Strategies for Heavy Rail Ridership Forecasting using Statistical, Machine Learning, and Ensemble Time Series Methods

Ashley Hightower
Department of Civil and Environmental Engineering
University of Tennessee, Knoxville
851 Neyland Drive, Knoxville TN 37996
Email: ahighto3@utk.edu

Candace Brakewood, PhD
Department of Civil and Environmental Engineering
University of Tennessee, Knoxville
851 Neyland Drive, Knoxville TN 37996
Email: cbrakewo@utk.edu

Word Count: 6,728 words + 3 tables (250 words per table) = 7,478 words

Submitted: July 17, 2024

ABSTRACT
Transit agencies could benefit from forecasting methods that are cheaper, quicker, and easier than the most commonly used methods, and time series methods are one potential alternative not widely used in practice. In this study, time series analysis was used to produce univariate, system-wide, monthly transit ridership forecasts for all heavy rail agencies in the continental United States. Using an automated moving window approach, nearly 3,000 models were generated to examine the changing forecasting performance over time of seven different methods: ARIMA, ETS, STL-ARIMA, STL-ETS, TBATS, a neural network, and an ensemble method. Three time periods were used to generate forecasts, which were the full series (2002 to 2023), pre-COVID period (pre March 2020), and post-COVID period (post March 2020). The MAPE and MASE were used to measure forecast accuracy. Using each method, the majority of the pre-COVID forecasts had acceptable performance. By comparison, the models underperformed using the full series and post-COVID periods. The ensemble and ARIMA methods tended to outperform the other methods for each time period. The neural network substantially underperformed for the pre-COVID and full series periods, but slightly outperformed the other methods for the post-COVID period. Based on these findings, time series forecasting is an efficient method to forecast ridership with stable seasonality, periodicity, and trends, which was typical of pre-pandemic ridership. However, the results suggest that for data exhibiting less stable ridership patterns, univariate time series forecasting of transit ridership is more challenging.

Keywords: public transit ridership, time series forecasting, heavy rail, COVID-19

INTRODUCTION

Even prior to the COVID-19 pandemic, overall transit ridership in the US was in decline (see **Figure 1**). The COVID-19 pandemic caused substantial ridership decreases due to reduced service, stay-at-home orders, and fears of infection, among other reasons. More than three years after the onset of the COVID-19 pandemic, ridership at nearly all US agencies in medium/large metro areas has still not recovered *(1)*, likely due to factors such as the increased popularity of telework, changes in travel patterns *(2)*, and service disruptions *(3)*. At the same time, other types of travel in the US such as air travel and interstate travel have recovered *(4; 5)*. Moreover, the COVID-19 pandemic has introduced more uncertainty to transit ridership forecasting by disturbing previously well-established features of ridership patterns, such as the presence of strong seasonality and daily peaks typical of commuter ridership *(2)*.

Transit agencies can benefit from considering alternative methods of ridership forecasting. The most widely-used methods of ridership forecasting in practice tend to be the four-step and activity-based travel demand models, which are time consuming to set up, require extensive calibration, and are not suited to sub-regional forecasting *(6-8)*. While more simple methods of ridership forecasting exist and are capable of achieving a finer resolution, for example, linear regression models *(6)*, time series models are particularly easy to set up due to their ability to produce univariate forecasts; in other words, transit agencies only need to know their previous ridership in order to begin forecasting. Given that most agencies already report their ridership levels to the Federal Transit Administration, they likely already have the data they need to produce time series forecasts. By using free, open-source software packages to automatically optimize the parameters of a time series model, agencies could quickly, frequently, and cheaply produce reasonably accurate forecasts for regular planning and budgeting activities.

In this study, monthly ridership for the 14 heavy rail agencies in the continental United States (US) was forecasted using seven different time series methods: ARIMA, ETS, STL-ARIMA, STL-ETS, TBATS, a neural network, and an ensemble method. Three time series periods were used. In the first time period, a baseline was established for pre-COVID forecast performance. In the second time period, the models were trained through the onset of the pandemic to examine the usefulness of pre-COVID data in post-COVID forecasts. In the third time period, the models were trained only on post-COVID data in order to test the ability of the models to identify trends and patterns in post-COVID ridership. A moving window was used to see the changing performance of the forecasts over time, which can provide insight into the factors that influence the accuracy of univariate ridership forecasts. By comparing forecasts for all US heavy rail agencies, any universal and reoccurring characteristics of the data that negatively impacted forecasting performance could be exposed. Additionally, by testing multiple forecasting periods, the effect of the length of the training data and a disruptive outside factor (COVID) was evaluated. This study can help inform transit agencies who seek to explore more dynamic and less computationally intensive methods to accurately forecast ridership.

This paper proceeds as follows. First, a literature review is presented that briefly explains traditional methods of forecasting transit ridership and recent studies that have forecasted ridership using time series methods. The forecasting software, periods of analysis, time series methods, and performance measures are then explained. Next, the results of the forecasts are presented; about 3,000 time series models were generated, which demonstrate their changing performance over time and due to the pandemic. Conclusions and areas for future research are discussed last.

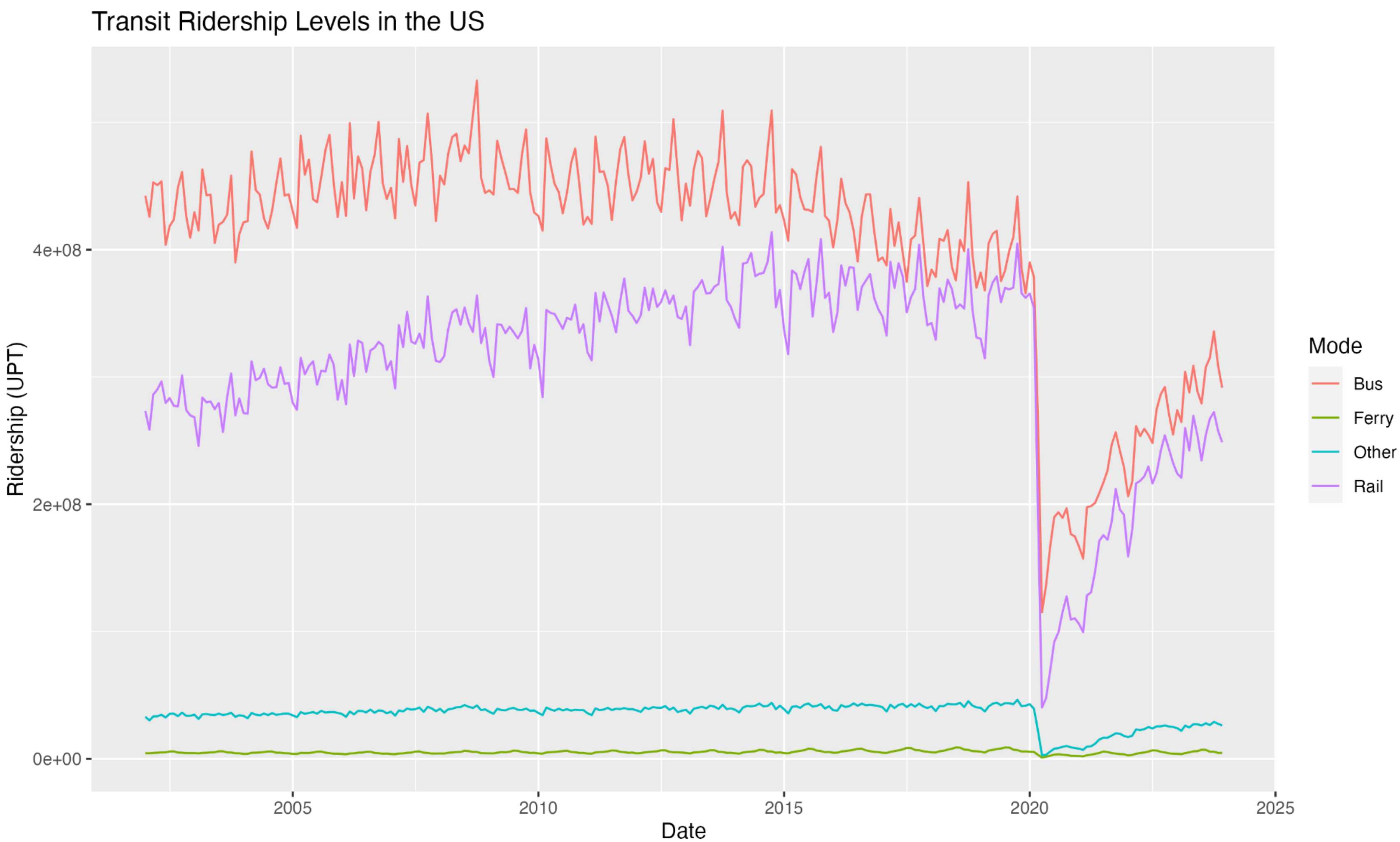

Figure 1 National Transit Ridership in the US by Mode (January 2002 to December 2023)

LITERATURE REVIEW

The four-step and activity-based travel demand models are common methods of estimating ridership demand in practice *(9-18)*, as are direct ridership models using regression-based techniques *(6)*. However, the four-step and activity-based models have some drawbacks; for example, they have extensive calibration requirements which can be time-consuming to complete *(7; 8)*, are best suited for the regional scale *(6)*, and their assumptions may not hold for post-COVID conditions *(7)*. Direct ridership models provide high-resolution ridership forecasts, but they require multiple variables, may require fine-tuning regarding which variables to include, and are often best-suited to produce order-of-magnitude estimates of ridership *(6)*.

