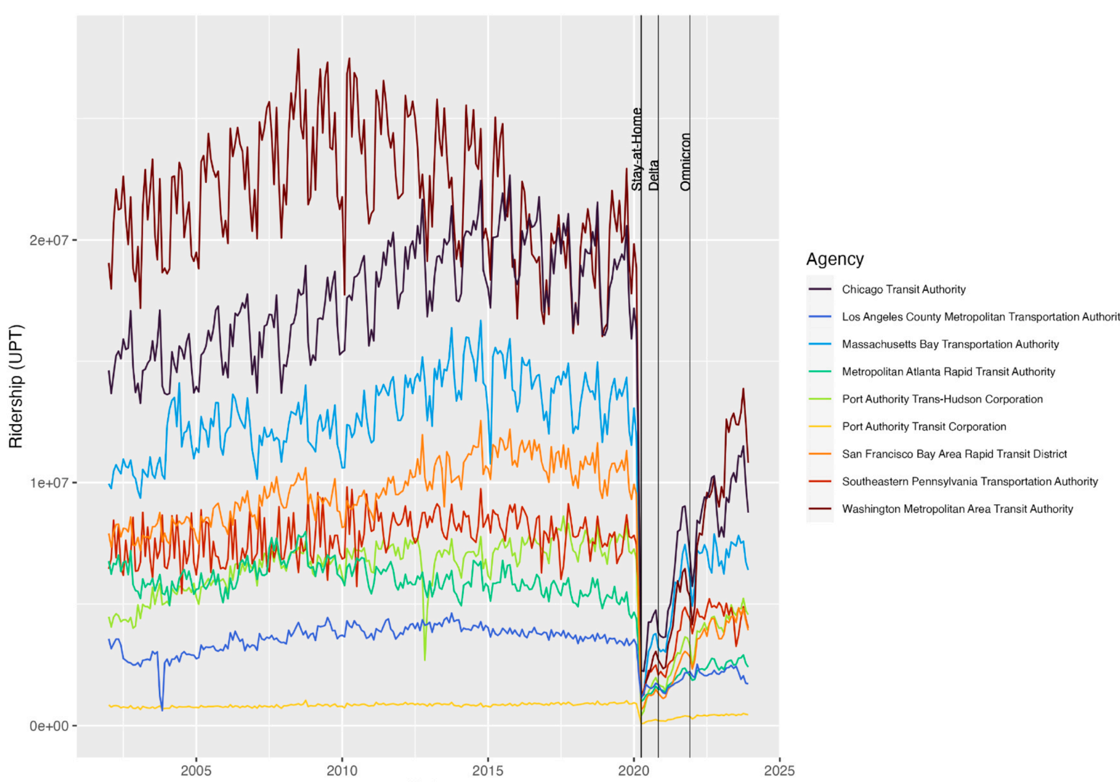

Fig. 2. Ridership at larger heavy rail agencies from January 2002 to December 2023.

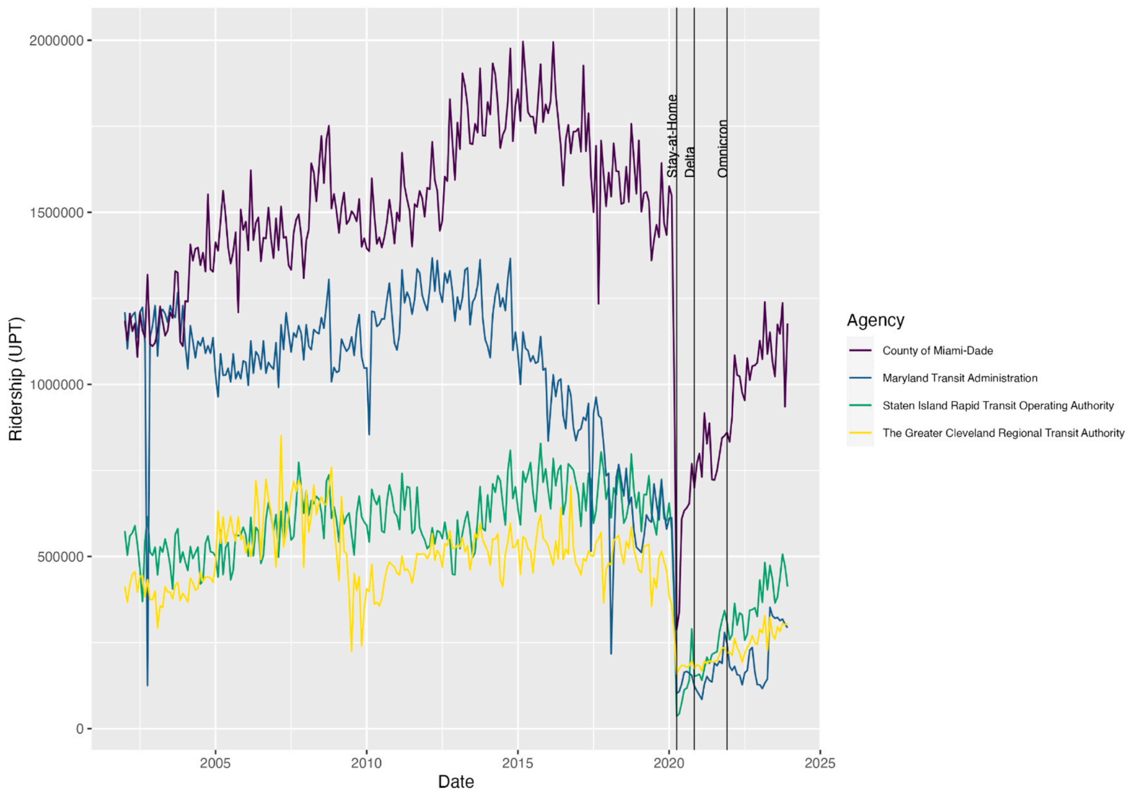

Fig. 3. Ridership at smaller heavy rail agencies from January 2002 to December 2023.

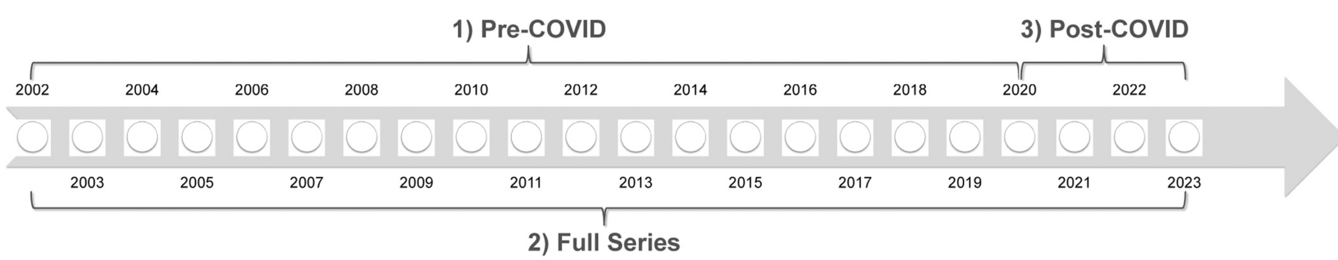

Fig. 4. The three forecasting time periods used in this study.

method was used because it decreased the mean absolute percent error of many of the post-COVID forecasts by over one percentage point. Also, the original GitHub code produces a single interactive plot using the "dygraphs" package (Vanderkam et al., 2018), but these were changed to non-interactive plots with seven graphs using the "ggplot2" package (Wickham, 2016). The functions used to produce the forecasts in this study have been made publicly available on Github at the following web address: https://github.com/tscore-utc/time-series_forecasting.

4.3. Time series models

Seven time series models were estimated in this study: ETS, ARIMA, STL-ETS, STL-ARIMA, TBATS, NNET, and a hybrid model with equal weights of the ARIMA, NNET, TBATS, and STL methods (Hybrid ANST). This study focused on the first four classical time series models because they are widely established, well understood, and familiar to academics and some practitioners. The last three models were selected due to their ability to handle complex seasonality in time series using non-linear functions. This study does not focus on other machine learning methods or deep learning methods for several reasons. First, the short length of the data used in this study is not likely to be suitable to such methods. Second, to the authors' knowledge, no US transit agency is using other machine learning or deep learning methods to conduct ridership forecasts, and this study is intended to be aimed at practitioners. Last, machine learning methods, especially deep learning methods, are expensive to conduct, require extensive set up, and the results are more difficult to interpret.

Each of the chosen forecasting methods are briefly discussed below. The equations have been omitted for brevity but interested readers are referred to the following references (Hyndman and Athanasopoulos, 2018d; Livera et al., 2011).

4.3.1. Exponential smoothing (ETS)

ETS is the exponential smoothing state space model; this method is computationally simple but the least robust among the methods used in this paper. There are 18 potential equations for an ETS model. A general notation can be denoted as ETS(error, trend, seasonal), and each component has multiple possibilities. Errors can be "A" additive or "M" multiplicative. Trends can be "N" none, "A" additive, or "Ad" additive damped. Seasonal components can be "N" none, "A" additive, or "M" multiplicative. By using the "ets()" function in the "forecast" package, the best fit model was selected from the 18 options by minimizing the corrected Akaike information criterion (AICc) (Hyndman and Athanasopoulos, 2018a; Hyndman and Khandakar, 2008; Hyndman et al., 2021). For more details on the ETS method, readers are referred to (Hyndman and Athanasopoulos, 2018d).

