Proceedings of CONF-CIAP 2026 Symposium: Applied Mathematics and Statistics
DOI: 10.54254/2753-8818/2026.CH31530

Forecasting Urban Traffic in Adelaide with Autoregressive Models and Exogenous Regressors

Iris Shi

Acton Boxborough Regional High School, Acton, USA
iris.lingshan.shi@gmail.com

Abstract. Predicting traffic volumes is crucial for reducing congestion on roads while avoiding unnecessary over-spending on road infrastructure. This study uses AutoRegressive (AR), AutoRegressive Integrated Moving Average (ARIMA), and Seasonal ARIMA models with eXogenous regressors (SARIMAX) models to make hourly traffic volume predictions at seven locations in the vicinity of the Adelaide Central Market; the AR and ARIMA models were trained on daily data, and the SARIMAX model was trained on hourly data. For improved performance, a logarithm return function was combined with a z-score test and an IQR test to remove data outliers before using time-weighted linear interpolation or exponential smoothing to fill in missing data. Statistical tests such as the Augmented Dickey-Fuller (ADF) and Partial Autocorrelation Function (PACF) were applied to the models to determine their optimal coefficients. The accuracy of each model was measured using R-square (R^2), root mean square error (RMSE), mean absolute percentage error (MAPE), and symmetric MAPE (sMAPE). All three models had lower MAPE and sMAPE values, all less than 20%. The SARIMAX produced the best R^2 results with all values at least 0.90, while the other two models performed poorly in this test. Overall, the SARIMAX model yielded the most productive results; urban planners can use this model to forecast future traffic volumes and make adjustments to road networks to mitigate congestion.

Keywords: Urban traffic forecasting, traffic volume prediction, time series analysis, ARIMA, SARIMAX, Fourier series

1. Introduction

Traffic is the movement of different forms of transportation—vehicles, pedestrians, and cyclists—along roads. Urban traffic is more specifically known for the dense population of people and activities in a small area. This type of traffic is often characterized by a more complex cohesion of different transportation systems, accompanied by more traffic signals, pedestrian and bike crossings, and other frequent interactions that slow movement and increase travel times.

With an increase in automobile production, more and more vehicles are occupying the roads, leading to greater traffic congestion. Congestion occurs when the volume of traffic exceeds the road's maximum capacity. It is usually identifiable when cars are stopped for a long time or go through a constant cycle of starting and stopping. Although uncontrollable events, such as road work and weather, can congest roads, the sheer number of cars on a roadway can cause stoppages. During

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.CH31530

peak morning commute-to-work hours, the average speed of traffic decreased to less than 15 km per hour (less than ten miles per hour) in the commercial and business districts of many cities [1]. In addition, automobile accidents are more likely to occur when traffic is slow, further increasing congestion. The start and stop cycle of cars also contributes to higher carbon emissions, leading to greater air pollution and environmental damage in general [2]. Zhang & Batterman found in their simulation of traffic congestion that as the number of vehicles per hour increased, the amount of NO_2 released into the atmosphere increased regardless of the morning or afternoon rush hour or highway or arterial traffic. Additionally, as the simulated emission of this air pollutant increased, the number of predicted emergency doctor visits, hospital admissions, and mortality also rose, affecting not only the population on the road but also those living nearby [3].

When people spend more time commuting, they make less productive contributions and, as a result, weaken the economy [4]. Simply sitting in traffic causes missed appointments, less accurate pick-up and delivery times, and an overall negative impact on businesses and consumers that rely on explicit timing [2]. In fact, from a survey of 1,528 residents in Beijing, China, Wang & Liu found that there was a 1.1% increase in the likelihood of depression for every additional ten minutes waiting in traffic [5]. Even a single pedestrian crossing the street can cause a stubborn blockage on a road that has not exceeded its maximum capacity. Therefore, it is increasingly important for urban planners to understand the general trends in road and intersection traffic to account for long-term congestion, not just day-to-day stops. These planners can use the results of this study to determine the best course of action: providing additional infrastructure to reduce congestion, such as implementing urban ring roads or car parks, or managing the demand for transportation, such as improving the traffic signal system or setting up high occupancy lanes [1].

AutoRegressive (AR) models have been used to generalize data into patterns and use those data to predict outcomes. In context, urban planners can use physical road features and time series data to predict the volume of traffic on a road or at an intersection at a given time [6-9]. This study looks at these models to predict hourly traffic volumes at various locations in the Adelaide Central Market neighborhood in Adelaide, Australia. The models used in this study are the AR, AR Integrated Moving Average (ARIMA), and Seasonal ARIMA with eXogenous regressors (SARIMAX) models.

This paper is broken into six main sections: the introduction, literature review, methodology, results and discussions, and conclusions and future work. The following section is the literature review, which examines the viability of using the aforementioned models to predict traffic and recognizes the differences between those prior studies and this study. The following module outlines the methodology, explaining how the initial raw data is processed into usable data for each model. After data processing, the introduction of the AR models follows, along with the statistical tests that determine the optimal parameters for the models and the performance evaluation statistics that assess their forecasting accuracy. Following the methodology, the results and discussions evaluate each model's outcomes and explain their significance. Finally, the conclusion and future work section summarizes the key findings, discusses the implications of the research, and suggests future directions with this study.

The main focus of this paper is to evaluate the accuracy of the models used to predict long-term traffic volumes and to determine which models made predictions with the highest accuracy. Therefore, the following hypotheses were constructed:

Hypothesis 1All of the AR models can predict hourly traffic volumes in the Adelaide Central Market area with at least moderate statistical significance.

Hypothesis 2The ARIMA model demonstrates superiority over the AR model due to its complexity, but overall, the SARIMAX model will make the most purposeful predictions and forecasts.

2. Literature review

Pun et al. used multivariable linear regression and random forest regression to estimate urban traffic flow in Hong Kong. In both regression models, they used the topological and geometric features of the roads to predict traffic flow. They used these models to predict traffic flow for three data sets: GPS-based truck trajectory data (representing private traffic flow), public transport route data (representing public traffic flow), and AADT data (representing total traffic flow). In their results, they found that the R values in the multivariable linear regression (0.80, 0.66, and 0.66, respectively in the training dataset and 0.85, 0.64, and 0.47, respectively in the test dataset) were often greater than the R values in the random forest regression (0.76, 0.60, and 0.52, respectively in the training dataset and 0.82, 0.63, and 0.57, respectively in the test dataset), suggesting that the relationship between the predictors and the traffic flow was largely linear; albeit for the most part, both of these models performed far better than the single-variable linear regressions for each predictor variable. However, their work does not consider traffic constraints, such as turns and the direction vehicles are moving, that may have a significant impact on urban traffic flow. Additionally, this study was done only in Hong Kong, so other studies are needed to generalize their conclusions [6]. Although linear regression seems like a viable method to predict traffic, it is not suitable for this study because there was not a sufficient amount of topological data collected.