Many prior studies of transit ridership forecasting have explored the performance and applicability of a variety of other methods, for example, time series multiple regression *(19)*, ordinary least squares regression *(20)*, and mode choice models *(21)*. One recent study compiled data from 164 transit projects in the US in order to evaluate the overall accuracy of transit forecasting. Heavy rail ridership forecasts were found to be highly optimistic, perhaps due to the scope of their service and the potential for high variability in demand *(22)*.

Although there exist many good forecasting methods, time series methods are especially user friendly, cost effective, and time efficient. In particular, univariate time series models can be used to quickly generate ridership forecasts at any level of resolution and are easily adaptable to new conditions, unlike other models which depend on assumptions or base parameters informed by experimental evidence (e.g., ridership elasticities).

There exist several relevant studies of time series methods to forecast transit ridership in dynamic conditions *(23-29)*. In the two most relevant studies, the effect of a disruptive event (the COVID-19 pandemic) on forecasting performance was evaluated using a number of statistical and machine learning methods *(30; 31)*. In the first of the two most relevant prior studies (which is also by the authors of this paper), the same seven time series methods used in this study were used to forecast pre- and post-COVID heavy rail ridership to evaluate the general suitability of time series forecasting for transit demand. Ridership forecasting was found to be straightforward pre-COVID, but the performance of the forecasts were negatively impacted by the onset of the pandemic at most agencies. Using a longer dataset for post-COVID forecasts, the classical and ensemble methods outperformed the other models, but using a short dataset, the more complex TBATS, neural network, and ensemble methods outperformed *(30)*. Similarly, another relevant study compared real-time forecasting performance for bus ridership in Bogota, Columbia using five different ARIMA and deep learning models. The models once again were found to all perform well in stable conditions but performed substantially worse in dynamic conditions. The LSTM model with adaptive training was found to adapt the quickest to dynamic conditions *(31)*.

This brief review shows that transit ridership forecasting is an important and challenging task, especially when conditions are unstable. The most common forecasting methods in practice tend to be demanding, time-consuming, and are often limited in their scope and by the context of the study area. Time series models require only one variable, previous ridership, in order to be estimated, but can also incorporate covariates associated with transit ridership, e.g., gas prices, change in population, fare changes, and income levels. This study builds upon prior research by investigating the long-term effects of COVID-19 on the predictability of heavy rail ridership through a moving window technique. By estimating nearly 3,000 forecasts for 14 agencies, this study sheds light on what universal factors, if any, impact the performance of ridership forecasts and helps to inform transit planners about strategies to forecast ridership in dynamic conditions.

DATA AND METHODS

This section discusses the data, period of analysis, software, models and performance measures used to conduct this analysis.

Data

Monthly ridership data in terms of unlinked passenger trips (UPT) were downloaded from the National Transit Database *(32)*. The fourteen agencies used in this study are shown in **Table 1**, alongside background information regarding each agency's service area population and size, annual UPT, and percent recovery in ridership from December 2019 to December 2023.

Period of Analysis

Three time periods were used for analysis. The first time period was the pre-COVID period, which included all NTD heavy rail ridership data prior to March 2020. This time period was isolated in order to establish a baseline performance of the time series methods under stable conditions. The second time period was the full data series from January 2002 to December 2023, which was the latest available data point at the time of analysis. The third time period was the post-COVID period, which included all data after March 2020. These periods of analysis are consistent with a prior study on time series ridership forecasting written by the authors of this paper *(30)*.

Software

Staff at Minneapolis Metro Transit published an online tool to produce univariate time series forecasts of monthly transit ridership *(33)*. Their tool is an open source, R-based Shiny app *(34; 35)*. In order to conduct a time series analysis using the same methods a transit agency may use, the same forecasting methods used by Minneapolis Metro Transit were selected for this study. The functions used in this study were written by the authors using R Version 4.3.1, and the code is publicly available at https://github.com/ashley2876/forecasting_repo.

Time Series Models

Shown in Figure 2, seven methods used in this study are: ARIMA, ETS, STL-ARIMA, STL-ETS, TBATS, a feed-forward autoregressive neural network with a single hidden layer (NNET), and an ensemble model with equal weights of the ARIMA, NNET, STL-ARIMA, and TBATS methods (Hybrid ANST). These seven methods were chosen because they are already used to forecast ridership by at least one US transit agency, Minneapolis Metro Transit. Other time series methods were not included because of their computational complexity, obscurity, and/or because they were not suitable for the data's length or pattern. A brief overview of the advantages and disadvantages of each method is provided in **Table 2**. For more details, readers are referred to *(30; 36; 37)*.

Performance Measures

The performance measures used in this paper were the mean absolute percent error (MAPE) and the mean absolute scaled error (MASE) of the testing data. These performance measures were selected because they are suitable to compare forecasts of datasets that have different scales. Based off popular convention and practice, forecasts with a MAPE less than 5% were considered to have "good" performance, and forecasts with a MAPE less than 10% were considered to have "acceptable" performance. A MASE less than one indicated that the forecast performed better than the seasonal naïve forecast, i.e., compared to simply setting all forecasted values to be equal to the last observed value from the same season, the forecast produced a more accurate result *(36)*.

Table 1 Summary Statistics of Heavy Rail Agencies in the Continental US

Agency Name	**UZA Name**	**Service Area Square Miles**	**Service Area Population**	**Unlinked Passenger Trips FY 2022**	**Recovery in Ridership Dec '19 to Dec '23**
Chicago Transit Authority	Chicago, IL--IN	310	3,207,635	103,524,858	55.1%
County of Miami-Dade	Miami--Fort Lauderdale, FL	306	2,701,767	11,446,854	82.1%
Los Angeles County MTA	Los Angeles--Long Beach--Anaheim, CA	4,099	10,394,849	25,075,130	50.7%
Maryland Transit Administration	Baltimore, MD	2,560	7,811,145	2,252,070	50.6%
Massachusetts Bay Transportation Authority	Boston, MA--NH	3,244	3,109,308	78,861,897	55.4%
Metropolitan Atlanta Rapid Transit Authority	Atlanta, GA	949	2,128,687	26,079,792	54.4%
MTA New York City Transit	New York--Jersey City--Newark, NY--NJ	321	8,804,190	1,788,363,060	73.3%
Port Authority Trans-Hudson Corporation	New York--Jersey City--Newark, NY--NJ	226	3,134,256	46,589,043	64.4%
Port Authority Transit Corporation	Philadelphia, PA--NJ--DE--MD	16	159,726	4,870,310	50.4%
San Francisco Bay Area Rapid Transit District	San Francisco--Oakland, CA	80	867,725	36,774,619	42.1%
Southeastern Pennsylvania Transportation Authority	Philadelphia, PA--NJ--DE--MD	844	3,475,337	52,499,263	52.0%
Staten Island Rapid Transit Operating Authority	New York--Jersey City--Newark, NY--NJ	59	495,747	3,757,728	68.2%
The Greater Cleveland Regional Transit Authority	Cleveland, OH	458	1,412,140	2,808,149	66.4%
Washington Metropolitan Area Transit Authority	Washington--Arlington, DC--VA--MD	1,349	5,089,918	76,077,714	60.8%
Source: National Transit Database Complete Monthly Ridership, December 2023 *(32)*					

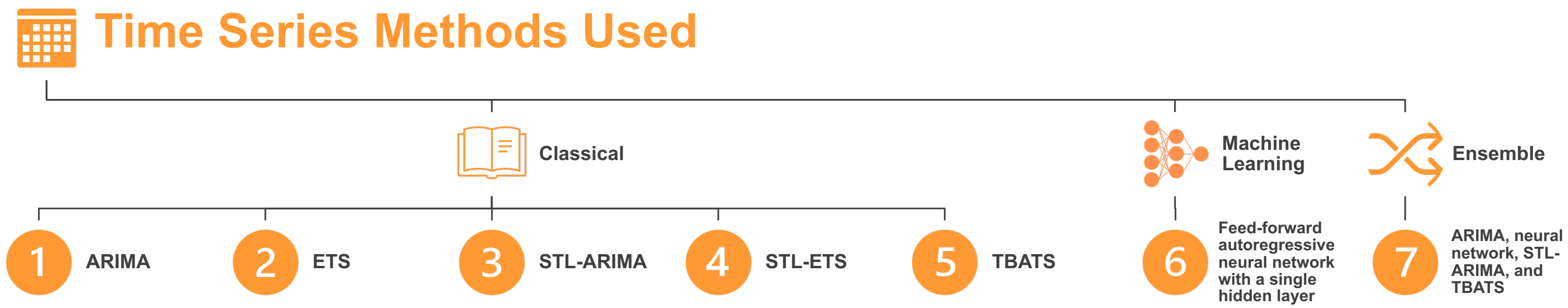

Figure 2 Seven Forecasting Methods

Table 2 Advantages and Disadvantages of Each Forecasting Method

Method	Advantages	Disadvantages	Reference
ARIMA	• Very robust, potentially infinite models • Relatively computationally simple • Very well-established and well-defined in literature and practice	• Assumes linear data	*(36; 38-41)*
ETS	• Adjustable parameter can place higher/lower weights on more recent observations • Suitable for data with no clear trend or pattern • Computationally simple	• Not robust, effectively has only nine models to choose from	*(36; 42)*
STL	• Allows the seasonal component to change over time • Robust to outliers • Can handle nonlinearity	• Does not automatically handle calendar variation	*(36; 43)*
TBATS	• Good for complex seasonality • Allows seasonality to change slowly over time • Can handle nonlinearity	• Long computation time	*(36; 37)*
Neural Network	• Good for complex nonlinear relationships	• Computationally complex • Longer computation time • Tendency to overfit	*(44-46)*
Ensemble	• Minimizes the errors from each method • Has been shown in literature to outperform pure statistical and pure machine learning methods	• Computationally complex • Longest computation time	*(46; 47)*

RESULTS AND DISCUSSION

In order to determine if there exist consistent and reoccurring characteristics of ridership data that tend to negatively impact univariate ridership forecasts, moving windows were used to estimate a consecutive series of forecasts with 12-month-long horizons. The MAPE and MASE were calculated for each forecast in order to record the changing performance over time of the seven time series methods used in this study. A total of 2,940 forecasts were estimated. The results of the pre-COVID forecasts are shown first in order to establish a baseline of performance for the time series forecasting methods. Next, the results of the full series forecasts are shown, followed by the post-COVID forecasts. Last, a summary of the overall performance of each method for each time period is given.