4.3.2. Autoregressive integrated moving average (ARIMA)

ARIMA is a state space model that is relatively computationally simple but assumes linear data to generate forecasts. There exist a potentially infinite number of possible equations for an ARIMA model. A general notation for a non-seasonal ARIMA model is ARIMA (p, d, q). The “p” is the order of the autoregressive (AR) part, the “d” is the degree of first differencing (the integrated (I) part), and the “q” is the order of the moving average (MA) part. A general notation for a seasonal ARIMA model is ARIMA(p, *d*, q)(P, D, Q)$_m$, where “P,” “D,” and “Q” are the seasonal AR, differencing, and MA orders, respectively, and “m” is the number of observations per year. In this study, the number of observations per year is always 12. Using the “auto.arima()” function from the “forecast” package with the argument “stepwise = FALSE,” R automatically chose the best fit ARIMA model based on a combination of the following criteria: unit root tests, minimizing the AICc, and minimizing the maximum likelihood estimation (Hyndman and Khandakar, 2008; Hyndman and Khandakar, 2021; Hyndman and Athanasopoulos, 2023b). For more details on the ARIMA method, readers are referred to (Hyndman and Athanasopoulos, 2018d).

4.3.3. Seasonal and trend decomposition using locally estimated scatterplot smoothing (LOESS) (STL) with ETS or ARIMA

STL is a method to decompose a time series into the seasonal, trend-cycle, and remainder components, where LOESS is a method of evaluating nonlinear relationships. The STL method allows the seasonal component to change over time and can be robust to outliers, but it does not automatically handle calendar variation. In this study, the STL method was used to de-seasonalize the data. This is useful for transit ridership data because transit ridership tends to vary substantially by season. The seasonal component “S_t” is forecasted separately from the non-seasonal components “T_t” and “R_t” using the seasonal naïve method. A seasonal naïve forecast means all forecasts were set equal to the last observed value from the same season (Hyndman and Athanasopoulos, 2018d). The non-seasonal components can then be forecast using any non-seasonal forecasting method (Hyndman and Athanasopoulos, 2018d). The “stlm()” function from the “forecast” package was used to forecast using the STL-ETS and STL-ARIMA methods (Hyndman and Athanasopoulos, 2018b; Hyndman and Khandakar, 2008; Hyndman et al., 2021). For more details about STL decomposition, readers are referred to (Hyndman and Athanasopoulos, 2018b).

4.3.4. Exponential smoothing state space model with trigonometric seasonality, Box-Cox transform, autoregressive moving average errors, trend, and seasonal periods (TBATS)

The TBATS model is an exponential smoothing state-space model capable of modelling complex seasonality; this method also allows the seasonal component to change slowly over time. TBATS can consider a Box-Cox transformation to handle non-linearity in the data, the autoregressive moving average model to de-correlate the residuals, trend components for long- and short-term trend (with and without dampening), and trigonometric functions based on Fourier representations with time-varying components to handle complex seasonality (Livera et al., 2011). By using the “tbats()” function in the “forecast” package, R selected the best fit model by adjusting the above parameters to find a model configuration that minimized the Akaike information criterion (AIC) (Razbash and Hyndman, N.D.). This method has a long computation time relative to the other state space models used in this study. For more details on the TBATS model, readers are referred to (Livera et al., 2011).

4.3.5. Feed-forward autoregressive neural network with a single hidden layer (NNET)

The neural network used in this study is capable of handling data with complex nonlinear relationships. R’s nnetar() function from the “forecast” package was used to estimate a feed-forward autoregressive neural network with a single hidden layer (Hyndman et al., 2021). In order to ensure reproducibility, the seed was set to “1234”; setting a seed was required for this study because the neural network randomly selects the initial parameters, which caused the performance measures for the same forecasts to vary slightly from run to run (Hyndman, 2017). The seasonal model is represented by NNAR(p,P,k)$_m$, where “P”, the number of seasonal lags, defaults to one, “p”, the number of non-seasonal lags, is selected from the optimal linear model fitted to the seasonally adjusted data, “k”, the number of neurons in the hidden layer, is equal to $(p + P + 1)/2$, and “m” is the number of observations per year (always 12 in this study). Prediction intervals were derived for the neural network forecasts by setting the forecast() function to simulate 1000 future sample paths (Hyndman and Athanasopoulos, 2018c; Hyndman, 2017). With a larger data set, the performance of the forecasts may be improved when “P,” the number of seasonal lags, is set to a context-appropriate value (e.g., 12 for monthly data). However, the short length of the data in this study caused the neural network to overfit the model when P was set to 12, so the default of one was used. For more details on the neural network model, readers are referred to (Hyndman and Athanasopoulos, 2018c).

4.3.6. Hybrid

Hybrid models produce forecasts by combining several methods to produce a single forecast result. The hybridModel() function from the “forecastHybrid” package was used to estimate a hybrid model with equal weights of the ARIMA, NNET, TBATS, and STL methods. Minneapolis Metro’s hybrid model uses ETS instead of ARIMA; for this study, the ARIMA model was selected instead because it improved the performance of many the post-COVID forecasts by decreasing the value of the mean absolute percent error by over one percentage point. The method for the STL component of the hybrid model was ARIMA. Parallel processing was used between models to shorten the computation time. The weights of each model were set to be equal because this has been shown empirically to give good performance. The hybrid method has the longest computation time among these seven methods. For more details on the hybrid model, readers are referred to (Shaub, N.D.).

4.4. Performance measures

The performance measures used in this study were the mean absolute percent error (MAPE) and the mean absolute scaled error (MASE) of the testing data. The MAPE is a percentage error and is unitless. Therefore, the MAPE is a good measure to compare between datasets that have different scales. The MASE is a scaled error and can be used to compare between datasets with different units. A MASE less than one indicates that, according to the mean absolute error (MAE), the forecast performed better than the average naïve forecast. In the case of seasonal data, a naïve forecast means all forecasts were set equal to the last observed value from the same season (Hyndman and Athanasopoulos, 2018d). The MAPE and MASE for the testing data were recorded to provide a measure of the forecast accuracy (Hyndman and Athanasopoulos, 2018d). The formula for the MAPE is provided in Eq. (1). The formula for the MASE for a seasonal forecast is given in Eq. (2).

$$MAPE = mean\left(\left|\frac{100e_t}{y_t}\right|\right) \quad (1)$$

Where

t = time (in months)

y_t = the fitted values (ridership in UPT)

e_t = the forecast error (the difference between an observed value and its forecast) (Hyndman and Athanasopoulos, 2023a).

$$MASE = mean\left(\left|\frac{e_j}{\frac{1}{T-m}\sum_{t=m+1}^{T}|y_t - y_{t-m}|}\right|\right) \tag{2}$$

Where
t = time (in months)
m = number of observations per year (always 12 in this study)
y_t = the fitted values (ridership in UPT)
$y_{t\text{-}m}$ = the fitted values from the previous season (ridership in UPT)
T = time at the last observation
e_j = the forecast error (the difference between an observed value and its forecast) (Hyndman and Athanasopoulos, 2023a).

5. Results and discussion

The results of the pre-COVID (from January 2002 to February 2020), full series (from January 2002 to December 2023), and post-COVID (from April 2020 to December 2023) analyses are presented in this section. The results are shown in this order to demonstrate the impact of the pandemic on the forecast accuracy, and to show the impact of the series length on the models' performance. For all periods, each of the 14 heavy rail agencies was analyzed using the seven forecasting methods, and the MAPE and MASE were calculated.

To evaluate the fit of the models, the Ljung-Box test was used to assess the white noise of the residuals for each model, and histograms of the residuals were examined to check for normality and influential outliers. The residuals were checked using the "checkresiduals()" function from the "forecast" package (Hyndman and Khandakar, 2008; Hyndman et al., 2021). For the majority of the pre-COVID models, the histograms were normally distributed with the means close to zero, i.e., the models were unbiased. The Ljung-Box test showed that for 51% of the models, some autocorrelation patterns remained that the models were not able to capture; this means that while the forecasts could be quite good, the prediction intervals may have been inaccurate. For the full series forecasts, the histograms were normally distributed with the means close to zero, and 73% of the results from the Ljung-Box test showed that the residuals were white noise. However, some of the models, especially the ETS and neural network methods, were not able to capture all the patterns in the data, meaning that while the these forecasts could be quite good, the prediction intervals may have been inaccurate. For the post-COVID-only series, the histograms were approximately normally distributed with the means close to zero, and the Ljung-Box test showed that 80% of the models captured all available patterns in the data, leaving the residuals with only white noise. The residuals therefore showed that the models generally produced unbiased results and captured most of the patterns in the data for each time period (Hyndman and Athanasopoulos, 2018d). Even though the pre-COVID residuals showed some remaining patterns, the results below show that the overwhelming majority of the models still produced good or acceptable forecasts.

5.1. Results of the pre-COVID analysis

For the pre-COVID analysis, the training dataset was from January 2002 through February 2019. The testing dataset was from March 2019 through February 2020. Visualizations of the forecasts for the transit agencies in Chicago and Cleveland are provided in Fig. 5 and Fig. 6 to show examples of the outputs that were created for each of the fourteen agencies. In the figures, the x-axis is the date, and the y-axis is ridership (UPT). The blue line is the forecasted ridership, and the red line is the actual ridership (testing data). The light purple shaded area is the 95% prediction interval, and the dark purple shaded area is the 80% prediction interval. Although the data was trained from 2002 to 2019, only the training data after 2010 is shown so that both the forecasts and the trends are shown more clearly. At many agencies, for example at the Chicago Transit Authority, the graphs showed clear seasonal patterns and a downward trend prior to the pandemic. Other agencies, such as the Greater Cleveland Regional Transit Authority, had weaker seasonal patterns and had a less dramatic downward ridership trend.