Akhtar et al. developed a seasonal ARIMA model to forecast short-term traffic volumes on the Eastern freeway westbound corridor of Melbourne using three months of data. Using fifty days of 15-minute time intervals, they used the model to test their predictions for the next thirteen days of data. They determined that their best model, ARIMA(6,0,0) x $(0,1,1)_{96}$, which has a seasonal lag of 96 data points or 24 hours, had a mean absolute percentage error (MAPE) of 8.38%, meaning the model has high accuracy, as the MAPE is below 10%, which makes sense because the data expressed daily patterns and limited volatility from how small the time period of the dataset was. They concluded that the model can accurately predict short-term traffic on that particular freeway. However, the authors noted that addressing additional parameters, such as traffic speed, may further improve accuracy [7]. This study supports the use of seasonal ARIMA models for predicting traffic volumes; however, it does not address predicting traffic in the long term. Additionally, the lack of other accuracy-measuring statistics, besides MAPE, is concerning. Note that this study calculated the MAPE on the logarithm-transformed data, which is much lower than the MAPE applied when the data is scaled up, back to the original [10].

Katambire et al. collected data at Muhima Junction in Kigali City, 15-minute intervals of data between 7:00 am and 8:00 pm from January 1, 2021, to December 31, 2022. There were two methods of traffic volume predictions: ARIMA and Long Short-Term Memory (LSTM). Their results showed that the LSTM model was more effective than the ARIMA model, as indicated by the better resultant error metrics for MAPE, mean absolute error (MAE), and root mean square error RMSE (22.5%, 10.2, and 5.8, respectively, for the LSTM model and 24.2%, 10.8, and 9.1, respectively, for the ARIMA model) [8]. However, both models achieved similar accuracy in their predictions, making ARIMA almost as viable a time series forecast model as LSTM. Additionally, LSTM is a deep learning model that requires large datasets and significant computational resources, neither of which is exactly present in this study of traffic in Adelaide. Nevertheless, this study further supports the use of the time series model ARIMA to predict traffic volumes.

Bazrafshan et al. conducted a study to forecast monthly acceptable particulate matter concentrations in Zabol, Iran, using time-series models. They used a classical SARIMA and

DOI: 10.54254/2753-8818/2026.CH31530

SARIMAX, where the exogenous regressor was a Fourier series. They used spectral analysis to determine the strong periodicity of seasons, incorporated into the Fourier terms. Out of their simpler, non-hybrid models, the SARIMAX with the Fourier series performed the best when comparing metrics such as RMSE, MAE, etc [9]. This study supports the use of the Fourier series as an exogenous regressor for SARIMA, but the complexities of traffic data and dust concentrations are different matters. Therefore, one purpose of this study will be to determine the viability of using Fourier terms to predict seasons in traffic data.

These prior studies support the use of autoregressive models in predicting traffic volumes, as well as other models such as linear regression and LSTM. However, because of resource limitations, only the autoregressive models will be used in this study.

3. Methodology

Figure 1 illustrates the general process of the study, from extracting the relevant data from the dataset to checking the model's accuracy with performance metrics. Recall that the three models used in this study are AR, ARIMA, and SARIMA. Before training each model, the data is filtered by extracting only the relevant information from the selected locations, and then outliers are removed. This process is designed to optimize the resulting error statistics while still preserving a sufficient amount of data. The AR and ARIMA models use the same training and test data, which consists of the hourly traffic at 8:00 am every day. The SARIMAX model uses the hourly traffic data from all twenty-four hours of each day.

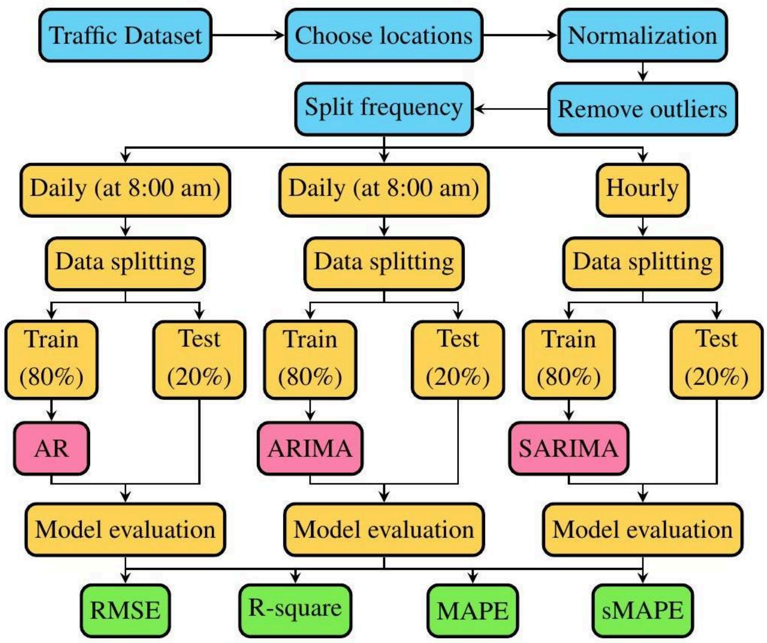

Figure 1. Traffic volume prediction pipeline

Seven various locations were studied to conclude that a single model works well for multiple locations. Figure 2 shows a map including all the intersections studied in this paper for the time series models. Out of all the locations, only seven were chosen for study because there were no long-term spikes or drops in their hourly traffic volumes, yet each location's traffic still varied from the others.