Results of the Pre-COVID Analysis

For the pre-COVID period, 12 sets of forecasts were estimated in order to establish a baseline performance considering the effect of a full year of seasonal changes in transit ridership. The forecasting windows were selected by setting the last window to use the last 12 months of observations prior to the COVID-19 pandemic; therefore, the last set of models used testing data from March 2019 to February 2020, and the first set of models used testing data from April 2018 to March 2019 (see
Figure 3).

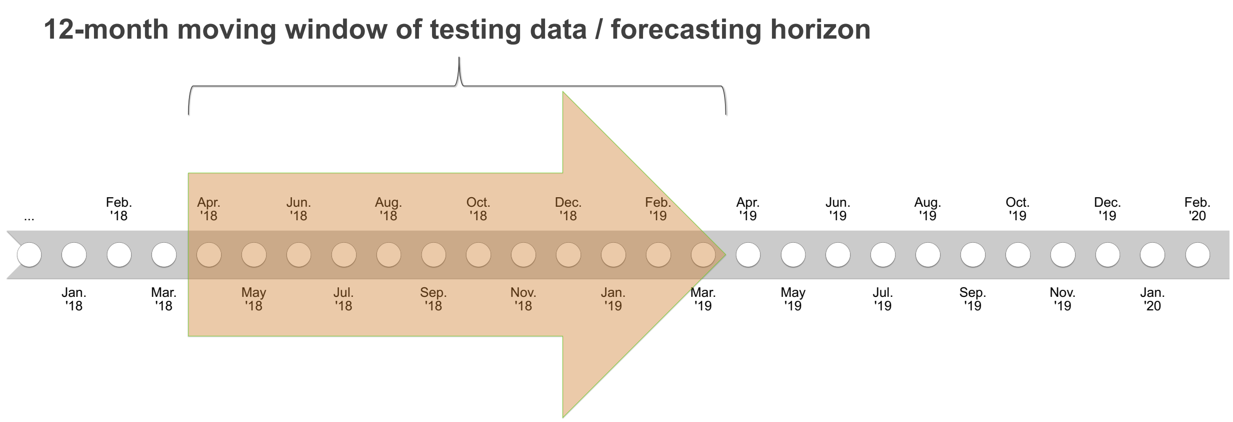

Figure 3 Moving Window for a Consecutive Series of 12-Month, Pre-COVID Forecasts

Each forecast estimated ridership for the following 12 months, and the estimates were compared to actual ridership in order to evaluate the forecasting performance using the MAPE and MASE (also referred to as “errors”). In this study, a forecast with a MAPE below 5% was considered to have “good” performance, and a forecast with a MAPE below 10% was considered to have “acceptable” performance. A MASE less than 1.00 means the forecast outperformed the simple naïve method. In **Figure 4**, the errors for each forecasting method at the Chicago Transit Authority were plotted alongside the time series of ridership. In **Figure 5**, the errors for the Greater Cleveland Regional Transit Authority were similarly plotted; these two agencies were selected as examples of the results found at each of the 14 agencies in this study. The change in errors is shown on the y-axis of the upper plots, and the x-axis represents which testing/forecasting window was used, using the first month of the testing data as the identifier. The bottom plot shows the time series of monthly ridership in unlinked passenger trips; although the models were trained on data beginning in January 2002, only the time series from January 2010 to the onset of the pandemic were included, so that the pre-COVID trend and seasonal patterns are clearly visible.

Figure 4 shows that for pre-COVID ridership at the Chicago Transit Authority, all but one model produced good forecasts. The ETS and ARIMA methods tended to outperform the other methods. The neural network consistently underperformed compared to the rest of the methods and struggled to produce forecasts that outperformed the naïve method. All models struggled to outperform the naïve method for the forecasting windows starting in November and December. Examining the ridership plot revealed a larger-than-previous fluctuation in ridership from late 2018 to early 2019, which likely contributed to the decrease in forecasting performance for those windows. However, thanks to Chicago's obvious strong seasonal patterns, the time series methods generally performed well.

Figure 5 shows that for pre-COVID ridership at the Greater Cleveland Regional Transit Authority, the time series methods tended to produce acceptable results until the testing window moved to begin in September 2018. The models' performance appeared to improve over the course of the first three testing windows, with the ETS and TBATS methods even producing good forecasts with a MAPE less than 5%. However, beginning with the fourth testing window, the performance steadily decreased. Examining the ridership plot revealed discontinuities in relative seasonality in the data, especially in 2019. Also in 2019, ridership decreases were steeper than in previous years, likely due to a closure on the heavy rail line for maintenance over the summer of 2019 *(48)*. Overall, there appeared to be fewer and weaker seasonal patterns in Cleveland compared to Chicago, which may have made forecasting more challenging in that region.

Figures for the other heavy rail agencies were not included for the sake of brevity. At many agencies, there was a decrease in forecasting performance that emerged with the testing window beginning in mid to late 2018. Often, the forecasts then improved, creating a concave shape on the plots. For the agencies whose forecasts improved, the changes in MAPE could be an arbitrary byproduct of measuring performance using an averaged percent error, i.e., the MAPE could be sensitive to unexpectedly high/low seasonal peaks/lulls in ridership, especially when the ridership has been overestimated, producing negative percent errors *(36)*. But for agencies with steadily worsening forecast performance, in many cases, the decline in forecast performance could be related to the trend in the data in 2019, i.e., larger changes in ridership compared to previous years. Nevertheless, the majority of the models produced good or acceptable forecasts for all agencies and all testing windows. Generally, about 90% of the models for each forecasting method had an acceptable MAPE, and about 60% of the models had a good MAPE; the one exception was the neural network, for which method only 43% of the models produced a good MAPE. Despite the overall acceptable results, there were some agencies for which ridership was consistently more difficult to forecast (Cleveland and Baltimore). This may be due to the limited heavy rail service in those two cities, which each have only one heavy rail line. No one method stood out as having particularly better performance compared to all the rest for pre-COVID heavy rail ridership, although at individual agencies, some methods did outperform the others.