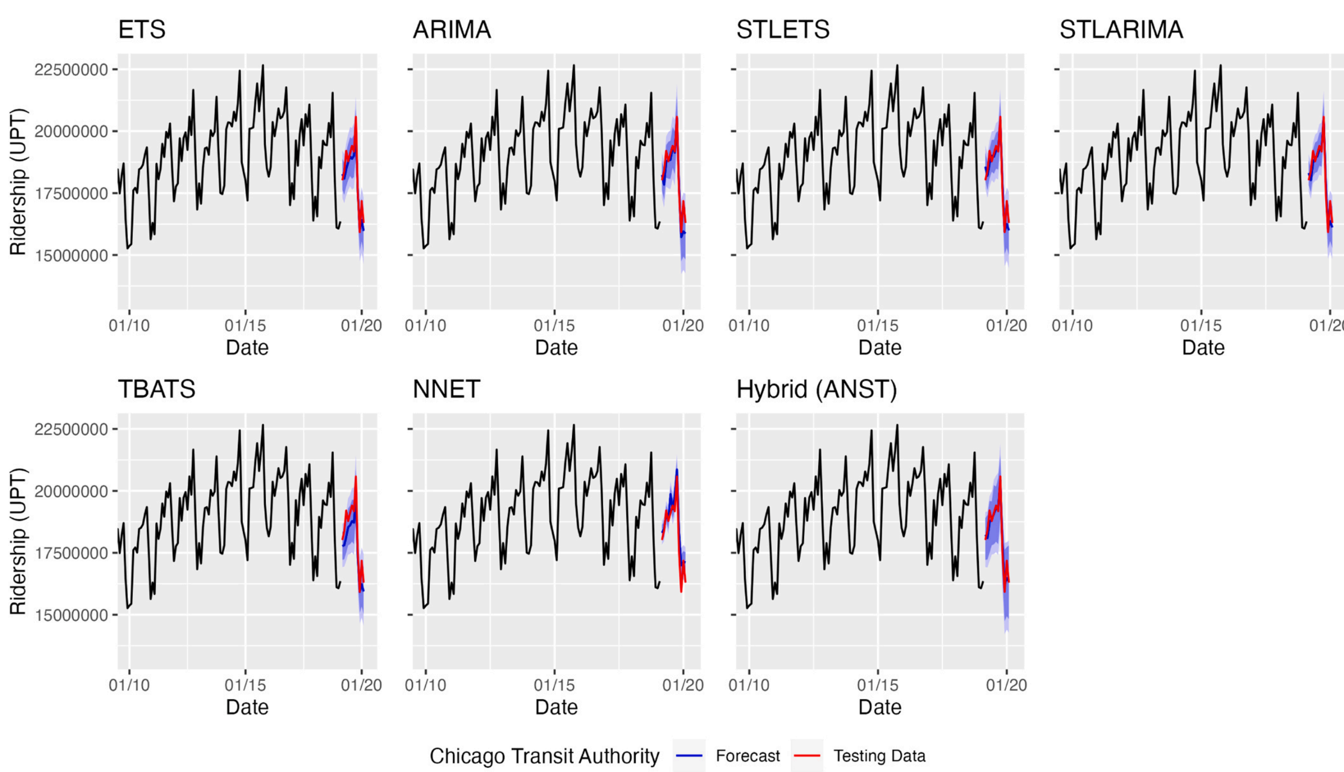

Fig. 5. Pre-COVID ridership forecasts for the Chicago Transit Authority.

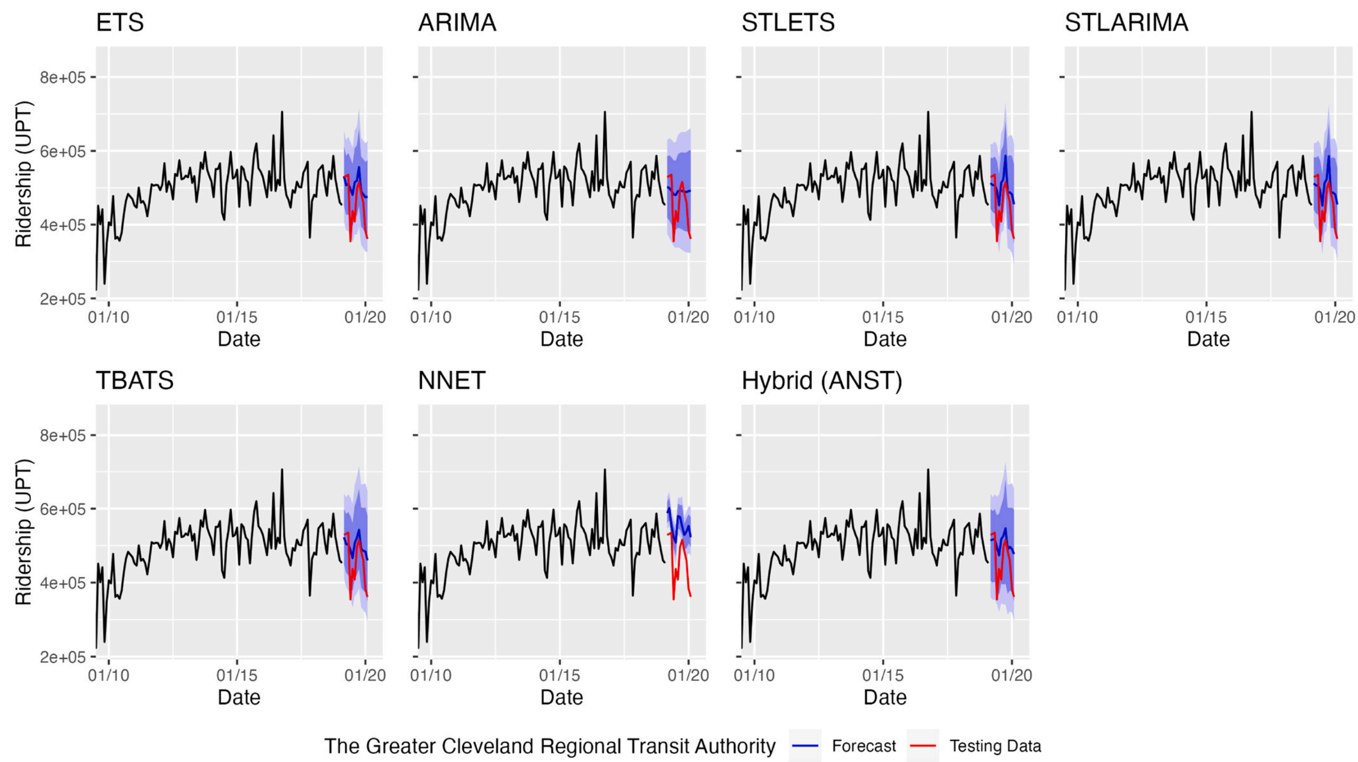

Fig. 6. Pre-COVID ridership forecasts for the Greater Cleveland Regional Transit Authority.

Table 2 presents the results of the pre-COVID model performance for each agency. The lowest/highest MAPE and MASE for the testing data (referred to simply as "errors") are shown, with the highest error indicating the poorest performance and the lowest error indicating the best performance. In this study, models with a MAPE less than 5% were considered to have good performance, and models with a MAPE less

Table 2
Performance of the pre-COVID forecasts.

Agency	Measure	ETS	ARIMA	STL-ETS	STL-ARIMA	TBATS	NNET	Hybrid (ANST)	Minimum	Maximum
Chicago Transit Authority	MAPE	1.79	1.76	1.64	1.49	2.65	2.40	1.33	Hybrid (ANST)	TBATS
	MASE	0.53	0.50	0.47	0.42	0.78	0.67	0.38		
County of Miami-Dade	MAPE	7.99	5.52	7.25	6.78	7.67	8.03	7.64	ARIMA	NNET
	MASE	1.22	0.84	1.12	1.04	1.17	1.22	1.17		
Los Angeles County Metropolitan Transportation Authority	MAPE	6.53	3.53	5.52	5.25	6.10	9.46	5.76	ARIMA	NNET
	MASE	0.94	0.50	0.79	0.75	0.87	1.35	0.82		
Maryland Transit Administration	MAPE	7.42	10.46	7.60	9.69	7.41	24.76	11.94	TBATS	NNET
	MASE	0.49	0.70	0.49	0.63	0.50	1.64	0.80		
Massachusetts Bay Transportation Authority	MAPE	2.92	5.07	3.39	3.39	3.09	4.48	2.96	ETS	ARIMA
	MASE	0.49	0.85	0.57	0.57	0.52	0.74	0.49		
Metropolitan Atlanta Rapid Transit Authority	MAPE	12.75	16.10	12.99	13.25	12.99	10.98	11.81	NNET	ARIMA
	MASE	1.71	2.16	1.74	1.78	1.75	1.45	1.58		
MTA New York City Transit	MAPE	7.71	7.87	7.44	8.09	8.05	10.25	8.65	STL-ETS	NNET
	MASE	1.99	2.03	1.92	2.08	2.08	2.64	2.23		
Port Authority Trans-Hudson Corporation	MAPE	2.82	3.36	2.03	1.96	2.27	3.01	1.90	Hybrid (ANST)	ARIMA
	MASE	0.49	0.57	0.35	0.33	0.40	0.52	0.33		
Port Authority Transit Corporation	MAPE	2.45	3.06	2.14	2.63	3.05	7.44	3.47	STL-ETS	NNET
	MASE	0.69	0.86	0.60	0.74	0.86	2.08	0.98		
San Francisco Bay Area Rapid Transit District	MAPE	2.81	2.40	2.55	1.78	2.11	3.11	2.05	STL-ARIMA	NNET
	MASE	0.65	0.53	0.59	0.41	0.48	0.68	0.46		
Southeastern Pennsylvania Transportation Authority	MAPE	3.35	3.19	2.87	2.87	6.09	5.93	4.29	STL-ETS	TBATS
	MASE	0.60	0.57	0.52	0.52	1.12	1.07	0.78		
Staten Island Rapid Transit Operating Authority	MAPE	4.88	7.07	5.23	4.60	4.66	5.60	5.16	STL-ARIMA	ARIMA
	MASE	0.63	0.89	0.67	0.59	0.60	0.72	0.66		
The Greater Cleveland Regional Transit Authority	MAPE	13.30	13.51	13.20	13.12	12.64	22.65	13.77	TBATS	NNET
	MASE	0.73	0.74	0.73	0.73	0.69	1.28	0.75		
Washington Metropolitan Area Transit Authority	MAPE	10.00	8.62	7.33	7.72	7.93	6.19	7.29	NNET	ETS
	MASE	2.04	1.73	1.49	1.57	1.62	1.25	1.48		
Green - MAPE below 5%, Yellow - MAPE below 10%, Red - MASE above 1.00										

than 10% were considered to have acceptable performance (Hyndman and Athanasopoulos, 2023a). In the table, a MAPE below 5% (good performance) is highlighted green, a MAPE below 10% (acceptable performance) is highlighted yellow, and a MASE over 1.00 (did not outperform the naïve method) is highlighted red.