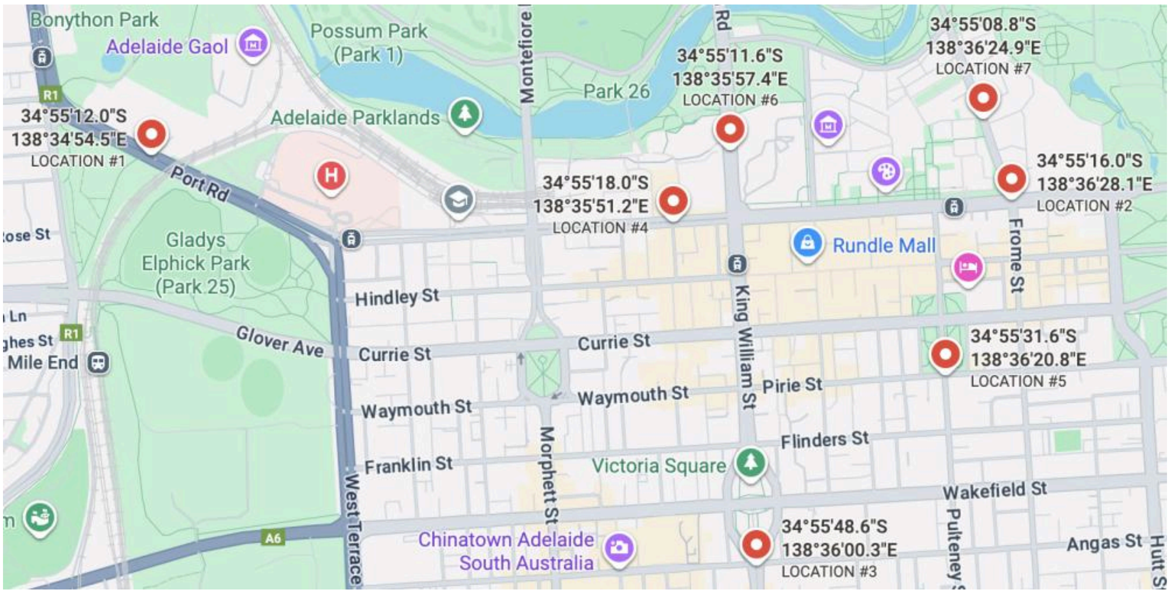

Figure 2. Locations of the intersections studied (map data: Google)

3.1. Data collection

The dataset used to train the models is from the South Australian Government Data Directory, titled "Traffic Intersection Volumes" and includes hourly data at 122 locations from January 1, 2010, to December 31, 2014 [11]. Table 1 shows the columns and a few rows of data extracted from the dataset. Some columns in the dataset, such as 'site_no', 't_detector_counts', and 't_error_counts', did not contribute to traffic prediction and were not imported with the rest of the data.

Table 1. Intersections in Adelaide, Australia (2010-2014) data

site_desc	lat	lon	rec_date	rec_hour	hourly_traffic
4 WAY INTERSECTION	-34.9292592	138.58786944	2011-01-18	16:00:00	6522.0
4 WAY INTERSECTION	-34.91759974	138.5930615	2013-04-04	20:00:00	1047.0
PEDESTRIAN CROSSING	-34.92769818	138.58914409	2010-04-25	3:00:00	375.0
SIGN. T-JUNCTION	-34.9217229	138.5959869	2010-06-09	18:00:00	1965.0
PEDESTRIAN CROSSING	-34.92104901	138.609361	2014-02-26	08:00:00	3808.0

3.2. Data precprocessing

Recall that the data used for the AR and ARIMA models is different from the data used in the SARIMAX model. Note that when removing outliers with the SARIMAX model, they are removed hourly and then recombined before the data is split into training and testing sets. Removing outliers from all the data combined at every hour would result in very few data points being removed. A logarithm transformation was applied to all the models because they performed better with it. Before conducting any statistical tests, all data from weekends were removed because weekends tended to have less hourly traffic than weekdays. Besides weekends, there are holidays and periods of road work, etc., where traffic volumes are lower than usual, so more statistical tests must be done to remove these outliers. After this removal, the logarithm return and the inter-quartile range (IQR) functions rigorously clean the data.

DOI: 10.54254/2753-8818/2026.CH31530

A logarithm returns z-score function, then an IQR function, was applied to the data to remove outliers. The logarithm return function examines consecutive times in a time series. In the case of these models, the function analyzed consecutive days in the data. The logarithm returns finds:

$$L_t = \ln \frac{T_{t+1}}{T_t} \tag{1}$$

where L_t is the logarithm return between times t and $t+1$ and T_t is the hourly traffic at time t (in days). A z-score is then calculated for each logarithm return with the formula:

$$z_t = \frac{L_t - \mu}{\sigma} \tag{2}$$

where z_t is the z-score at time t , μ is the mean of all the returns, and σ is the standard deviation (SD) of all the returns. Note that if n is the total number of returns, then the SD is calculated with the formula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (L_i - \mu)^2}{n}} \tag{3}$$

Now, if the absolute value of a z-score exceeds 1.96, the z-score at the 95% confidence level, or in other words, the logarithm return value is not within 1.96 standard deviations from the mean, then the data point T_t is removed. T_t is removed instead of T_{t+1} because all outliers are low outliers, so a high return value would mean T_t was a low outlier, while T_{t+1} was closer to the expected hourly traffic. However, the logarithm return function cannot detect a large cluster of outliers, as it only looks at consecutive values. The purpose of this outlier detection was to remove the single significant outliers, leaving the IQR function to remove the less obvious clusters of outliers.

The IQR is the difference between Q3 and Q1 or IQR = Q3 – Q1, where Q1 and Q3 are the first and third quartiles, respectively. Note that quartiles mark what fourth of the data is below that point (i.e., 25% of hourly traffic data falls below Q1 and 75% of data falls below Q3). Therefore, the IQR measures the range of the middle 50% of data. Any points that fall below Q1 – 1.5 × IQR or above Q3 + 1.5 × IQR are considered outliers and are removed.

However, after removing outliers, some data is missing, and a time series model cannot work with an incomplete set of data, i.e., there are holes in the data. Therefore, filler values were added to patch up the holes. For the AR and ARIMA models, wherever a hole was present, a time-weighted linear interpolation filled it. In other words, a line of data is drawn between the last known time and the immediately next known time, filling the holes with their respective points on that line. The data patching for the SARIMAX model is more elaborate. The first step of the process is to fill each data hole with the prior day's data. For example, the data for Saturday (and Sunday) is a copy of Friday's data; the traffic volumes at 10/31/09 7:00 pm are the same as the volumes at 10/30/09 7:00 pm, etc. However, filling the missing data with copies of prior data is not ideal because it can artificially reduce variability and introduce bias [12]. Therefore, exponential smoothing is used to add variance to the artificially-filled data.