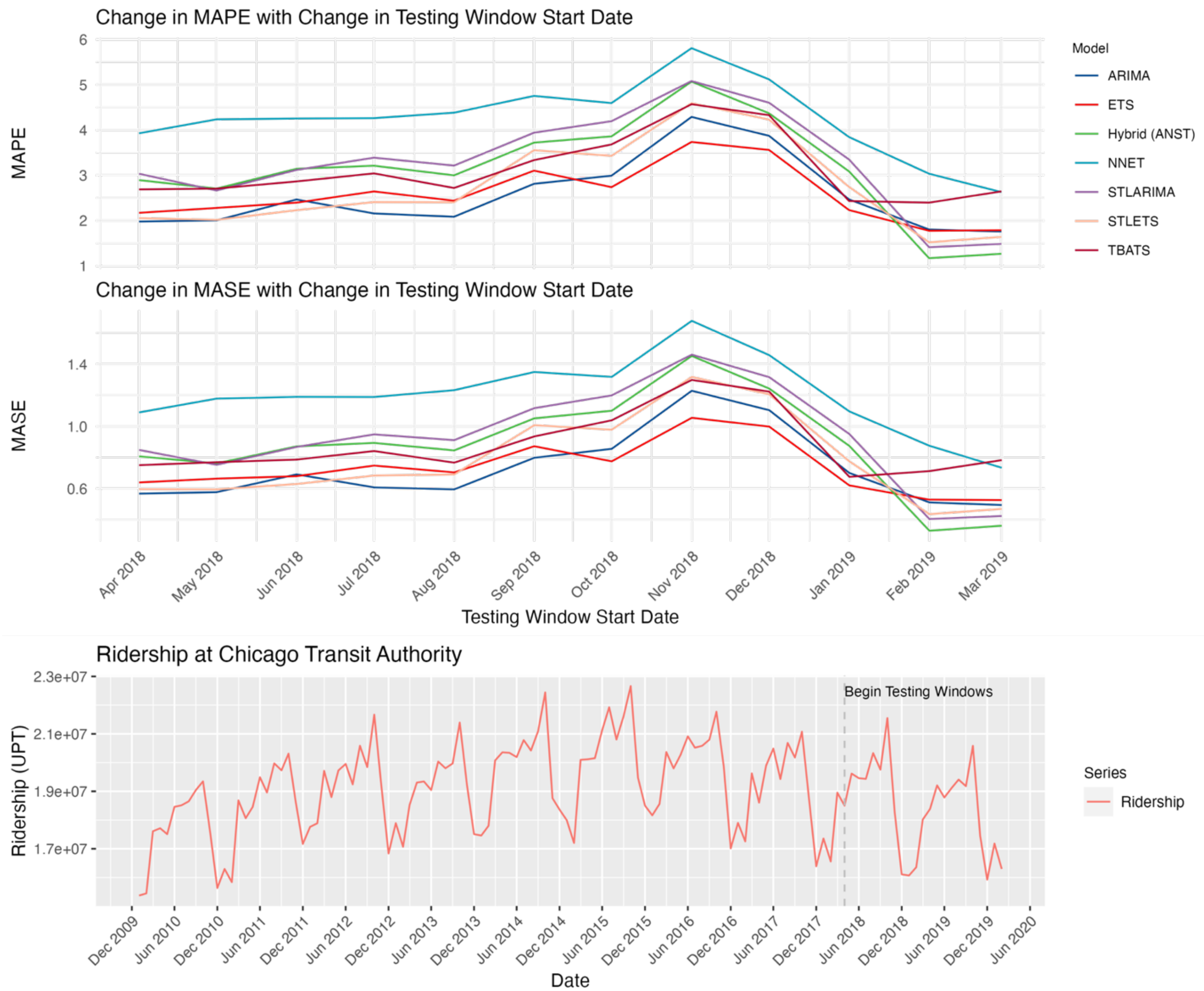

Figure 4 Pre-COVID Change in Errors at the Chicago Transit Authority

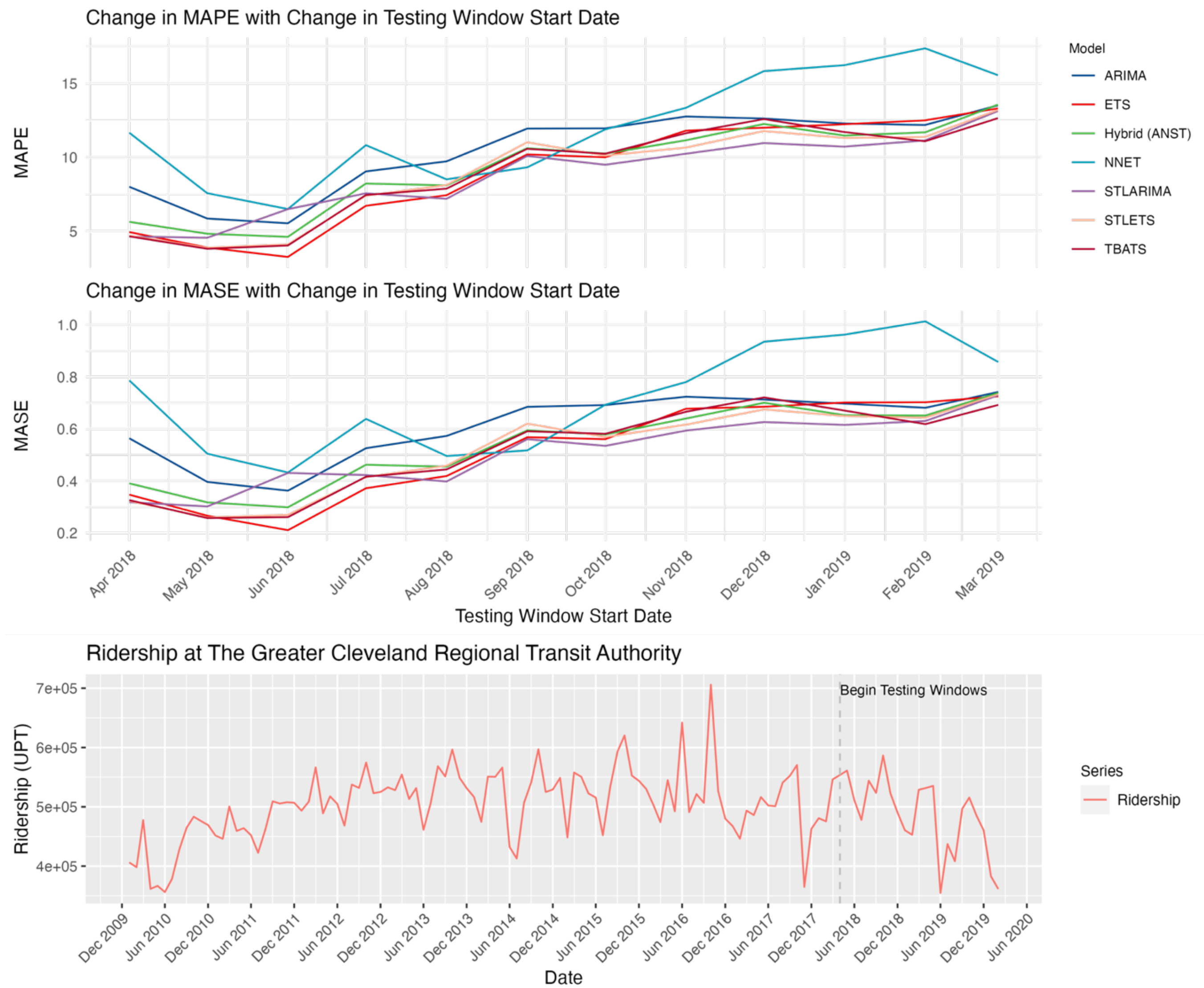

Figure 5 Pre-COVID Change in Errors at the Greater Cleveland Regional Transit Authority

Results of the Full Series Analysis

Nine sets of forecasts were estimated for post-COVID ridership using the full data series (January 2002 to December 2023). The STL methods require at least 25 observations in order to produce forecasts. Therefore, the post-COVID series' forecasts needed to train on the data at least from April 2020 to April 2022. In order to compare between the forecasts from the full series and post-COVID series, both periods' first testing/forecasting window began with May 2022, allowing for nine sets of forecasts to be estimated (see **Figure 6**).

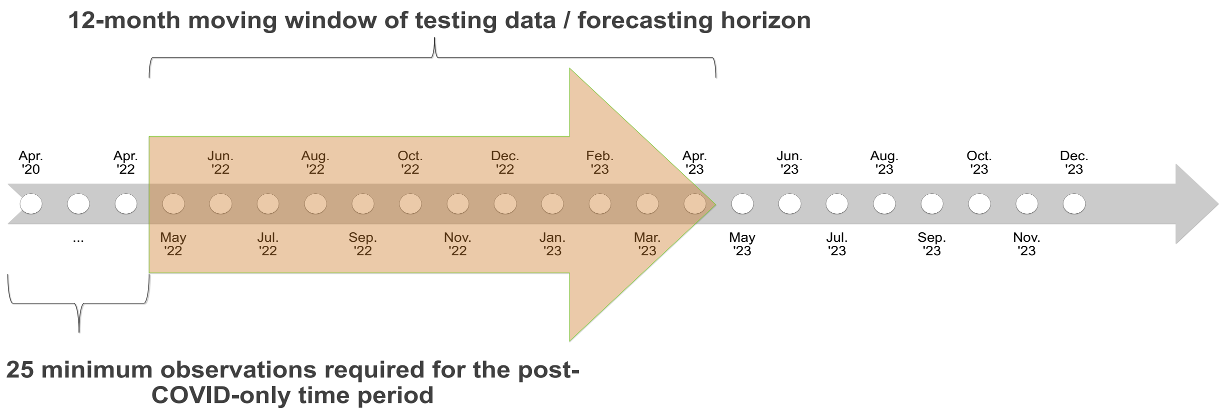

Figure 6 Moving Window for a Consecutive Series of 12-Month Post-COVID Forecasts Using the Full Series and Post-COVID Time Periods

Each of the 882 models produced for the full series period were used to forecast ridership for the following 12 months, and the errors were recorded. The errors for the Chicago Transit Authority and the Greater Cleveland Regional Transit Authority are shown in **Figure 7** and **Figure 8,** respectively; in the plots, only post-COVID ridership is shown in order to view the trends and patterns more clearly.

Figure 7 shows that for the full series period, most of the models for the Chicago Transit Authority produced acceptable forecasts, with the exception of the neural network models. Most of the models also outperformed the naïve method, with the exception of a few of the neural network models. No one method stood out as having particularly better performance than all the other methods, although the ETS and TBATS methods produced good forecasts earlier in the post-COVID period and the ensemble (hybrid) method produced good forecasts later in the post-COVID period. The time series plot of ridership shows clear seasonal patterns and a steady, positive trend, which likely contributed to the ability of the time series methods to produce acceptable forecasts.

Figure 8 shows that, for the Greater Cleveland Regional Transit Authority, few of the models produced acceptable results; however, all of the models outperformed the naïve method. No one method stood out as having better performance than all the rest, but the neural network method underperformed. The time series plot of ridership shows a generally positive trend but lacks clear seasonal patterns.