For 81.6% (80 out of 98) of the models, the MAPEs were acceptable (below 10%), and for 42.9% (42 out of 98) of the models the MAPEs were good (below 5%). These results show that, using R software to automatically select model parameters, the seven time series forecasting methods were capable of producing acceptable pre-COVID transit ridership forecasts at most of the selected agencies.

Despite generally acceptable performance, at four of the agencies, no model produced a MAPE below 5% (MTA New York City Transit, Washington Metropolitan Area Transit Authority, Maryland Transit Administration, and the County of Miami-Dade). At those four agencies, the MASE values were generally over 1.00, indicating that the models did not outperform the naïve method. At the Metropolitan Atlanta Rapid Transit Agency and Greater Cleveland Regional Transit Authority, none of the models produced a MAPE below 10%. None of the Metropolitan Atlanta Rapid Transit Agency models outperformed the naïve method, while almost all the Greater Cleveland Regional Transit Authority models did. The difficulty in forecasting at these agencies appeared to be related to changes in ridership trends in 2019; in New York City, ridership rose much higher than in prior years, whereas in the rest of the cities, ridership decreased more than in prior years. In the case of Cleveland, ridership was over estimated due to closures on their one and only heavy rail line over the summer of 2019.

Among the seven methods, no one method stood out as having the best performance, as they all performed well for most agencies. The neural network underperformed compared to the other methods. In general, the pre-COVID results suggest that these seven methods of time series forecasting perform well when the ridership data have stable seasonal patterns and trends. However, for agencies without those features, ridership forecasting with time series methods may be more challenging.

5.2. Results of the full series analysis

For the full series analysis, the training dataset was from January 2002 through December 2022, and the testing dataset was from January 2023 through December 2023. Similar to the pre-COVID results, visualizations of the forecasts for the transit agencies in Chicago and Cleveland are provided in Fig. 7 and Fig. 8 to show examples of the outputs that were created for each of the fourteen agencies. Although the data was trained from 2002 to 2022, only the training data after 2020 is shown so that the post-COVID forecasts can be shown clearly. Once again, most of the forecasts, like those for the Chicago Transit Authority, appear to have captured fairly well the patterns in the data and have forecasted approximately the correct scale of ridership. However, for a small number of agencies, such as the Greater Cleveland Regional Transit Authority and the Maryland Transit Administration, the forecasts appeared to over- or under-estimate ridership and failed to capture much of a pattern at all.

Table 3 presents the results of the full series model performance for each agency. The lowest/highest errors were again used to measure the performance of each method. For 52% (47 out of 91) of the models, the MAPEs were below 10%, and for 14% (13 out of 91) of the models the MAPEs were below 5%. It should be noted that the data from the Southeastern Pennsylvania Transportation Authority had a quality control issue that impacted the performance of the forecasts; therefore, the results for the Southeastern Pennsylvania Transportation Authority were omitted from the discussion.

Most of the methods produced acceptable MAPEs at around half the agencies using the full series. The neural network and STL-ARIMA methods underperformed for this set of forecasts. There was once again a division of model performance between transit agencies; at three of the smaller agencies, the Maryland Transit Authority, Greater Cleveland Regional Transit Authority, and Staten Island Rapid Transit Operating Authority, none of the models produced acceptable forecasts; with the exception of the models for the Greater Cleveland Regional Transit Authority, almost none of the forecasts for these three agencies outperformed the naïve method. For the agencies in Los Angeles, San

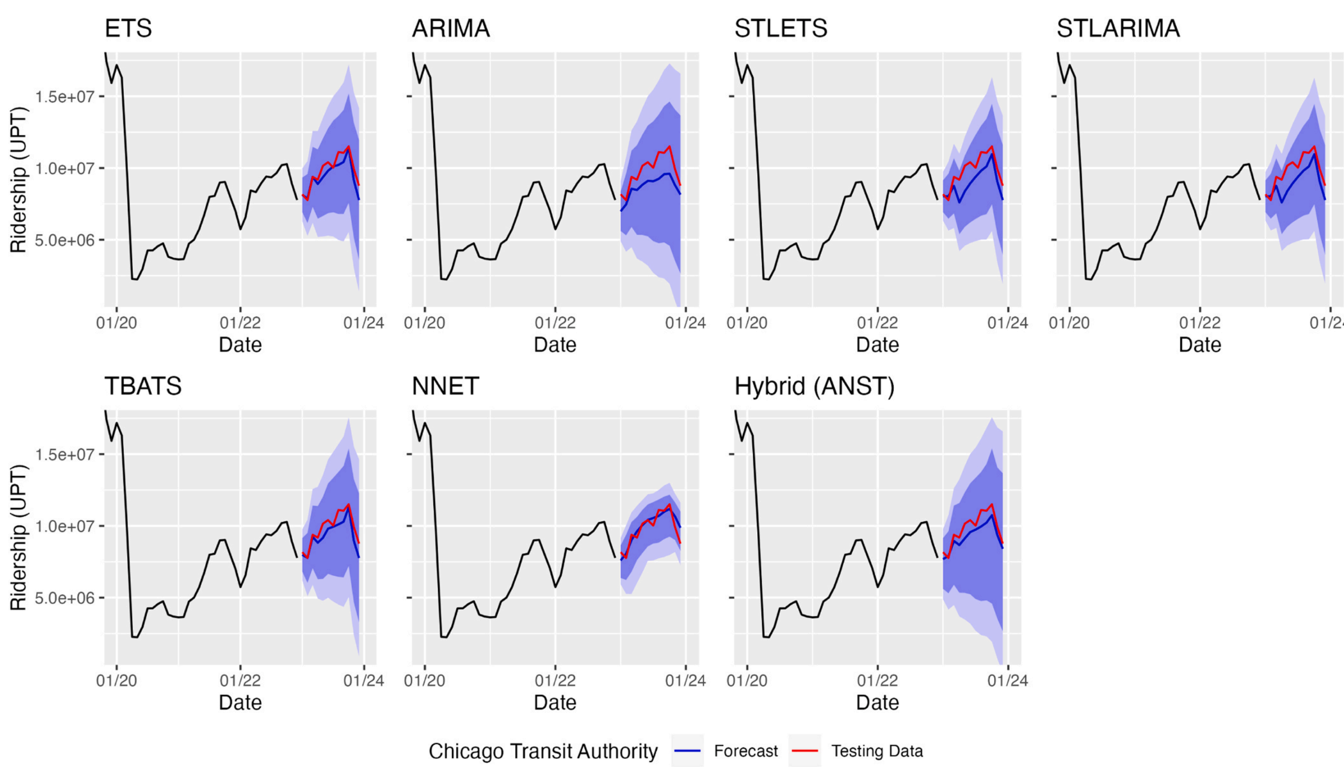

Fig. 7. Full series ridership forecasts for Chicago Transit Authority.

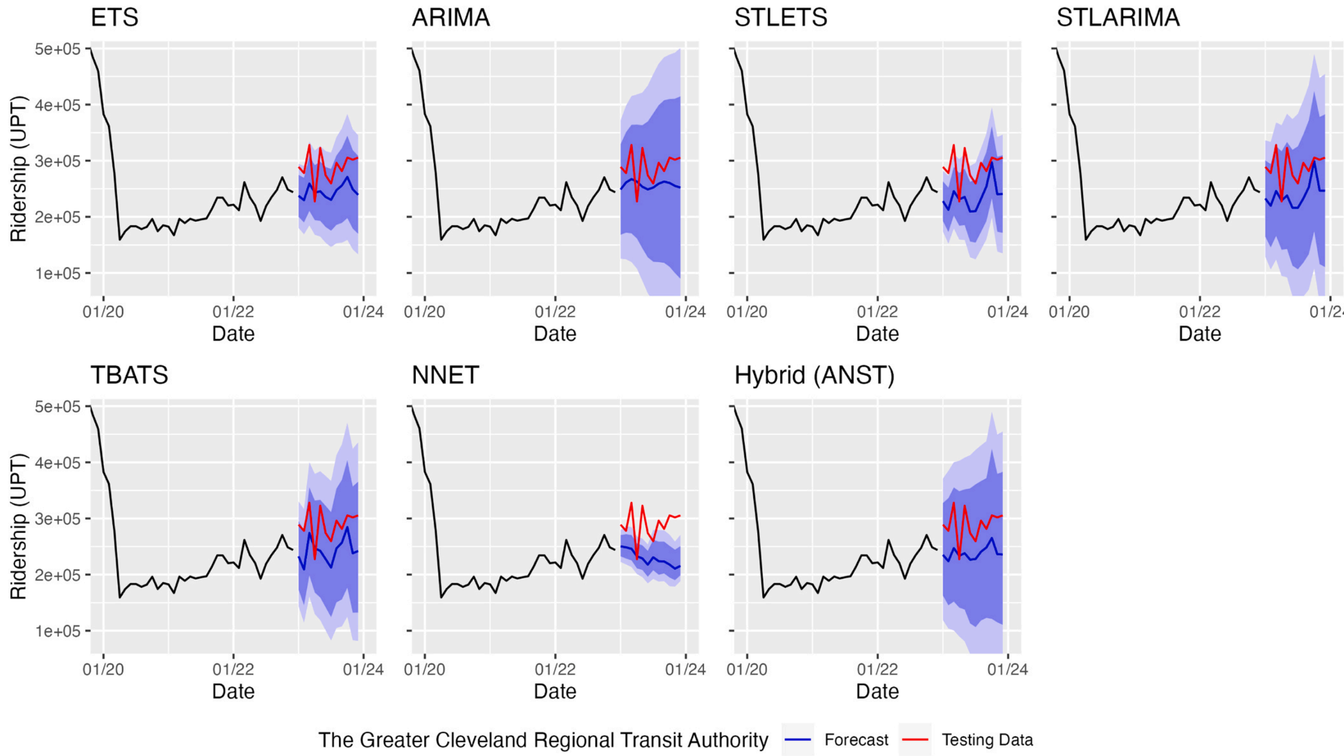

Fig. 8. Full series ridership forecasts for the Greater Cleveland Regional Transit Authority.