Exponential smoothing is exceptional at ignoring minor discrepancies in the data, such as those found in copy-filled data, because the model aims to create a smoother prediction, reducing the impact of inconsistencies on the overall model. The specific smoothing model used in this study is the Holt-Winters model, which considers level, the average traffic volume during a particular period of time; trend, the direction in which the level moves over time, i.e., an increase or decrease; and seasonality, repeating patterns at fixed intervals.

The main seasonal period is 24 hours, determined using a variety of statistical tests explained in the next section, "Applied models". The Holt-Winters model performed the best when the trend was additive—meaning a linear increase/decrease in the level over time—, and the seasonality was additive—meaning the deviation of a data point from the level was calculated as a difference instead of as a proportion. After the Holt-Winters model was applied, the accuracy statistics applied to the SARIMAX model increased, on average, compared to when the model used the artificially-filled training data.

3.3. Applied models

3.3.1. Autoregressive model

AR models are time series models, meaning the only predictor for hourly traffic volumes is past data on traffic volumes. AR models specifically look at lags. For example, if an AR model uses lag 3, it examines the past three days' hourly traffic to predict the current day's traffic. Recall that the traffic volume data was logarithmically transformed before being trained in the model. This transformation allows for the parameters to have a multiplicative relationship instead of an additive relationship. After testing the AR model on both the non-transformed and transformed data, the model trained with the transformed data outperformed the other model for all seven studied locations.

The Partial Autocorrelation Function (PACF) helps determine the optimal parameter for the AR model—the variable name of this parameter is p . Figure 3 shows a diagram of the PACF plot for location #4. The diagrams for the other six locations have similar plots, featuring a large, positive peak at lag 0, followed by moderate peaks at lags 5, 6, 7, 8, and 9. Note that a cut-off after a certain lag usually indicates that that lag is the optimal value for p . Because there are peaks repeating every seven lags, it is safe to assume that $p = 7$ creates a good model. After testing $p = 5, 6, 7, 8, 9$, it was in fact true that $p = 7$ performed the best. Note that $p = 7$ means the past week of traffic data is used to predict each day's traffic. This finding makes sense because traffic has relatively consistent weekly patterns; for example, last Friday's traffic volume should be similar to this Friday's volume.

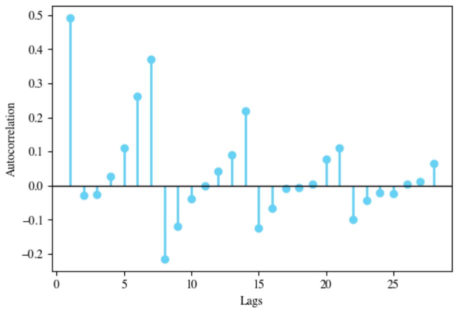

Figure 3. PACF plot at location #4

3.3.2. Autoregressive integrated moving average model

There are three components to an ARIMA model: the autoregressive (AR), integrated (I), and moving average (MA) parts. The integration portion determines how many times the data is differentiated before it is trained and tested. The primary purpose of this component is to prevent nonstationarity in the data, meaning the data has an increase or decrease in mean over time or any other kind of inconsistency. Note that applying a logarithm transformation makes the data appear more stationary than before the transformation. Typically, the more data is differenced, the more stationary it becomes. However, over-differencing may cause unexpected spikes or dips, reducing the accuracy of predictions. The parameter for integration is d , e.g., if then the data is differenced once before training and testing. Mathematically, given that T_{t-1} and T_t are the traffic volumes at two consecutive hours, their difference is:

$$T_t^{'} = T_t - T_{t-1} \tag{4}$$

Because the same data is used in the AR model as in the ARIMA model, their PACF plots are identical; hence, $p = 7$ is the optimal AR parameter. The MA portion is very similar to the AR, but this part looks at previously made errors instead of previous traffic volumes. The error for each day is the residual, calculated as the difference between the predicted and observed hourly traffic. q is the parameter for MA. For example, if $q = 1$, the model only considers the previous day's error to adjust the current day's prediction. Mathematically, the overall ARIMA model can be written as:

$$T_t^{(d)} = \phi_0 + \Sigma_{i=1}^{P} \phi_i T_{t-i}^{(d)} + \epsilon_t + \Sigma_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{5}$$

where $T_t^{(d)}$ is T_t after differencing d times, ϕ_0 is the constant term of the AR, ϕ_i is the coefficient of the hourly traffic at time $t-i$, ϵ_t is the error term for time t , and θ_i is the coefficient of the error term for hourly traffic at time $t-i$.

The Augmented Dickey–Fuller (ADF) test determines if the data is stationary. If the p-value of the ADF test is greater than 0.05, then the data is considered nonstationary and needs to be differenced. If the data is still not stationary—that is, this new p-value is still greater than 0.05—after this differencing, the data will be differenced once more. The data will never be differenced

DOI: 10.54254/2753-8818/2026.CH31530

more than twice, as overdifferencing can create unnecessary noise, lose information, and add extra AR and MA terms [13]. For all seven locations, the ADF test calculated p-values magnitudes smaller than 0.05, so none of these datasets required differencing. However, the ADF value was 0.003 for location #1 and 0.0006 for location #6, which are relatively close to 0.05. Comparing the results for the optimal ARIMA model at this location with both $d=0$ and $d=1$ showed that the model with $d=1$ performed better for all the accuracy statistics for both locations—significantly better for location #1 and slightly better for location #6. Therefore, $d=0$ was used for all the locations except #1 and #6, which had $d=1$.

The Autocorrelation Function (ACF) helps determine the optimal parameter q , which is the parameter of the MA that accounts for past errors. Figure 4 shows the ACF plot for location #1. Similar to the PACF plot, a cut-off after a certain lag usually indicates that that lag is the optimal value for q . However, with location #1 and location #6 (not shown), there is no obvious cut-off. This constant decrease suggests that the data is not stationary. This observation further supports the differentiation of the data at locations #1 and #6.

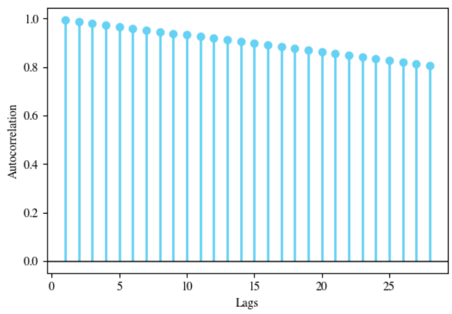

Figure 4. ACF plot at location #1

Now, other ACF plots will be observed to determine the optimal number of lags for q . Looking at the ACF plot at location #3, shown in Figure 5, it seems there are significant peaks at lags 1, 7, 14, 21, and 28. Just like with the lags for p , lags 1, 2, 3, 4, 5, 6, and 7 were tested, and $q=7$ showed the best results. Therefore, the model $ARIMA(p,d,q) = ARIMA(7,d,7)$ was trained and tested for results—recall that $d=0$ for all the locations except location #1 and #6, which had $d=1$.