Overall for the 14 agencies, the ARIMA and ensemble methods outperformed the other methods for post-COVID ridership forecasting using the full data series. Roughly 50-55% of the models for all methods produced acceptable forecasts, but for the ARIMA

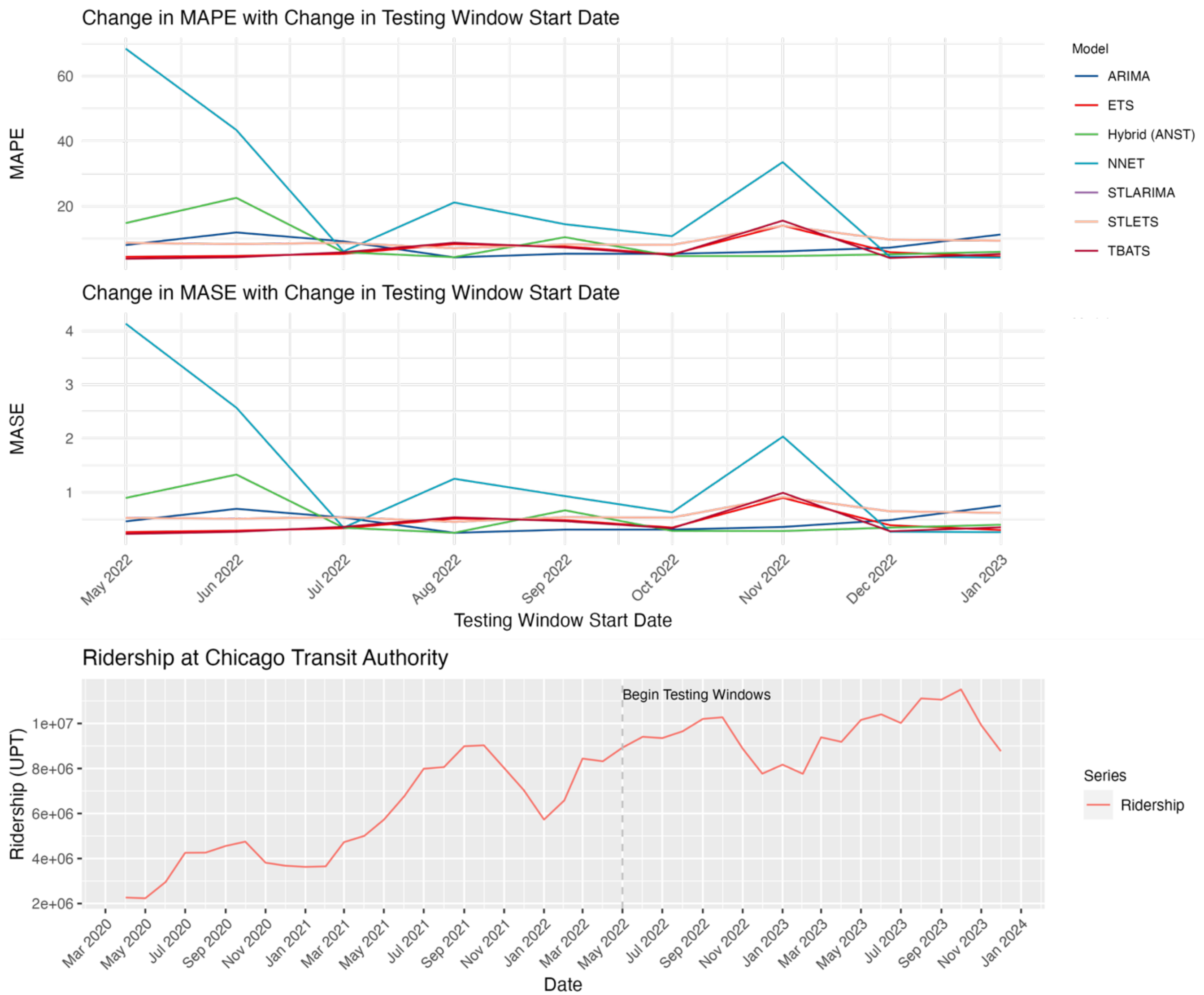

Figure 7 Full Series Change in Errors at the Chicago Transit Authority

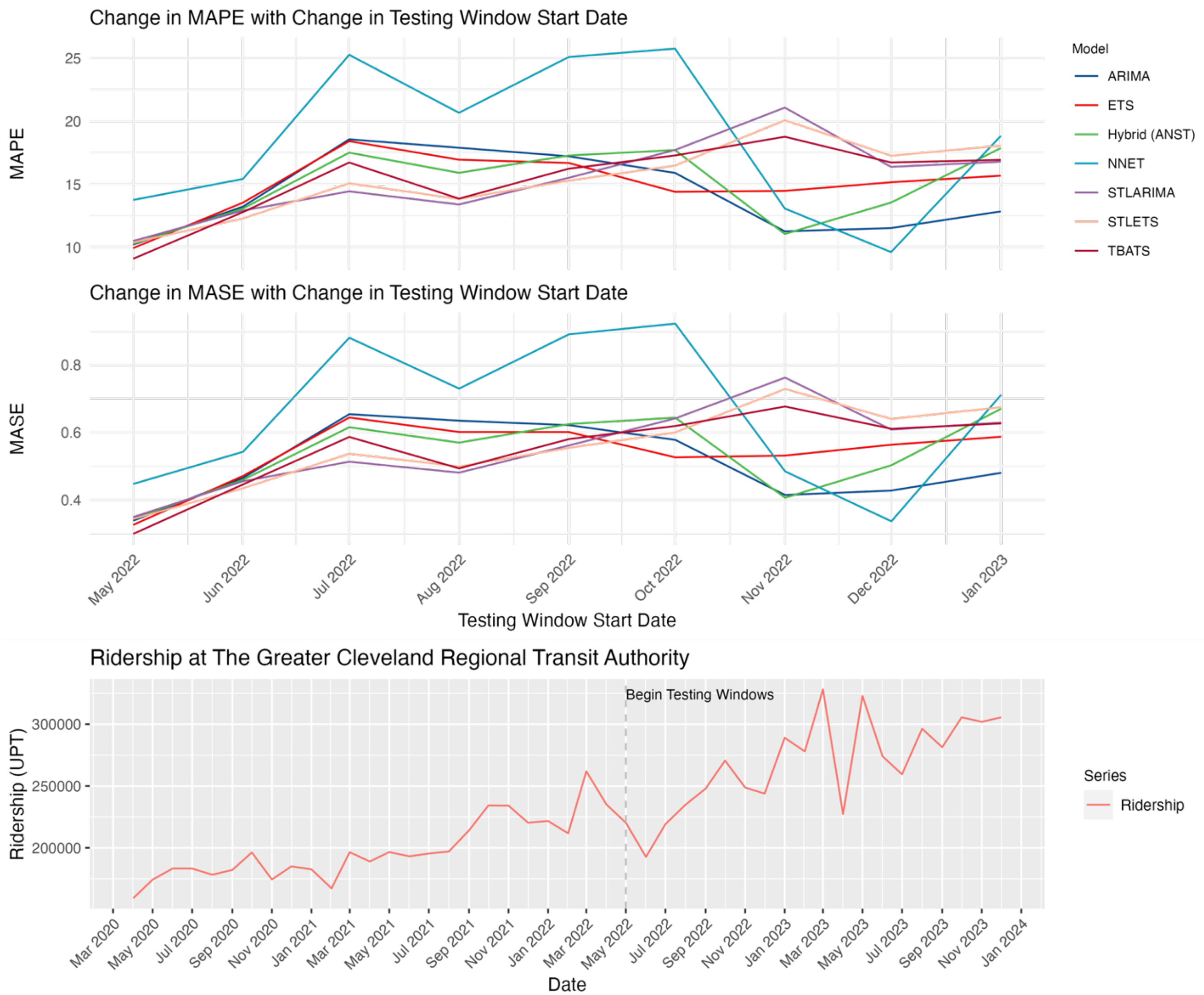

Figure 8 Full Series Change in Errors at the Greater Cleveland Regional Transit Authority

method, 64% of the models produced acceptable forecasts. Roughly 5-10% of the models for all methods produced good forecasts, but roughly 25% of the models using the ARIMA or ensemble methods produced good forecasts. The neural network underperformed, with less than 25% of the neural network models producing an acceptable forecast. The neural network also struggled to outperform the naïve method.

Results of the Post-COVID Analysis

Similar to the previous section, nine sets of forecasts were estimated for post-COVID ridership, but now using only post-COVID data (May 2020 to December 2023). As described previously, because of the short time span between the onset of the pandemic and the time of this analysis, there were only 45 monthly observations for the post-COVID period, 25 of which were required to train the models. Therefore, only nine sets of forecasts could be estimated. The errors for the Chicago Transit Authority and the Greater Cleveland Regional Transit Authority are once again shown in **Figure 9** and **Figure 10**, respectively; the upper plots are the change in error according to which testing/forecasting window was used, and the bottom plot is the ridership in unlinked passenger trips.

Figure 9 shows that most of the models for the Chicago Transit Authority produced acceptable forecasts when the training data included observations at least through July 2022. However, for the rest of the models, the forecasts did not produce acceptable MAPEs. The ETS and STL-ETS methods were the exception; opposite to the rest of the methods, the ETS method produced acceptable forecasts when the training data did not include observations beyond July 2022. The STL-ETS method never produced an acceptable forecast. Notably, the neural network method outperformed the rest of the methods, with all but one forecast producing acceptable MAPEs. All models outperformed the naïve method, likely due to Chicago's clear seasonal patterns and steady, positive trend.