Table 3
Performance of the full series forecasts.

Agency	Measure	ETS	ARIMA	STL-ETS	STL-ARIMA	TBATS	NNET	Hybrid (ANST)	Minimum	Maximum
Chicago Transit Authority	MAPE	4.49	11.25	9.37	9.37	5.25	4.43	6.09	NNET	ARIMA
	MASE	0.30	0.75	0.62	0.62	0.35	0.27	0.41		
County of Miami-Dade	MAPE	5.14	6.89	9.38	9.48	5.59	5.94	5.09	Hybrid (ANST)	STL-ARIMA
	MASE	0.40	0.51	0.73	0.74	0.44	0.44	0.39		
Los Angeles County Metropolitan Transportation Authority	MAPE	9.58	13.23	14.27	14.16	10.42	13.35	12.18	ETS	STL-ETS
	MASE	0.60	0.81	0.94	0.92	0.66	0.76	0.74		
Maryland Transit Administration	MAPE	41.39	49.25	38.31	50.46	34.28	40.33	38.19	TBATS	STL-ARIMA
	MASE	1.03	1.36	1.01	1.30	0.90	1.08	1.09		
Massachusetts Bay Transportation Authority	MAPE	5.01	5.20	4.79	4.79	4.41	8.05	4.18	Hybrid (ANST)	NNET
	MASE	0.27	0.27	0.26	0.26	0.23	0.43	0.22		
Metropolitan Atlanta Rapid Transit Authority	MAPE	5.92	4.25	5.81	10.27	5.47	31.12	11.53	ARIMA	NNET
	MASE	0.28	0.20	0.28	0.49	0.26	1.49	0.55		
MTA New York City Transit	MAPE	5.25	6.97	6.94	6.94	9.31	3.66	4.18	NNET	TBATS
	MASE	0.44	0.55	0.58	0.58	0.78	0.30	0.35		
Port Authority Trans-Hudson Corporation	MAPE	8.49	5.71	11.67	12.80	9.64	4.79	6.65	NNET	STL-ARIMA
	MASE	0.50	0.34	0.69	0.75	0.57	0.28	0.40		
Port Authority Transit Corporation	MAPE	3.57	4.98	7.83	7.86	3.26	17.20	5.83	TBATS	NNET
	MASE	0.21	0.30	0.46	0.47	0.19	1.02	0.35		
San Francisco Bay Area Rapid Transit District	MAPE	9.46	15.50	12.44	14.11	9.14	50.03	18.23	TBATS	NNET
	MASE	0.43	0.72	0.57	0.65	0.41	2.31	0.84		
Southeastern Pennsylvania Transportation Authority [1]	MAPE	14.20	26.98	11.55	11.23	18.46	16.98	18.72	STL-ARIMA	ARIMA
	MASE	0.75	1.43	0.62	0.60	1.03	0.89	1.00		
Staten Island Rapid Transit Operating Authority	MAPE	24.03	22.82	28.41	29.32	24.76	32.76	27.56	ARIMA	NNET
	MASE	1.35	1.32	1.60	1.66	1.40	1.86	1.57		
The Greater Cleveland Regional Transit Authority	MAPE	15.69	12.85	18.07	16.80	16.94	20.44	17.45	ARIMA	NNET
	MASE	0.59	0.48	0.67	0.63	0.63	0.77	0.66		
Washington Metropolitan Area Transit Authority	MAPE	8.36	11.11	15.92	16.85	12.20	15.92	7.12	Hybrid (ANST)	STL-ARIMA
	MASE	0.49	0.66	0.95	1.01	0.72	0.91	0.42		

Green - MAPE below 5%, Yellow - MAPE below 10%, Red - MASE above 1.00, Gray – Quality Control Issue
[1] Southeastern Pennsylvania Transportation Authority's ridership data appear to have a quality control issue that has negatively impacted the performance of the models

Francisco, and Washington D.C., only the ETS method was consistently able to produce an acceptable forecast.

The results for the pre-COVID period showed that none of the MTA New York City Transit models produced a MAPE below 5%. For the full series forecasts, the neural network and hybrid models have now produced good forecasts, and the rest of the forecasts for MTA were acceptable. Similarly, for the Metropolitan Atlanta Rapid Transit Authority, none of the pre-COVID forecasts in the previous section were acceptable or outperformed the naïve method, but half of the forecasts were acceptable for the full-series forecasts in this section. For these two agencies, most of the full series forecasts that trained through the onset of the pandemic actually outperformed the last set of forecasts that trained off of only pre-COVID data. A possible explanation for the improved performance using the full series compared to the pre-COVID series could be that in New York, there was an unexpected period of higher ridership before the pandemic; the opposite situation is true of Atlanta.

The residual tests showed that most of the full series models captured all patterns that were present in the data, leaving only white noise. However, for many of the pre-COVID models, the residuals still contained some patterns. These findings may reflect a continued interruption in the strong seasonal and periodic transit ridership trends that were present at most agencies pre-COVID, which has made it more difficult to extract patterns from the post-COVID data. As a result, the full series forecast results are mixed, reflecting much more uncertainty; for some agencies, the methods were able to predict the general periodic trend and produce good or acceptable forecasts, but for other agencies, the methods were not able to capture much of a pattern at all, and the forecasts underperformed. Nevertheless, the methods were able to produce acceptable forecasts at about half of the agencies, and for some agencies (e.g., MTA New York City Transit and the Metropolitan Atlanta Rapid Transit Authority) the models produced better forecasts than were produced using only the pre-COVID period. In the next section, the post-COVID observations are isolated in order to better understand the new trends and patterns that may exist in post-COVID ridership.

5.3. Results of the post-COVID period

For the post-COVID period, the training dataset was from April 2020 through December 2022, and the testing dataset was from January 2023 through December 2023. A visualization of the forecasts for Chicago Transit Authority and the Greater Cleveland Regional Transit Authority are provided in Fig. 9 and Fig. 10, as examples of the outputs that were created for each of the 14 heavy rail agencies. Despite having used only the post-COVID data, the forecast results are overall somewhat similar to the forecasts produced using the full data series. The major differences between the full series plots and the post-COVID plots are that the post-COVID forecasts tend to be smoother and have more of an upward trend compared to the full series forecasts. However, the ETS models struggled to identify much trend or pattern at all, reflecting the ETS method's relative computational simplicity and lack of robustness. The prediction intervals for the two STL methods, the TBATS method, and the neural network also tended to be much smaller for the post-COVID forecasts, which could indicate that the models may have struggled with overfitting the data (Halloway, N.D.).

Table 4 presents the results of post-COVID model performance for each agency. The lowest/highest errors were again used to measure the performance of each method. For 55% (50 out of 91) of the models, the MAPEs were acceptable (below 10%), and for 13% (12 out of 91) of the models the MAPEs were good (below 5%). Once again, the results for Southeastern Pennsylvania Transportation Authority were omitted from the discussion due to a quality control issue in their data.

The overall performance of the models is similar to that of the full series; 47 of the full series models and 50 of the post-COVID models had at least an acceptable MAPE. Also, 13 of the full series models and 12 of the post-COVID models had a good MAPE. Despite potentially overfitting the data, the TBATS and neural network methods performed rather well relative to the other methods, having produced the most acceptable and good MAPEs, respectively. The STL-ETS and hybrid methods also performed relatively well, each producing good/acceptable MAPEs at eight of the agencies. The STL-ARIMA method underperformed, having produced the worst MAPE at just under half the

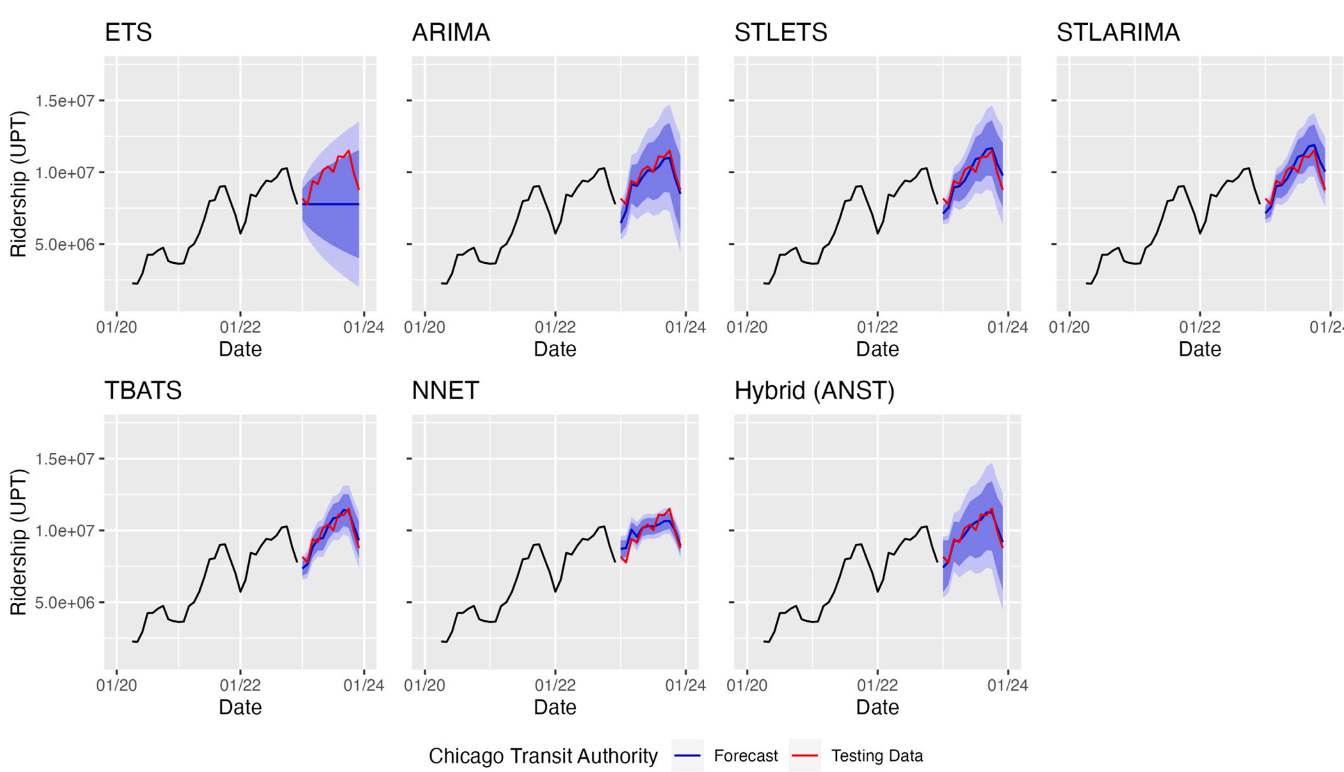

Fig. 9. Post-COVID ridership forecasts for Chicago Transit Authority.