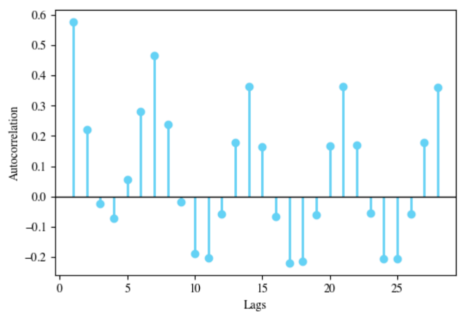

Figure 5. ACF plot at location #3

DOI: 10.54254/2753-8818/2026.CH31530

3.3.3. Seasonal autoregressive integrated moving average model with exogenous regressors

SARIMAX models are ARIMA models with a seasonal component and exogenous regressors. In this study, they also differ in that the ARIMA model predicts daily frequencies while the SARIMAX model predicts hourly frequencies. However, the process of finding the optimal model is the same, not considering the additional components. The data for each location is checked for stationarity using the ADF test. For each site, the ADF p-value was far less than 0.05, so it can be assumed that the data for all of them are stationary ($d = 0$). The general SARIMA model, the SARIMAX model without the exogenous regressors, can be written as:

$$T_t^{(d,D)} = \left(\phi_0 + \Sigma_{i=1}^{p}\phi_i T_{i=1}^{(d,D)} + \Sigma_{i=1}^{P}\Phi_i T_{t-is}^{(d,D)}\right) + \left(\epsilon_t + \Sigma_{i=1}^{q}\theta_i\epsilon_{t-i} + \Sigma_{i=1}^{Q}\Theta_i\epsilon_{t-is}\right) \tag{6}$$

Note that P , D , and Q are parameters for the seasonal AR, I, and MA, respectively. s is the length of the season (in hours), Φ_i is the seasonal coefficient of the hourly traffic at time t (i.e., i "seasons" before t), and Θ_i is the seasonal coefficient of the error term for hourly traffic at time $t - is$. Note that the data is differenced d times normally and D times seasonally, meaning the difference is found between times that are s hours apart instead of just 1 hour apart, as with standard differencing. Note that there is no statistical test, such as the ADF test, to determine if a model needs seasonal differencing, so a guess-and-check method is used to optimize the model. The seasonal AR examines P lags that are each s hours apart, while the seasonal MA examines Q lags of errors that are also each s hours apart.

Figure 6 shows the PACF plot for location #4. All of the locations have similar high peaks at lags 1 and 2, with almost negligible peaks for later lags, except for moderate peaks around 20 lags. However, these peaks are ignored because the resources in this study did not allow for such complex models to be tested. After testing $p = 1$ and $p = 2$, it was found that all of these SARIMAX models had very similar, almost indistinguishable results, but $p = 1$ performed slightly better than $p = 2$ so the nonseasonal AR is $p = 1$. Additionally, there is a noticeable negative peak at lag 24, suggesting that the season could have a length of 24. The length of the season is more evident on the ACF plot.

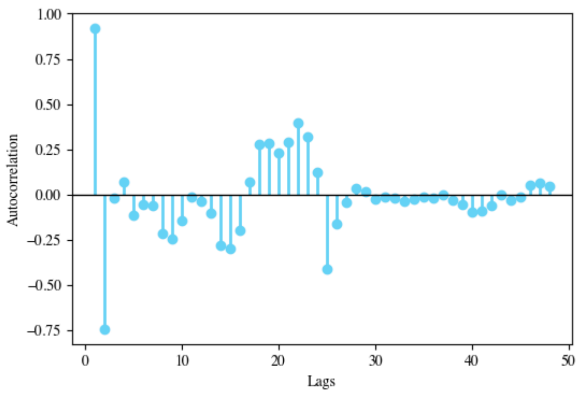

Figure 6. SARIMAX PACF plot at location #4

DOI: 10.54254/2753-8818/2026.CH31530

Figure 7 shows the ACF plot for location #2, although all the ACF plots have similar significant peaks every 24 lags. Therefore, it is strongly implied that the length of each season is 24 hours, so $s = 24$.

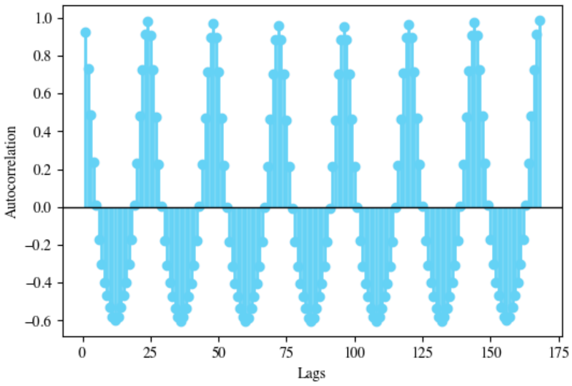

Figure 7. SARIMAX ACF plot at location #2

Recall that the "seasonality" of the ARIMA model was 7 days, which implies that there may be 168-hour (7-day) seasonality. Therefore, the final SARIMAX model uses both these seasonal lengths to predict traffic. Traditional SARIMA models cannot consider more than one season, unless a model like $SARIMA\,(p,d,q),(P,D,Q)_s = SARIMA\,(1{,}0,1),(7{,}0,7)_{24}$ was implemented—although, this model was tested and performed poorly compared to the final SARIMA model, which used an exogenous regressor. The final SARIMA model is $SARIMA\,(1{,}0,1),(0{,}0,0)_0$ without the exogenous regressor. The SARIMAX has a Fourier series exogenous regressor. A Fourier series is the sum of multiple sinusoidal graphs. The equation for it is:

$$f(t)_l = \Sigma_{i=1}^{K}\left[\alpha_i \sin\left(\frac{2\pi i t}{l}\right) + \beta_i \cos\left(\frac{2\pi i t}{l}\right)\right] \tag{7}$$

where K is the number of Fourier terms, l is the seasonal period, and α_i and β_i are the coefficients for the sine and cosine components, respectively. These coefficients are estimated through regression as part of the exogenous regressor matrix passed to the SARIMAX model. Then the general SARIMAX model can be written as:

$$T_t^{(d)} = ARIMA_t + f(t)_{24} + f(t)_{168} \tag{8}$$

Recall that the ARIMA portion of the SARIMAX model is $ARIMA\,(1,d,1)$, and the model has seasonal lengths of 24 and 168 hours. The ARIMA portion models the nonseasonal dynamics, while $f(t)_{24}$ and $f(t)_{168}$ handle the seasonality.