Figure 10 shows that, for the Greater Cleveland Regional Transit Authority, almost none of the models produced acceptable results, nor did many of them outperform the naïve method. Interestingly, about half of the models with forecasts starting on November 2022 produced an acceptable MAPE and outperformed the naïve method; all but the neural network improved when moving the testing/forecast window from October to November 2022. This result contrasted with the full series forecast for the same forecasting windows; in **Figure 8,** the ETS, STL, and TBATS models worsened or stayed the same, while the ARIMA, neural network, and ensemble models improved. Also, none of the full series models for this testing window produced acceptable forecasts, whereas four of the post-COVID models produced acceptable forecasts. Nevertheless, no method stood out as having the best performance for Cleveland. As discussed previously, this may be partially due to the data's positive trend but only vague seasonal patterns.

Overall for the 14 agencies, the forecasts for post-COVID ridership using only post-COVID data underperformed compared to the same forecasts using the full data series. Most methods generally produced acceptable forecasts for around 45-50% of the models; the exceptions were the ETS and STL-ARIMA methods, which only produced acceptable forecasts for about 29% and 35% of the models, respectively. The neural network produced good forecasts for about 15% of the models, and the ensemble and STL-ETS methods produced good forecasts for about 13% and 11% of the models, respectively. The ETS and STL-ARIMA methods only produced good forecasts for just over 1% of the models.

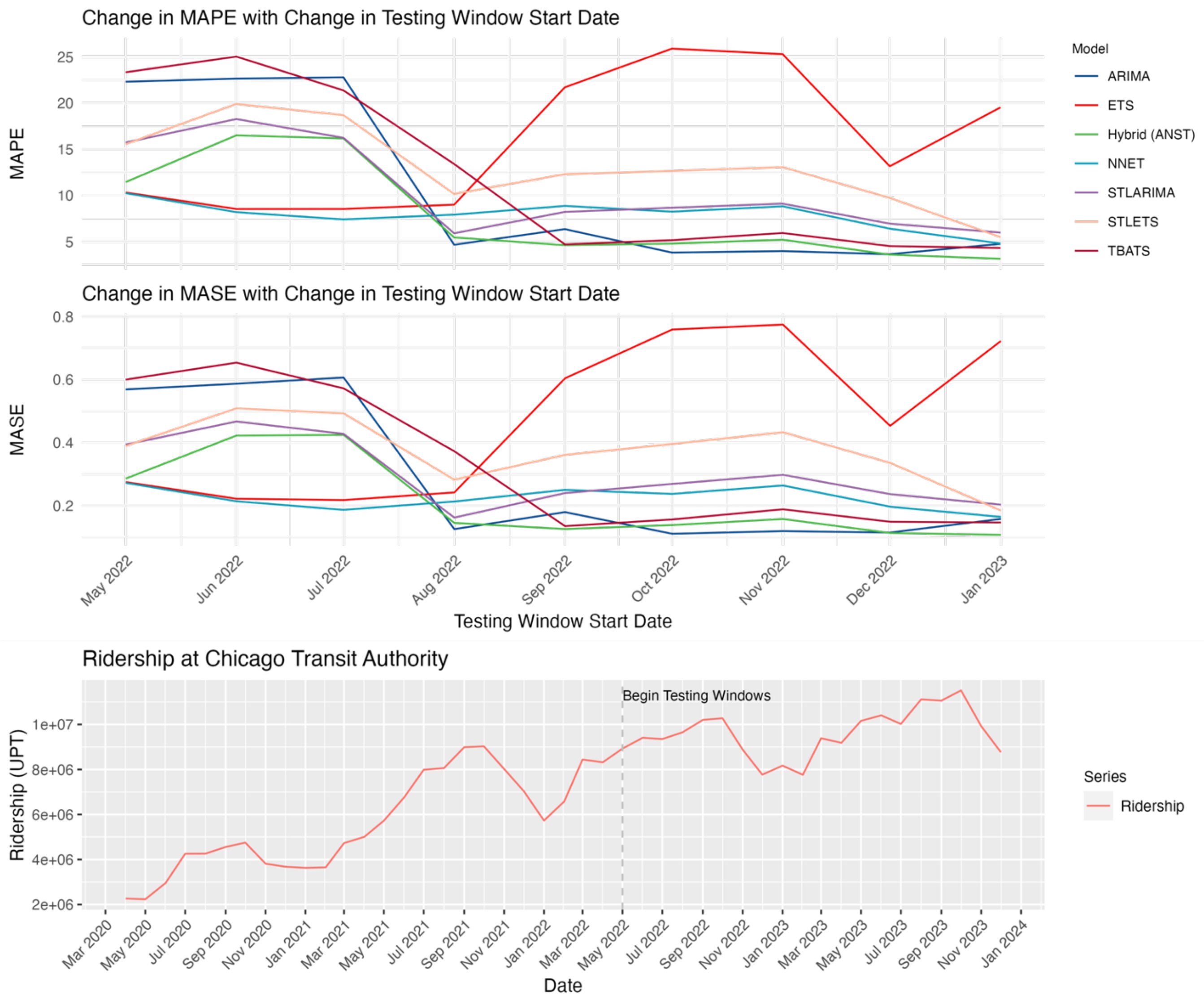

Figure 9 Post-COVID Change in Errors at the Chicago Transit Authority

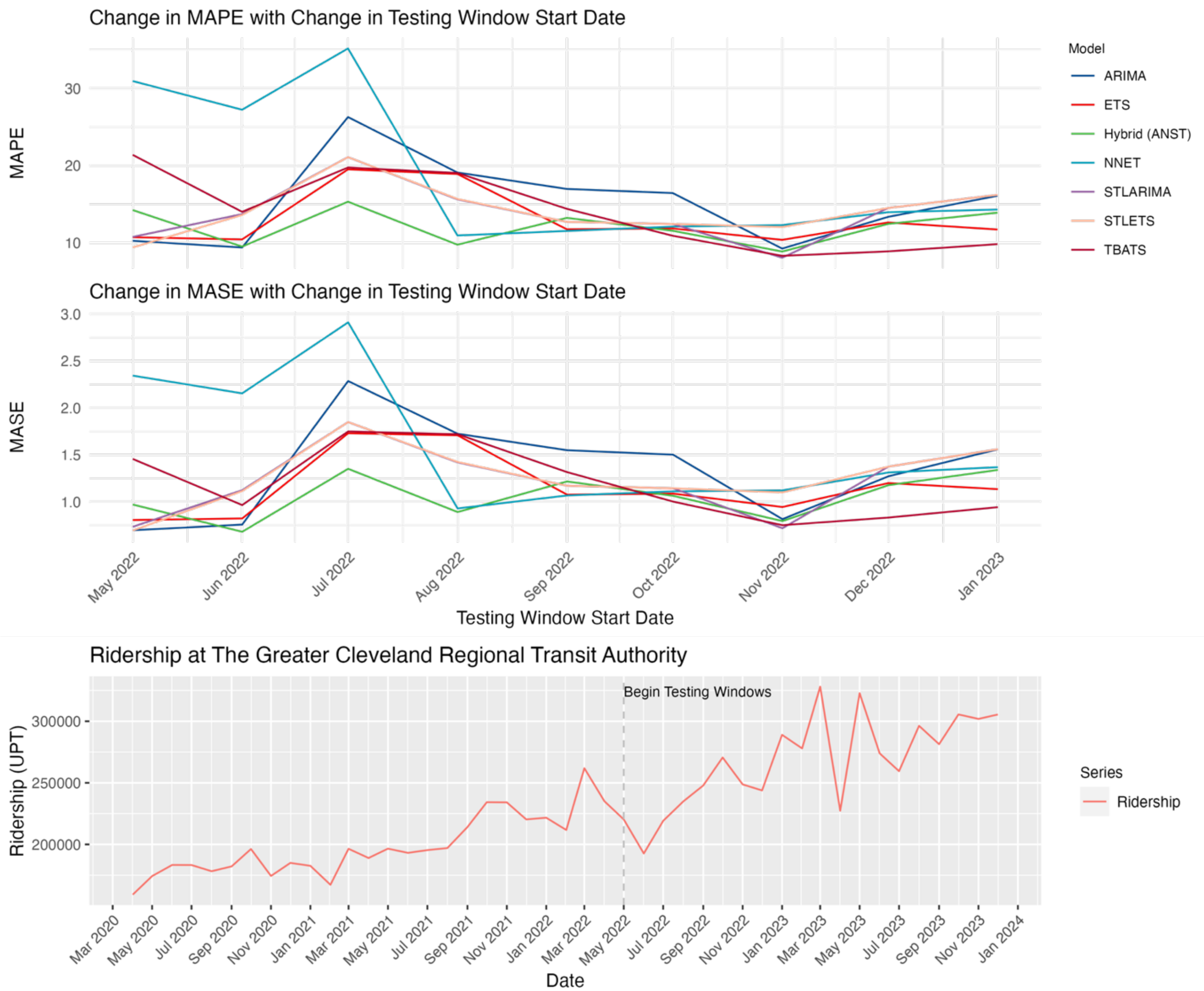

Figure 10 Post-COVID Change in Errors at the Greater Cleveland Regional Transit Authority

Overall Results

In summary, the time series methods used in this study produced good or acceptable forecasts for most heavy rail agencies before the onset of the COVID-19 pandemic. After COVID, ridership forecasting for heavy rail agencies was more challenging. In all cases, the good and acceptable forecasts tended to be concentrated at specific agencies, suggesting that the performance of time series forecasts may be dependent on the characteristics of the transit agency and its ridership. For example, for all forecasting periods, the forecasts for the agencies in Cleveland and Baltimore tended to be poor. In contrast, forecasting was more straightforward at some agencies such as those in Chicago, Miami, and Los Angeles. The characteristics that influenced forecasting performance appeared to be the stability of the trend and seasonal periods, which in turn were likely affected by factors like service disruptions, the scope of the service area, and disruptive events like the COVID-19 pandemic.