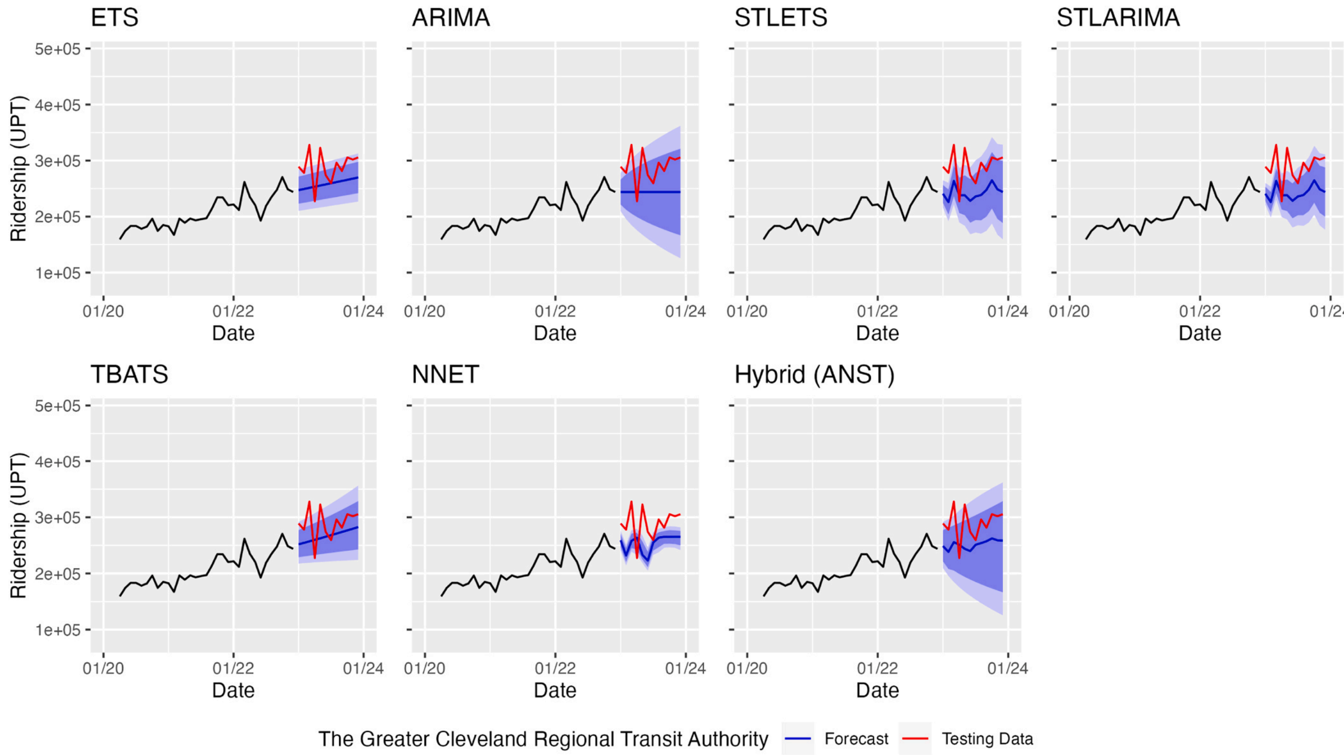

Fig. 10. Post-COVID ridership forecasts for the Greater Cleveland Regional Transit Authority.

Table 4
Performance of post-COVID ridership forecasting models.

Agency	Measure	ETS	ARIMA	STL-ETS	STL-ARIMA	TBATS	NNET	Hybrid (ANST)	Minimum	Maximum
Chicago Transit Authority	MAPE	19.53	4.76	5.49	5.98	4.31	4.58	3.11	Hybrid (ANST)	ETS
	MASE	0.72	0.16	0.18	0.20	0.15	0.16	0.11		
County of Miami-Dade	MAPE	6.68	11.68	5.63	9.49	8.29	7.19	5.81	STL-ETS	ARIMA
	MASE	0.37	0.62	0.31	0.50	0.47	0.41	0.31		
Los Angeles County Metropolitan Transportation Authority	MAPE	10.79	10.99	14.11	14.11	14.13	10.62	12.47	NNET	TBATS
	MASE	0.59	0.61	0.79	0.79	0.77	0.58	0.69		
Maryland Transit Administration	MAPE	42.00	42.00	45.53	45.53	43.15	33.73	36.75	NNET	STL-ARIMA
	MASE	2.50	2.50	2.54	2.54	2.56	1.64	2.19		
Massachusetts Bay Transportation Authority	MAPE	14.32	8.50	14.52	14.86	8.58	6.18	7.31	NNET	STL-ARIMA
	MASE	0.43	0.26	0.45	0.46	0.26	0.19	0.22		
Metropolitan Atlanta Rapid Transit Authority	MAPE	6.23	4.99	4.92	7.37	4.57	5.85	4.62	TBATS	STL-ARIMA
	MASE	0.28	0.23	0.22	0.34	0.21	0.27	0.21		
MTA New York City Transit	MAPE	16.98	7.83	6.40	12.59	6.22	3.73	7.67	NNET	ETS
	MASE	0.58	0.27	0.22	0.44	0.22	0.13	0.26		
Port Authority Trans-Hudson Corporation	MAPE	8.78	5.11	6.78	8.87	7.80	3.63	4.77	NNET	STL-ARIMA
	MASE	0.29	0.16	0.22	0.29	0.26	0.12	0.15		
Port Authority Transit Corporation	MAPE	12.07	6.04	7.38	12.80	9.17	3.10	6.97	NNET	STL-ARIMA
	MASE	0.43	0.21	0.26	0.46	0.33	0.11	0.25		
San Francisco Bay Area Rapid Transit District	MAPE	7.68	13.51	8.36	7.53	8.11	26.08	10.99	STL-ARIMA	NNET
	MASE	0.24	0.41	0.25	0.22	0.24	0.79	0.33		
Southeastern Pennsylvania Transportation Authority [1]	MAPE	36.27	24.79	34.88	32.36	25.43	14.42	23.98	NNET	ETS
	MASE	1.01	0.67	0.96	0.89	0.70	0.38	0.65		
Staten Island Rapid Transit Operating Authority	MAPE	12.75	10.53	20.00	16.29	10.53	21.02	12.66	TBATS	NNET
	MASE	0.57	0.45	0.85	0.69	0.46	0.90	0.56		
The Greater Cleveland Regional Transit Authority	MAPE	11.72	16.08	16.20	16.20	9.82	13.96	13.94	TBATS	STL-ARIMA
	MASE	1.13	1.56	1.56	1.56	0.94	1.33	1.34		
Washington Metropolitan Area Transit Authority	MAPE	14.36	8.37	8.02	7.77	8.66	12.02	7.59	Hybrid (ANST)	ETS
	MASE	0.54	0.28	0.28	0.27	0.30	0.44	0.26		

Green - MAPE below 5%, Yellow - MAPE below 10%, Red - MASE above 1.00, Gray – Quality Control Issue
[1] Southeastern Pennsylvania Transportation Authority's ridership data appear to have a quality control issue which has negatively impacted the performance of the models

agencies. The ETS method also underperformed, having produced an acceptable MAPE at only four agencies.

Generally, all models outperformed the naïve method, with the exception of two of the three smallest agencies, the Greater Cleveland Regional Transit Authority and the Maryland Transit Administration. As discussed in the previous section, ridership at these two agencies may be more difficult to forecast due to the limited heavy rail service provided in Cleveland and Baltimore, which each have only one heavy rail line. Different methods of forecasting may be necessary to overcome the challenges associated with identifying patterns at these agencies.

5.4. Results summary

In summary, pre-COVID, forecasting transit ridership with time series methods was straightforward, and each method performed fairly well, with the models collectively producing acceptable results for about 80% of the forecasts. Forecasting in the post-COVID era was more challenging, with the models that trained on either the full series or post-COVID datasets producing acceptable results for about half the forecasts. Generally, the classical methods outperformed the other methods when trained on the full data series, but the more complex methods outperformed the other methods when trained only on the post-COVID data. For the models with training data starting from January 2002 (the pre-COVID and full series periods), the neural network method notably underperformed; however, the neural network outperformed the other methods when it trained only off of post-COVID data. The performance of the neural network may be due to the use of only one seasonal lag as opposed to 12; in the pre-COVID and full series periods, only using one seasonal lag could have given the neural network a disadvantage compared to the other models. However, the same parameter appeared to give the neural network an advantage in the post-COVID era when the previous year's ridership would likely be a worse predictor of future ridership compared to the previous month's ridership. The hybrid method tended to perform relatively well for all time periods.