DOI: 10.54254/2753-8818/2026.CH31530

3.4. Performance evaluation

3.4.1. R-square (R^2)

$$R^2 = 1 - \frac{\frac{1}{n}\Sigma_{i=1}^{n}\left(T_i - \widehat{T_i}\right)^2}{\frac{1}{n}\Sigma_{i=1}^{n}\left(T_i - \overline{T}\right)^2} \tag{9}$$

where $\widehat{T_1}, \widehat{T_2}, \ldots, \widehat{T_n}$ are the predicted values, $T_1, T_2, \ldots, T_n$ are the observed values, $\overline{T}$ is the mean of all the observed hourly traffic, and n is the total number of observations. Note that the numerator is the square of the root mean squared error (RMSE), and the denominator measures the variance of the observed data from their mean. R^2 values range from 0 to 1. Models with a higher R^2 value are more reliable, as they show a higher correlation between the observed and predicted traffic volumes. Models with a low R^2 value show poor model performance. Note that when a model has a negative R^2 value, its prediction is worse than just predicting $\overline{T}$ for all observed values in terms of capturing the variance of the observed data.

3.4.2. Root mean square error

The RMSE measures the average magnitude of errors between the predicted and observed values in the models. The formula is stated below:

$$RMSE = \sqrt{\Sigma_{i=1}^{n}\frac{\left(\widehat{T_i} - T_i\right)^2}{n}} \tag{10}$$

Note that the difference between the predicted values and their respective observed values had to be squared to account for positive and negative numbers. A lower RMSE means a more accurate model, as it shows a smaller difference between the predicted and observed values.

Recall that the original traffic volumes went through a logarithm transformation, so the calculated expected RMSE of the transformed values will not actually show anything. Simply finding the exponent of the predicted values will result in an underestimate from the true RMSE; therefore, a smearing estimator is used [14]. Note that ϵ_i is the residual on the logarithm scale, calculated as $\epsilon_i = T_i - \widehat{T_i}$. Equation 11 and Equation 12 show the calculations for the smearing factor and the prediction on the regular, non-transformed scale, respectively:

$$SF = \frac{1}{n}\Sigma_{i=1}^{n} e^{\epsilon_i} \tag{11}$$

DOI: 10.54254/2753-8818/2026.CH31530

$$\left(\widehat{T_o}\right)_i = e^{\widehat{T_i}} \cdot SF \tag{12}$$

The resulting $\left(\widehat{T_o}\right)_i$ is then compared with $(T_o)_i$, the hourly traffic in the original, untransformed data, in the RMSE calculation (stated in Equation 10) in place of $\widehat{T_i}$ and T_i, respectively. Note that the smearing estimator is used similarly for the mean absolute percentage error (MAPE) and symmetric MAPE (sMAPE) calculations too, as the MAPE and sMAPE of the logarithm-transformed data are always going to be smaller than their respective results on the original data [10].

3.4.3. Mean absolute percentage error

The accuracy of a forecasting model is measured by the MAPE, which calculates the average percentage difference between predicted and actual values. Mathematically, it is represented as:

$$MAPE = \frac{1}{n} \Sigma_{i=1}^{n} \frac{\left|T_i - \widehat{T_i}\right|}{|T_i|} \tag{13}$$

The absolute value ensures that the magnitude of both negative and positive differences contributes to the error. The sum adds up all the errors across all n differences, then the whole sum is divided by n to find the mean of the errors. Note that this result is a decimal proportion, not the actual percentage. A lower MAPE indicates a more accurate forecasting prediction.

3.4.4. Symmetric mean absolute percentage error

The sMAPE measures the same thing as the MAPE, which is the accuracy of a model, except smaller observed traffic volumes are penalized less. This occurrence is true because, instead of the observed traffic volume being in the denominator alone for each summed term, the average of the observed and predicted volume is in the denominator:

$$SMAPE = \frac{1}{n} \Sigma_{i=1}^{n} \frac{\left|T_i - \widehat{T_i}\right|}{\frac{\left|T_i + \widehat{T_i}\right|}{2}} \tag{14}$$

Although most of the locations for each model show little difference in the MAPE and sMAPE values, sMAPE is essential for identifying systematic bias. Specifically, if sMAPE is smaller than MAPE, the model may have overestimated traffic volumes on average; conversely, if sMAPE is larger, it may tend to underestimate traffic volumes. However, MAPE and sMAPE are measured on different scales and are not directly comparable, so these observations should be interpreted with caution.

DOI: 10.54254/2753-8818/2026.CH31530

4. Results and discussions

The accuracy of all the AR models was evaluated using the R^2, RMSE, MAPE, and sMAPE statistics, where the data was first back-transformed to the original scale using the smearing estimator for the RMSE, MAPE, and sMAPE tests because applying these statistic tests on the logarithmically transformed data would produce better-than-expected results. Note that models with below 10% MAPE values have very accurate estimations, between 10% and 20% have relatively accurate estimations, between 20% and 50% have relatively reasonable estimations, and above 50% have inaccurate estimations [15]. sMAPE can be interpreted similarly.

4.1. Results with the AR model

Table 2 shows the performance evaluation statistics for all the locations. The average R^2, RMSE, MAPE, and sMAPE of all the locations are -0.14, 185.6, 10.6%, and 10.4%, respectively. The negative average R^2 and overall minimal R^2 values for all the locations indicate that the AR model did not capture the variance of the data at all. However, variance is not the factor that determines how viable the model is for forecasting future data; the MAPE and sMAPE results determine the model's forecasting feasibility. Locations #1, 2, 3, and 4 all have MAPE and sMAPE values less than 10%, meaning they are highly accurate. Locations #5, 6, and 7 have MAPE and sMAPE values less than 20%, so their AR models are considered relatively accurate. While the AR model achieved low MAPE values, its predictive accuracy is likely superficial, as it failed to capture fluctuations in the data. For every location, except for location #5, the sMAPE was slightly less than the MAPE, indicating that the model tended to overestimate its forecast.