Table 3 summarizes the percent of good and acceptable forecasts by method and time period. The forecasts performed better when the training series was longer; however, the performance of the forecasts appeared to be sensitive to either the last training or testing observation (i.e., misleading training observations or greater changes in ridership month-over-month). For example, many of the forecasts in the full series period had a decrease in performance when the testing/forecasting window began with the observation from October or November 2022. In most cases, there did not appear to be one common factor among all transit agencies that caused better or worse performance than other months; rather, the changing performance appeared to be random. However, one explanation for the changing performance at individual agencies could be competing trends in the time series data, e.g., seasonal fluctuations in ridership being masked by smaller increases in ridership recovery in late 2022. Additionally, the neural network method underperformed for the pre-COVID and full series time periods, but slightly outperformed the other methods for the post-COVID time period. One potential reason for the neural network's mixed performance could be that its "P" parameter for the number of seasonal lags was set to one (as opposed to 12) in order to avoid extreme overfitting when forecasting only with post-COVID data. In the post-COVID data, there may have been a loss of seasonal patterns at many agencies, so using only one seasonal lag could have given the neural network models an advantage, assuming the previous month's ridership became a better predictor of future ridership than that of the previous year *(30)*.

Table 3 also shows that the ARIMA and ensemble (hybrid) methods exhibited relatively strong performance for the full series time period. Moreover, the ensemble method performed relatively well for all time periods. The good performance of the ensemble method may imply that no one forecasting method was able to adequately capture all patterns in the time series data, but by combining the methods, perhaps the errors of the individual models could be canceled out *(46)*.

Table 3 Percent of Models that were Good (MAPE < 5%) or Acceptable (MAPE < 10%) by Method and Time Period

Time Period	Performance	ARIMA	ETS	STL-ARIMA	STL-ETS	TBATS	NNET	Hybrid (ANST)
Pre-COVID	MAPE <10%	91.1	89.9	91.1	90.5	89.9	87.5	92.9
	MAPE < 5%	56.5	63.1	67.3	67.3	58.9	42.9	67.3
Full Series	MAPE < 10%	64.3	54	52.4	52.4	52.4	24.6	49.2
	MAPE < 5%	24.6	9.5	4	4.8	10.3	6.3	25.4
Post-COVID	MAPE < 10%	44.4	28.6	35.7	46	47.6	51.6	50
	MAPE < 5%	7.9	2.4	1.6	11.1	7.9	15.1	12.7

These findings are consistent with previous studies of time series forecasting performance, both for transit ridership as well as other applications. Some pre-COVID studies that compared the performance of time series forecasting methods for transit ridership likewise found that combined methods of ridership forecasting, like the ensemble method, tended to outperform the other methods *(28; 29)*. Through the Makridakis Competitions, a series of open forecasting competitions, the performance of time series methods has been empirically evaluated and compared using over 100,000 diverse datasets and every major time series method available *(46; 49)*. These competitions have consistently found that machine learning methods generally do not outperform simpler methods *(46; 50)*. One of the reasons behind the underperformance of pure machine learning methods may be their tendency to overfit the data. Additionally, pure statistical and machine learning methods were found to both underperform compared to hybrid methods, especially those that combine statistical and machine learning methods *(46)*. Another relevant finding was that the performance of the forecasts depended on the length of the forecasting horizon *(50)*.

CONCLUSIONS AND FUTURE RESEARCH

This study compared the performance of seven time series methods for univariate ridership forecasting at the 14 heavy rail agencies in the continental US. Three forecasting time periods were examined: pre-COVID (prior to March 2020), full series (January 2002 to December 2023), and post-COVID data (after March 2020). Nearly 3,000 forecasts were estimated in order to understand the changing performance of time series forecasting methods for transit ridership over time. The analysis revealed four major findings.

First, in the pre-COVID era, forecasting transit ridership using univariate time series methods was relatively straightforward; 90% of the models produced good or acceptable pre-COVID forecasts.

Second, the performance of the time series methods decreased in the post-COVID era, although the models that trained off the full data series (i.e., pre- and post-COVID data) overall produced slightly better forecasts compared to the models that trained only off post-COVID data.

Third, each individual method had changing performance according to which time period was used for model training. In the pre-COVID period, no one method outperformed all the rest, but the neural network slightly underperformed. Using the full series, the ARIMA and ensemble (hybrid) methods outperformed the rest of the models, and the neural network notably underperformed. However, using only post-COVID data to train the models, the neural network and ensemble methods outperformed all the other methods. The neural network may have outperformed the other methods when trained only on the post-COVID data because it was set to use only one seasonal lag. This implies that, in the post-COVID era, at some agencies, the previous month's ridership may be a better indicator of future ridership compared to the previous year's ridership. The ensemble method consistently performed relatively well for all time periods, perhaps due to its ability to minimize the effect of the errors from any one forecasting method

Fourth, there were differences in overall performance of time series forecasting methods by agency. At some agencies, the time series forecasting methods used in this study tended to produce acceptable results regardless of which time period was used to train the models. In contrast, at other agencies with more limited heavy rail service, for every combination of method or time period almost none of the models produced an acceptable forecast, and many of the forecasts did not even outperform the simpler naïve method.

In summary, this study demonstrates the general applicability of time series forecasting for heavy rail ridership. Univariate time series forecasts like the ones utilized in this paper are likely to be most suitable for agencies whose ridership data exhibit stable seasonality, periodicity, and trends. Agencies who are interested in making time- and cost-efficient forecasts need only to understand their own ridership data in order to forecast with a univariate method. By using either Minneapolis Metro Transit's forecasting tool or the publicly available code written for this paper, transit planners and practitioners could immediately begin generating their own time series forecasts to inform decision making regarding annual budgets, service levels, staffing needs, and other similar tasks.

Several areas for future research have emerged from this study. Future research should focus on forecasting ridership at smaller agencies, such as those in Baltimore and Cleveland. Additional methods of forecasting should be considered, and the performance of time series forecasts should be compared to that of the most common methods, such as the four-step, activity-based, and regression-based models. Last, time series forecasting should be tested for other transit modes, such as bus, light rail, or commuter rail.

ACKNOWLEDGEMENTS

ChatGPT was used to help debug the functions written in R. This work was funded by the Dwight D. Eisenhower Transportation Fellowship Program Award No. 693JJ32445014.

AUTHOR CONTRIBUTION STATEMENT

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

REFERENCES

1. Hammon, M. The State of Public Transit Ridership Across the US. *Planetizen*. https://www.planetizen.com/news/2023/12/126731-state-public-transit-ridership-across-us. 2023. Accessed March 5, 2024.

2. Rowlands, D., and T. H. Loh. Ensuring the intertwined post-pandemic recoveries of downtowns and transit systems. Brookings Institute, 2023. https://www.brookings.edu/articles/ensuring-the-intertwined-post-pandemic-recoveries-of-downtowns-and-transit-systems/. Accessed March 5, 2023.

3. Ziedan, A., C. Brakewood, and K. Watkins. Will transit recover? A retrospective study of nationwide ridership in the United States during the COVID-19 pandemic. *Journal of Public Transportation,* Vol. 25, 2023. https://doi.org/10.1016/j.jpubtr.2023.100046. Accessed March 5, 2024.

4. OAG. *COVID-19 Air Travel Recovery: How Did the COVID-19 Pandemic Affect Global Air Travel?* https://www.oag.com/coronavirus-airline-schedules-data#:~:text=At%20the%20very%20lowest%20point,in%202022%20in%20some%20regions. Accessed March 6, 2024.

5. USAFacts. *Has vehicle traffic returned to pre-pandemic levels?* USAFacts. https://usafacts.org/articles/has-vehicle-traffic-returned-to-pre-pandemic-levels/. Accessed March 6, 2024.

6. Cervero, R. Alternative Approaches to Modeling the Travel-Demand Impacts of Smart Growth. *Journal of the American Planning Association,* Vol. 72, 2006, pp. 285-295. https://doi.org/10.1080/01944360608976751. Accessed March 9, 2024.

7. Miller, E. The current state of activity-based travel demand modelling and some possible next steps. *Transport Reviews,* Vol. 43, No. 4, 2023. https://doi.org/10.1080/01441647.2023.2198458. Accessed March 9, 2024.

8. Pougala, J., T. Hillel, and M. Bierlaire. Parameter estimation for activity-based models. Presented at 22nd Swiss Transport Research Conference, Monte Verità, Ascona, Switzerland, 2022. Accessed May 9, 2024.

9. Federal Transit Administration. *STOPS – FTA's Simplified Trips-on-Project Software.* Federal Transit Administration, Washington, D.C.

https://www.transit.dot.gov/funding/grant-programs/capital-investments/stops. Accessed March 9, 2024.

10. Central Transportation Planning Staff. Appendix A Travel Demand Modeling. Boston Region Metropolitan Planning Organization. https://www.ctps.org/data/pdf/plans/LRTP/needs/2035NA_AppA.pdf. Accessed March 9, 2024.