The performance of the methods also varied by agency; for some agencies like the Chicago Transit Authority and the Port Authority Trans-Hudson Corporation, forecasting with time series methods was relatively straightforward regardless of the period used to train the models. For a few agencies, like those in New York City and Washington D.C., the methods actually performed better when forecasting post-COVID ridership compared to forecasting pre-COVID ridership. At a couple of agencies, like those in Cleveland and Baltimore, time series forecasting was challenging for all time periods and using all methods; those agencies generally had very limited heavy rail service, e.g., only a single heavy rail line.

These results demonstrate the need for agency- and context-specific solutions to forecast ridership. In order to use time series forecasting methods, an agency needs to understand their data in order to properly set up a model; for example, weekly data require different parameters than monthly data because the seasonal period has a different number of observations (monthly=12). Post-COVID, the prior year's ridership became a worse predictor of ridership compared to the previous month's ridership, making it possible to produce acceptable forecasts while essentially ignoring the seasonal period. However, as ridership stabilizes, this will likely no longer be the case for new observations, which may be expected to once again return to a strong seasonal, yearly pattern. At agencies with unstable ridership levels or a lack of strong seasonal ridership patterns, simpler forecasting methods are more appropriate, such as simple ETS methods. More advanced time series methods may be appropriate at these agencies if the models are adjusted to account for other variables besides ridership, such as line closures for repairs, fare changes, service changes, or special events. At agencies with more stable ridership levels and seasonal patterns, time series methods that combine statistical and machine learning methods are more appropriate, especially when there exists a sufficiently large number of data points. For data with a very large number of observations, more complex methods like the TBATS method may be more appropriate, especially if the data contain both long- and short-term seasonal patterns, e.g., daily or weekly data that span many years.

The findings of this research are consistent with the findings of other recent studies. Prior studies found substantial decreases in forecasting performance in situations with dynamic conditions relative to stable conditions, such as before and after the onset of the COVID-19 pandemic and a major national protest (Azimian and Jiao, 2021, Caicedo et al., 2023). The performance measures of the forecasts in prior studies, when comparing pre- and post-COVID conditions, were found to decrease in performance by anywhere from about 200% to about 600%, depending on the method used to forecast. In one prior study, at least around 45 new observations were required for the models to learn a new demand pattern (Caicedo et al., 2023); in this study, the maximum number of observations used to train the models on the new demand pattern was 33, which could partially explain the full series and post-COVID models' overall poor performance relative to the pre-COVID condition. Other prior studies have found that the performance of a time series method varied by transit agency, with ridership being more challenging to predict at agencies that have more limited or low-quality service (Gao et al., 2023). Prior studies have also found better performance of classical time series methods relative to machine learning methods given more uncertain conditions (Azimian and Jiao, 2021).

6. Conclusions and future research

This study compared the performance of seven time series forecasting methods for pre-COVID and post-COVID transit ridership for the 14 heavy rail providers in the continental United States. Three forecasting time periods were examined, which consisted of pre-COVID (prior to March 2020), full series (January 2002 to December 2023) and post-COVID data (after March 2020). Several key findings have emerged from this research.

First, time series forecasting prior to the COVID-19 pandemic could be done in a relatively straightforward manner for most agencies. Ridership data at most heavy rail agencies in the US contained steady, seasonal trends that could be modeled fairly well using classical time series forecasting methods as well as newer, more complex methods like the TBATS, neural network, and hybrid methods. From the findings, it can be concluded that time series forecasting is a simple, quick, and efficient method to forecast ridership at agencies with strong and regular seasonality, periodicity, and/or trends, which was typical prior to the COVID-19 pandemic.

A second noteworthy conclusion is that forecasting monthly ridership post-COVID is still challenging, likely due to competing trends, changing travel behavior, limited datasets, and other factors not reflected in the time series data. Nevertheless, time series forecasting methods such as those used in this study were capable of producing acceptable post-COVID ridership forecasts for about half of the models using both the full series of data as well as only post-COVID data. Relative to the other models, the classical methods tended to perform well for the longer data series that contained pre-COVID ridership (the pre-COVID and full series periods), and the more complex methods tended to perform well for the post-COVID period, which used only post-COVID data to train the models. The hybrid method, which was modeled using equal weights of two of the classical time series methods (ARIMA and STL-ARIMA) and two of the more complex methods (TBATS and neural network), tended to perform relatively well for all time periods. The neural network underperformed for the pre-COVID and full series periods, but for the post-COVID-only period, it outperformed the other methods. One reason for this could be that the neural network was set to only have one seasonal lag, instead of using 12 seasonal lags for monthly data. This may have given the neural network an advantage at some agencies in the post-COVID period due to a continued lack of clear seasonal patterns, which may be caused by competing trends in the data due to factors like ever-recovering ridership, especially now that COVID-related restrictions and trends of social isolation have relaxed. Particularly at agencies with more limited heavy rail service (i.e., smaller heavy rail agencies like the Greater Cleveland Regional Transit Authority), there seemed to be a lack of patterns in the heavy rail ridership data, as demonstrated by the white noise test and the performance of the forecasts, which were generally poor.

From these findings, it can be concluded that univariate time series forecasting is a simple, quick, and efficient method to forecast ridership at agencies with strong and regular seasonality, periodicity, and/or trends, which was typical of transit ridership prior to the COVID-19 pandemic. However, based on the findings of this study, the strong, stable ridership patterns found at many US heavy rail agencies prior to the pandemic have still not returned, which could be due to external factors such as telework or people relocating. Interestingly, at some agencies in major US cities (e.g., New York City and Washington D.C.), monthly transit ridership has become more predictable or has remained about as predictable post-COVID as compared to before the pandemic. Additionally, in some cities with more limited heavy rail service (for example, Cleveland and Baltimore), ridership remains just as unpredictable as it was prior to the pandemic. These mixed findings demonstrate a need for agency-specific solutions to ridership forecasting (i.e., picking the most suitable forecasting method based on the features of the dataset).

Several areas for future research have emerged from this study. First, more post-COVID data will become available in the future, allowing for further exploration of the post-pandemic trends. Future research should compare forecasts between cities to understand the differences in trends at some agencies, especially at smaller agencies where the time series methods struggled to produce good forecasts. Additionally, other methods may be used in future research to identify external factors that have an impact on heavy rail ridership and to quantify their impact; for example, a recently published study quantified the impact of external factors on post-COVID bus ridership levels using a multiple mediation analysis and found that unemployment, telework, and people relocating explained 13–38% of the ridership declines (Ziedan et al., 2023). Future studies could consider manually adjusting the model parameters or including exogenous regressors in the models to account for other factors impacting ridership and possibly improve the performance of the forecasts. Last, future studies may use larger datasets, e.g., daily or weekly data with a length of at least 1,000 datapoints, which would make suitable the use of more complex time series forecasting methods such as deep learning methods.

Based on the findings of this study, time series forecasting is likely to be a promising method of predicting system-level transit ridership at many heavy rail agencies, even in the post-COVID era. By making the code used in this study publicly available, as well as bringing more visibility to Minneapolis Metro's forecasting app and the code that powers it, transit agencies may find these methods of transit ridership forecasting more accessible, helping them to produce accurate and low-cost ridership forecasts more frequently and with a wider variety of methods. This is critical to practitioners, planners, and policy makers as they attempt to forecast system-wide ridership levels in the post-COVID era to set budget levels, plan schedules, and perform other important real world tasks using transit ridership forecasts.

Author contribution statement

The authors confirm contribution to the paper as follows: study conception and design: A. Hightower, A. Ziedan, C. Brakewood; data collection: A. Hightower, A. Ziedan; analysis and interpretation of results: A. Hightower, A. Ziedan, X. Zhu, and C. Brakewood; draft manuscript preparation: A. Hightower, A. Ziedan, J. Guo, and C. Brakewood. All authors reviewed the results and approved the final version of the manuscript.

Funding

This work was supported by the Dwight D. Eisenhower Transportation Fellowship Program Award No. 693JJ32445014, as well as the Transit-Serving Communities Optimally, Responsively, and Efficiently (TSCORE) University Transportation Center Grant No. 69A3552047141 with matching funds from the Tennessee Department of Transportation (TDOT RES2021-15) and the University of Tennessee, Knoxville.

CRediT authorship contribution statement

Candace Brakewood: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Xiaojuan Zhu:** Writing – review & editing, Writing – original draft, Methodology. **Jing Guo:** Writing – original draft, Visualization, Software, Methodology, Investigation. **Abubakr Ziedan:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Ashley Hightower:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

None.

References

Ayman, A., J. Martinez, P. Pugliese, A. Dubey, and A. Laszka, 2022. Neural Architecture and Feature Search for Predicting the Ridership of Public Transportation Routes. Presented at 2022 IEEE International Conference on Smart Computing (SMARTCOMP), Helsinki, Finland. Accessed December 15, 2023. 10.1109/SMARTCOMP55677.2022.00023.

Azimian, A., Jiao, J., 2021. Prediction of "L" Train's daily ridership in downtown Chicago during the COVID-19 Pandemic. Findings. https://doi.org/10.32866/001c.30181. Accessed July 15, 2022.

Babcock, M.W., Lu, X., Norton, J., 1999. Time series forecasting of quarterly railroad grain carloadings. Transp. Res. Part E: Logist. Transp. Rev. Vol. 35 (No. 1), 43–57. Accessed July 15, 2022.

Boyle, D., 2006. TCRP Synthesis 66: Fixed-Route Transit Ridership Forecasting and Service Planning Methods. Transit Cooperative Research Program. https://doi.org/10.17226/14001. Accessed December 14, 2023.

Caicedo, J.D., M.C. Gonzàlez, and J.L. Walker, 2023. Public Transit Demand Prediction During Highly Dynamic Conditions: A Meta-Analysis of State-of-the-ART Models and Open-Source Benchmarking Infrastructure. Accessed December 15, 2023.https://doi.org/10.48550/arXiv.2306.06194.