Table 2. AR statistical accuracy test results for each location (#1-7)

Location	R^2	RMSE	MAPE	sMAPE
#1	-0.12	199.2	7.9%	7.5%
#2	0.05	204.1	9.7%	9.5%
#3	-0.07	141.7	7.4%	7.0%
#4	-0.12	248.4	9.1%	8.7%
#5	-0.66	286.2	17.8%	19.1%
#6	-0.09	139.5	11.9%	11.3%
#7	0.04	80.2	10.5%	9.8%

However, looking at Figure 8, it seems like the model is not actually exceptional at making predictions. There are some predictions for about 28 days, after which they flatline for the rest of the forecast. The low MAPE may be because the traffic volumes already have low variation to begin with. However, their magnitudes are high, with most hours having over a thousand vehicles in the original data. Therefore, it may seem positive to have such low MAPE and sMAPE values. Examining the MAPE and sMAPE results from a model that merely forecasted the mean of the training data, it is evident that the "accurate" MAPE and sMAPE results may be due to almost-uniform data.

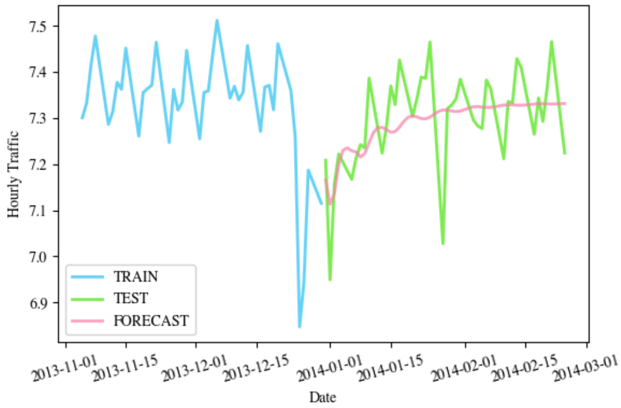

Figure 8. AR model forecast at location #3 results

Table 3 compares the MAPE and sMAPE of the AR model and the model that forecasts the mean. The results show that there is either no difference or only a minimal improvement from the mean model to the AR model, except for location #1. Therefore, the AR model is not very effective at predicting fluctuations in traffic. Recall that location #1 was close to being considered nonstationary by the ADF test. This nonstationarity is apparent in the graph of this AR model, as instead of a flat horizontal line, it is a linearly increasing line. Therefore, it makes sense that the mean model performed significantly worse for only location #1, as it is evident that an increasing linear forecast performs better than a horizontal line on increasing-nonstationary data.

Table 3. AR model MAPE and sMAPE for each location (#1-7) compared toMAPE and sMAPE of the training-data-mean-forecast model

	AR Model		Mean Model	
Location	MAPE	sMAPE	MAPE	sMAPE
#1	7.9%	7.5%	87.4%	155.4%
#2	9.7%	9.5%	10.2%	9.9%
#3	7.4%	7.0%	7.7%	7.3%
#4	9.1%	8.7%	9.1%	8.7%
#5	17.8%	19.1%	17.7%	18.9%
#6	11.9%	11.3%	12.9%	12.0%
#7	10.5%	9.8%	11.0%	10.1%

4.2. Results with the ARIMA model

Table 4 displays the RMSE, R^2, MAPE, and sMAPE of each of the seven locations. The average of the R^2, RMSE, MAPE, and sMAPE over all the locations are -0.13, 183.2, 10.4%, 10.4%, respectively. Most of these results are at least slightly better than the AR model's accuracy statistics (Table 3). This difference indicates that the ARIMA model is vaguely better than the mean model in some locations, i.e., locations #1, #3, and #4.

DOI: 10.54254/2753-8818/2026.CH31530

Table 4. ARIMA statistical accuracy test results for each location (#1-7)

Location	R^2	RMSE	MAPE	sMAPE
#1	-0.77	250.7	10.6%	11.1%
#2	0.13	195.2	9.3%	9.0%
#3	0.27	116.9	5.2%	4.8%
#4	0.20	209.2	7.2%	6.9%
#5	-0.74	293.5	18.8%	20.4%
#6	-0.05	137.2	11.2%	10.8%
#7	0.06	79.4	10.3%	9.6%

Figure 9 shows the last few points of training data, the first few points of test data, as well as the predicted (forecast) values from the ARIMA model at location #3. Compared to Figure 8, the forecast never flatlines; instead, it shows a 7-day periodic pattern. Even though the ARIMA model is not explicitly given a seasonal parameter, the model still predicts seasons because the data shows a repeating pattern.

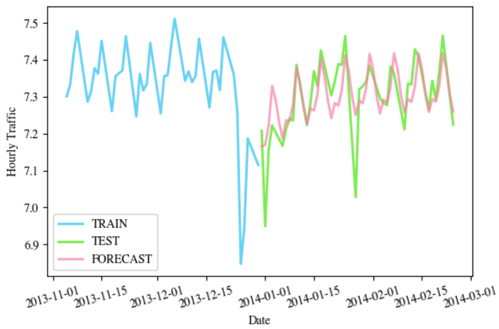

Figure 9. ARIMA model forecast at location #3 results

However, the ARIMA model modeling data at location #1 showed no seasonal pattern (Figure 10). Although the reasons for this discrepancy are unknown, one possible explanation is that the data is too sporadic, which prevents the model from identifying the 7-day periods and thus predicts the mean instead. This inconsistency suggests that the MAPE and sMAPE are better for the AR model than the ARIMA model at location #1 because the ARIMA model could not make a good prediction.

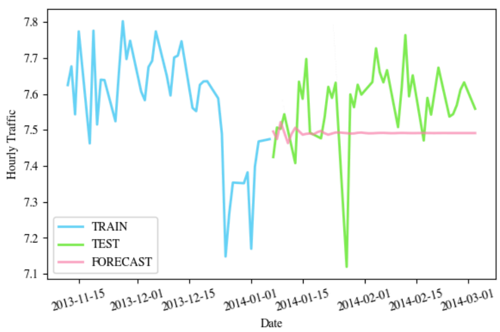

Figure 10. ARIMA model forecast at location #1 results

DOI: 10.54254/2753-8818/2026.CH31530

4.3. Results with the SARIMAX model

The optimal number of Fourier terms was 4 for the 24-hour season and 7 for the 168-hour season. Table 5 shows the R^2, RMSE, MAPE, and sMAPE values for all locations (#1-7) for the SARIMAX model and its mean model. Recall that the mean model for the SARIMAX data is different from the AR and ARIMA models because the SARIMAX uses hourly data instead of daily data. The average R^2, RMSE, MAPE, and sMAPE over the locations are 0.92, 306.5, 14.4%, and 14.0%, respectively. The SARIMAX model tended to make overestimates because the sMAPE was usually slightly less than the MAPE.