11. Connecticut Department of Transportation. *Travel Demand Forecasting.* Connecticut Department of Transportation, Connecticut. https://portal.ct.gov/DOT/PP_SysInfo/Travel-Demand-Forecasting. Accessed March 9, 2024.

12. Sound Transit. ST3 Regional High-Capacity Transit System Plan: Transit Ridership Forecasting Methodology Report. Sound Transit. https://www.wsdot.wa.gov/partners/erp/background/ST3%20Draft%20RidershipForecastingMethodologyReport_6March2015.pdf. Accessed March 9, 2024.

13. Metropolitan Washington Council of Governments. *Travel Demand Modeling.* Metropolitan Washington Council of Governments, Washington, D.C. https://www.mwcog.org/transportation/data-and-tools/modeling/. Accessed March 9, 2024.

14. Twin Cities Metropolitan Council. *Travel Forecasting.* Twin Cities Metropolitan Council, St. Paul, MN. https://metrocouncil.org/Transportation/Performance/Travel-Forecasting.aspx. Accessed March 9, 2024.

15. Chicago Metropolitan Agency for Planning. *Modeling Data.* Chicago Metropolitan Agency for Planning, Chicago, IL. https://www.cmap.illinois.gov/data/transportation/modeling. Accessed March 9, 2024.

16. Miami-Dade Transportation Planning Organization. *ETAP Tools.* Miami-Dade Transportation Planning Organization. Accessed March 9, 2024.

17. Greenwood, C. City of Atlanta 2023 Q3 Update. MARTA. https://www.itsmarta.com/uploadedFiles/More/About_MARTA/2023%20Q3%20COA%2008.30.2023_presentation-v2.pdf. Accessed March 9, 2024.

18. New York Metropolitan Transportation Council. *Data and Modeling.* New York Metropolitan Transportation Council, New York City, NY. https://www.nymtc.org/en-us/DATA-AND-MODELING. Accessed March 9, 2024.

19. Doi, M., and W. B. Allen. A Time Series Analysis of Monthly Ridership for an Urban Rail Rapid Transit Line. *Transportation,* Vol. 13, 1986, pp. 257-269. https://doi.org/10.1007/BF00148619. Accessed July 1, 2022.

20. Mucci, A., and G. D. Erhardt. Evaluating the Ability of Transit Direct Ridership Models to Forecast Medium-Term Ridership Changes: Evidence from San Francisco. *Transportation Research Record,* Vol. 2672, No. 46, 2018. https://doi.org/10.1177/0361198118758632. Accessed October 23, 2022.

21. Chen, C.-H. P., and G. A. Naylor. Development of a Mode Choice Model for Bus Rapid Transit in Santa Clara County, California. *Journal of Public Transportation,* Vol. 14, 2011, pp. 41-62. https://doi.org/10.5038/2375-0901.14.3.3. Accessed October 23, 2022.

22. Hoque, J. M., I. Zhang, D. Schmitt, and G. D. Erhardt. Are Public Transit Investments Based on Accurate Forecasts? An Analysis of the Improving Trend of Transit Ridership Forecasts in the United States. Published, No. Issue, 2023. SSRN. https://dx.doi.org/10.2139/ssrn.4557555. Accessed Access Date.

23. Kyte, M., J. Stoner, and J. Cryer. A Time-Series Analysis of Public Transit Ridership in Portland, Oregon, 1971-1982. *Transportation Research Part A: General,* Vol. 22, No. 5, 1988, pp. 345-359. https://doi.org/10.1016/0191-2607(88)90012-X. Accessed July 1, 2022.

24. Ding, C., J. Duan, Y. Zhang, and G. Yu. Using an ARIMA-GARCH Modeling Approach to Improve Subway Short-Term Ridership Forecasting Accounting for Dynamic Volatility. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 19, No. 4, 2018, pp. 1054-1064. http://dx.doi.org/10.1109/TITS.2017.2711046. Accessed June 30, 2022.

25. Wang, X., N. Zhang, Y. Chen, and Y. Zhang. 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, 2018. doi:10.1063/1.5039099. Accessed June 22, 2022.

26. Ge, S. Y., C. J. Zheng, and M. M. Hou. Forecast of Bus Passenger Traffic Based on Exponential Smoothing and Trend Moving Average Method. *Applied Mechanics and Materials,* Vol. 433-435, 2013, pp. 1374-1378. http://dx.doi.org/10.4028/www.scientific.net/AMM.433-435.1374. Accessed July 1, 2022.

27. Egu, O., and P. Bonnel. Medium-term Public Transit Route Ridership Forecasting: What, how and why? A Case Study in Lyon. *Transport Policy,* Vol. 105, 2021, pp. 124-133. https://doi.org/10.1016/j.tranpol.2021.03.002. Accessed June 30, 2022.

28. Chen, D., J. Zhang, and S. Jiang. Forecasting the Short-Term Metro Ridership with Seasonal and Trend Decomposition Using Loess and LSTM Neural Networks. *IEEE Access,* Vol. 8, 2020. 10.1109/ACCESS.2020.2995044. Accessed June 30, 2022.

29. Chiang, W.-C., R. A. Russell, and T. L. Urban. Forecasting ridership for a metropolitan transit authority. *Transportation Research Part A: Policy and Practice,* Vol. 45, No. 7, 2011, pp. 696-705. https://doi.org/10.1016/j.tra.2011.04.016. Accessed March 20, 2024.

30. Hightower, A., A. Ziedan, J. Guo, X. Zhu, and C. Brakewood. A comparison of time series methods for post-COVID transit ridership forecasting. *Journal of Public Transportation*, 2024. https://doi.org/10.1016/j.jpubtr.2024.100097. Accessed June 30, 2024.

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

32. 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.

33. Minneapolis Metro Transit. *Route Trends and Info*, Minneapolis. https://metrotransitmn.shinyapps.io/route-trends/. Accessed July 13, 2022.

34. R Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/. Accessed July 25, 2022.

35. Lind, E., J. Reid, and K. E. Ky. *MetroTransit / Route-Trends, Build Complete Base Function App*. https://github.com/metrotransit/route-trends/commit/752d7b50fd250d66e334c8dc31487890f8d1b8ab. Accessed December 14, 2023.

36. Hyndman, R. J., and G. Athanasopoulos. *Forecasting: Principles and Practice, 2nd Edition*. Melbourne, Australia. OTexts.com/fpp2. 2018. Accessed December 15, 2023.

37. Livera, A. M. D., R. J. Hyndman, and R. D. Snyder. Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association,* Vol. 106, No. No. 496, 2011, pp. 1513-1527. Accessed November 11, 2023.

38. Hyndman, R., and G. Athanasopoulos. 8.8 Forecasting. In *Forecasting: principles and practice, 2nd edition*, OTexts, Melbourne, Australia, 2018. Accessed December 15, 2023.

39. Yule, G. U. On the Time-Correlation Problem, with Especial Reference to the Variate- Difference Correlation Method. *Journal of the Royal Statistical Society,* Vol. 84, 1921, pp. 497-537. https://doi.org/10.2307/2341101. Accessed March 10, 2024.

40. Yule, G. U. On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character,* Vol. 226, 1927, pp. 267-298. https://www.jstor.org/stable/91170. Accessed March 10, 2024.

41. Polhemus, N. W. Time series analysis of local fluctuations in traffic parameters. *Transportation Research,* Vol. 10, No. 5, 1976, pp. 311-317. https://doi.org/10.1016/0041-1647(76)90110-6. Accessed March 10, 2024.

42. Hyndman, R. J., and G. Athanasopoulos. A taxonomy of exponential smoothing methods. In *Forecasting: principles and practice, 2nd edition*, OTexts, Melbourne, Australia, 2018. Accessed December 15, 2023.

43. Hyndman, R. J., and G. Athanasopoulos. 6.8 Forecasting with decomposition. In *Forecasting: principles and practice, 2nd edition*, OTexts, Melbourne, Australia, 2018. Accessed December 15, 2023.

44. Hyndman, R. J., and G. Athanasopoulos. 11.3 Neural network models. In *Forecasting: principles and practice, 2nd edition*, OTexts, Melbourne, Australia, 2018. Accessed December 15, 2023.

45. Hyndman, R. J. *Prediction Intervals for NNETAR Models*. https://robjhyndman.com/hyndsight/nnetar-prediction-intervals/. Accessed July 13, 2022.

46. Makridakis, S., E. Spiliotis, and A. Assimakopoulos. The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting,* Vol. 36, No. 1, 2020, pp. 54-74. https://doi.org/10.1016/j.ijforecast.2019.04.014. Accessed March 20, 2024.

47. Shaub, 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.

48. Greater Cleveland RTA. *RTA to re-open Red Line between Airport and West Blvd. beginning Monday morning.* Greater Cleveland RTA, Cleveland, OH. https://www.riderta.com/news/rta-re-open-red-line-between-airport-and-west-blvd-beginning-monday-morning. Accessed March 24, 2024.

49. Makridakis Open Forecasting Center. *Makridakis Competitions*. https://mofc.unic.ac.cy/history-of-competitions/.

50. Hyndman, R. J. A brief history of forecasting competitions. *International Journal of Forecasting,* Vol. 36, No. 1, 2020, pp. 7-14. https://doi.org/10.1016/j.ijforecast.2019.03.015. Accessed March 20, 2024.