Chen, C.-H.P., Naylor, G.A., 2011. Development of a mode choice model for bus rapid transit in Santa Clara County, California. J. Public Transp. Vol. 14, 41–62. https://doi.org/10.5038/2375-0901.14.3.3. Accessed October 23, 2022.

Chen, D., Zhang, J., Jiang, S., 2020. Forecasting the short-term metro ridership with seasonal and trend decomposition using loess and LSTM Neural Networks. IEEE Access Vol. 8. https://doi.org/10.1109/ACCESS.2020.2995044. Accessed June 30, 2022.

Ding, C., Duan, J., Zhang, Y., Yu, G., 2018. Using an ARIMA-GARCH modeling approach to improve subway short-term ridership forecasting accounting for dynamic volatility. IEEE Trans. Intell. Transp. Syst. Vol. 19 (No. 4), 1054–1064.. https://doi.org/10.1109/TITS.2017.2711046. Accessed June 30, 2022.

Doi, M., Allen, W.B., 1986. A time series analysis of monthly ridership for an urban rail rapid transit line. Transportation Vol. 13, 257–269. https://doi.org/10.1007/BF00148619. Accessed July 1, 2022.

Egu, O., Bonnel, P., 2021. Medium-term public transit route ridership forecasting: what, how and why? A Case Study in Lyon. Transp. Policy Vol. 105, 124–133. https://doi.org/10.1016/j.tranpol.2021.03.002. Accessed June 30, 2022.

Federal Transit Administration. Monthly Module Adjusted Data Release. Federal Transit Administration. https://www.transit.dot.gov/ntd/data-product/monthly-module-adjusted-data-release. Accessed November 15, 2023.

Gao, Y., Li, J., Wang, J., Yang, Z., 2023. U.S. public transportation ridership analysis and prediction based on COVID-19. Int. Soc. Opt. Eng. Vol. 12596 https://doi.org/10.1117/12.2672649. Accessed November 15, 2023.

Ge, S.Y., Zheng, C.J., Hou, M.M., 2013. Forecast of bus passenger traffic based on exponential smoothing and trend moving average method. Appl. Mech. Mater. Vol. 433-435, 1374–1378. https://doi.org/10.4028/www.scientific.net/AMM.433-435.1374. Accessed July 1, 2022.

Halloway, B.S. Understanding Prediction Intervals. Bryan S. Halloway. https://www.bryanshalloway.com/2021/03/18/intuition-on-uncertainty-of-predictions-in

troduction-to-prediction-intervals/#a-few-things-to-know-about-prediction-intervals. Accessed December 14, 2023.
Hoque, J.M., Zhang, I., Schmitt, D., Erhardt, G.D., 2023. Are public transit investmentsbased on accurate forecasts? An analysis of the improving trend of transit ridership forecasts in the United States. Available at SSRN: https://ssrn.com/abstract=4557555. Accessed December 14, 2023. doi:https://doi.org/10.2139/ssrn.4557555.
Hyndman, R.J., Athanasopoulos, G., 2018a. A taxonomy of exponential smoothing methods. In Forecasting: Principles and Practice, second edition. OTexts, Melbourne, Australia. Accessed December 15, 2023.
Hyndman, R.J., Athanasopoulos, G., 2018b. 6.8 Forecasting with decomposition. In Forecasting: Principles and Practice, second edition. OTexts, Melbourne, Australia. Accessed December 15, 2023.
Hyndman, R.J., Athanasopoulos, G., 2018c. 11.3 Neural network models. In Forecasting: Principles and Practice, second edition. OTexts, Melbourne, Australia. Accessed December 15, 2023.
Hyndman, R.J., Athanasopoulos, G., 2018d. Forecasting: Principles and Practice, 2nd Edition. OTexts, Melbourne, Australia. Accessed December 15, 2023. https://otexts.com/fpp2
Hyndman, R., Athanasopoulos, G., 2018e. 8.8 Forecasting. In Forecasting: principles and practice, 2nd edition. OTexts, Melbourne, Australia. Accessed December 15, 2023.
Hyndman, R., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. J. Stat. Softw. Vol. 26 (3), 1–22 https://www.jstatsoft.org/article/view/v027i03. Accessed December 15, 2023.
Hyndman, R.J, 2017. Prediction intervals for NNETAR models. Rob J. Hyndman. https://robjhyndman.com/hyndsight/nnetar-prediction-intervals/. Accessed December 15, 2023.
Hyndman, R.J., G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeen, 2021. forecast: Forecasting functions for time series and linear models. R package version 8.15. Accessed July 18, 2022.
Jayanthi, G., Jothilakshmi, P., 2021. Traffic time series forecasting on highways - a contemporary survey of models, methods and techniques. Int. Jounal Logist. Syst. Manag. Vol. 39 (No. 1), 77–110. https://doi.org/10.1504/IJLSM.2021.115068. Accessed July 15, 2022.
Kyte, M., Stoner, J., Cryer, J., 1988. A time-series analysis of public transit ridership in Portland, Oregon, 1971-1982. Transp. Res. Part A Gen. Vol. 22 (No. 5), 345–359. https://doi.org/10.1016/0191-2607(88)90012-X. Accessed July 1, 2022.
Lind, E., J. Reid, and K.E. Ky, N.D. MetroTransit / Route-Trends, Build Complete Base Function App. 〈https://github.com/metrotransit/route-trends/commit/752d7b50fd250d66e334c8dc31487890f8d1b8ab〉. Accessed December 14, 2023.
Livera, A.M.D., Hyndman, R.J., Snyder, R.D., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. J. Am. Stat. Assoc. 106 (No. 496), 1513–1527. Accessed November 11, 2023.
Minneapolis Metro Transit. Route Trends and Info, Minneapolis. https://metrotransitmn.shinyapps.io/route-trends/. Accessed July 13, 2022.
Moghimi, B., Kamga, C., Safikhani, A., Mudigonda, S., Vicuna, P., 2022. Non-stationary time series model for station-based subway ridership during COVID-19 Pandemic: case study of New York City. Transp. Res. Rec. Vol. 2677 (No. 4) https://doi.org/10.1177/03611981221084698. Accessed June 21, 2022.
Moreland, A., C. Herlihy, M. A. Tynsn, G. Sunshine, R. F. McCord, C. Hilton, J. Poovey, A. K. Werner, C. D. Jones, E. B Fulmer, A. V. Gundlapalli, H. Strosnider, A. Potvien, M. C. Garcia, S. Honeycutt, G. Baldwin, CDC Public Health Law Program, and CDC COVID-19 Response Team Mitigation Policy Analysis Unit, 2020. Timing of state and territorial COVID-19 stay-at-home orders and changes in population movement - United States, March 1-May 31, 2020. CDC. https://www.cdc.gov/mmwr/volumes/69/wr/mm6935a2.htm. Accessed March 22, 2024.
Mucci, A., Erhardt, G.D., 2018. Evaluating the ability of transit direct ridership models to forecast medium-term ridership changes: evidence from San Francisco. Transp. Res. Rec. Vol. 2672 (No. 46) https://doi.org/10.1177/0361198118758632. Accessed October 23, 2022.
Nihan, N.L., Holmesland, K.O., 1980. Use of the Box and Jenkins time series technique in traffic forecasting. Transportation Vol. 9, 125–143. https://doi.org/10.1007/BF00167127. Accessed July 15, 2022.
Perry, J., 2017. Measuring the accuracy of bus rapid transit forecasts. J. Public Transp. Vol. 20, 119–138. https://doi.org/10.5038/2375-0901.20.1.7. Accessed October 23, 2022.
Plotch, P, 2022. Transit Ridership: Not Expected to Return to Pre-Pandemic Levels This Decade. Eno Center for Transportation. https://www.enotrans.org/article/transit-ridership-not-expected-to-return-to-pre-pandemic-levels-this-decade/. Accessed July 10, 2022.
R Core TeamR: A language and environment for statistical computing R Foundation for Statistical Computing N.D. Vienna Austria.Accessed July 25, 2022http://www.r-project.org.
Razbash, S., R.J. Hyndman, N.D. tbats: TBATS model (Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components). RDocumentation.org. https://www.rdocumentation.org/packages/forecast/versions/8.21.1/topics/tbats. Accessed November 11, 2023.
Shaub, D, N.D. Using the “forecastHybrid” package. cran.r-project.org. https://cran.r-project.org/web/packages/forecastHybrid/vignettes/forecastHybrid.html#forecasting. Accessed November 11, 2023.
Vanderkam, D., Allaire, J., Owen, J., Gromer, D., Thieurmel, B., 2018. dygraphs: Interface to “Dygraphs” Interactive Time Series Charting Library. R. Package Version 1. 1. 1. 6. Accessed February 2, 2022.
Wang, X., N. Zhang, Y. Chen, and Y. Zhang, 2018. Short-Term Forecasting of Urban Rail Transit Ridership Based on ARIMA and Wavelet Decomposition. Presented at 6th International Conference on Computer-Aided Design, Manufacturing, Modeling and Simulation, Busan, South Korea. AIP Conference Proceedings 1967. doi:10.1063/1.5039099. Accessed June 22, 2022.
Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., 2020. Statistical and Econometric Methods for Transportation Data Analysis. CRC Press, New York, NY.
Wickham, H, N.D. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, New York, 2016. Accessed July 25, 2022.
Wickham, R.R, 1995. Evaluation of Forecasting Techniques for Short-Term Demand of Air Transportation. Boston, MA. Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/11584. Accessed July 15, 2022.
Zheng, X., Wei, P., 2020. Air transportation direct share time series forecasting: a hybrid model. J. Adv. Transp. Vol. 3, 1–13. https://doi.org/10.1155/2020/8924095. Accessed July 15, 2022.
Ziedan, A., Lima, L., Brakewood, C., 2023. A multiple mediation analysis to untangle the impacts of COVID-19 on nationwide bus ridership in the United States. Transp. Res. Part A Policy Pract. Vol. 173 https://doi.org/10.1016/j.tra.2023.103718. Accessed December 15, 2023.