Table 5. SARIMAX and mean model statistical accuracy test results for each location (#1-7)

	SARIMAX Model				Mean Model			
Location	R^2	RMSE	MAPE	sMAPE	R^2	RMSE	MAPE	sMAPE
#1	0.94	265.3	12.0%	12.9%	-0.01	1095.3	73.3%	47.8%
#2	0.92	343.4	14.5%	13.6%	-0.15	1345.5	265.1%	62.6%
#3	0.90	277.2	11.1%	10.7%	-0.01	880.0	74.5%	47.9%
#4	0.92	438.0	13.4%	12.5%	-0.06	1611.3	140.9%	54.2%
#5	0.90	409.3	17.3%	18.5%	-0.01	1284.4	281.6%	71.5%
#6	0.90	228.6	16.6%	15.1%	-0.21	775.7	229.8%	58.5%
#7	0.93	183.9	15.8%	15.0%	-0.26	761.4	465.8%	69.0%

Although the accuracy statistics, except for the R^2, are generally worse for the SARIMAX model than for both the AR and ARIMA models, these models should be compared cautiously because the data used in the SARIMAX model is more intricate. When comparing the SARIMAX model and its mean model, the SARIMAX model significantly improves every statistic at every location, making it more meaningful for predictions than the AR and ARIMA models. Therefore, the SARIMAX can make relatively accurate predictions at all locations, as the MAPE and sMAPE values are less than 20%, and the model strongly predicts data variation, with R^2 values greater than or equal to 0.90.

Figure 11 shows some training and testing data, as well as the forecast of the model. Clearly, this model's forecast is more accurate than the other models because it better aligns with the test data. However, do note that the scales of the graphs are not the same. Additionally, there is an apparent 24-hour seasonality shown in both the training and test data.

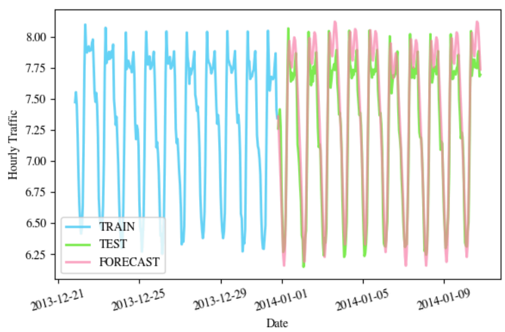

Figure 11. SARIMAX model forecast at location #3 results

DOI: 10.54254/2753-8818/2026.CH31530

5. Conclusions

5.1. Evaluations for each model

The horrible results are expected of the AR model because it is so simple. The poor R^2 results indicate that the forecast does not adequately model the data variations, meaning the model does not capture the short-term peaks and troughs. These results suggest that the model lacks the complexity to capture the dynamic, often seasonal, features in the traffic data. Because the model only uses past traffic volume values to predict future values, it can only forecast short-term autocorrelation. As a result, it produces smooth forecasts that cannot handle rapid changes in data, which are often prevalent in traffic data. The purpose of the study was to predict long-term traffic, but the AR model does not perform well in this aspect, as it tends to revert to the mean over just a few weeks. This model served as the baseline, emphasizing the need for more sophisticated approaches to predicting traffic.

The moving average (MA) component of ARIMA helps address the AR model's problem of smoothing out fluctuations. The MA component models the influence of past forecast residuals (errors), smoothening out random noise to learn usable patterns. In general, ARIMA can adapt to the short-term deviations and forecast predictions that better match the peaks and troughs of the data. Therefore, the MA portion of ARIMA improves the model's ability to capture short-term fluctuations, leading to generally more accurate performance statistic results.

The SARIMAX model extends ARIMA by incorporating both seasonal components and exogenous variables, which allows it to capture the recurring daily and weekly patterns better. The inclusion of Fourier terms as exogenous regressors provides a method to model two seasonal periods at once, resulting in forecasts that align much more closely with the observed data. Compared to ARIMA, SARIMAX is better at determining seasonal periodicities and maintaining long-term accuracy. Overall, SARIMAX made the most purposeful predictions, making it the most reliable model for predicting traffic.

5.2. Study limitations and future work

The most significant limitation in this study was the memory (RAM), CPU, and other system resources. Many complex AR models and other statistical models, such as the LSTM model used by Katambire et al., could not be run because of these restrictions. The dataset itself was a restriction, too. Because there were so few variables other than the times and their respective hourly traffic, the only reasonable models to use in this study were autoregressive models. Machine learning models, such as linear regression, could not produce results as accurate as those in the research conducted by Pun et al.

Another problem with the limited data was the problem of only being able to select specific locations to study. Many of the sites that were not chosen were unsuitable for this study because they exhibited long periods of sudden increases or decreases in traffic volumes, which autoregressive models cannot handle well. Although the models worked relatively well with the chosen locations, they are likely to have produced insignificant results with data from other sites. Therefore, the results of this study can really only be generalized to the chosen locations, not the whole Adelaide Central Market neighborhood.

Recall the plot of the ARIMA model at location #1 (Figure 10). This location is a special case where ARIMA performed noticeably worse than AR because ARIMA failed to capture the 7-day seasonal pattern. As future work, it would be beneficial to directly use a SARIMA model instead of trying to get the ARIMA model to recognize the seasonal pattern. Implementing a SARIMA model using the same daily data would also improve the predictions, as the current 7-day periods essentially repeat the exact seven predictions; SARIMA would incorporate day-to-day variations, too.

Another major problem was overfitting the data. An abundance of data was removed during the data processing, and the artificial data that filled those gaps may have been overfit to the model, especially with the Holt-Winters function for the SARIMAX model. It is not easy to balance an accurate model with preserving as much of the original data as possible. In the future, it may be beneficial to look at how the models perform with data with the weekends included or with weaker z-score and IQR parameters, e.g., testing at a 99% confidence level instead of at 95%.

It is interesting how the model tended to overestimate the data at most locations. Although this trend may be due to including a smearing estimator in the MAPE and sMAPE calculations, implementing a hybrid model that consists of a Least Absolute Shrinkage and Selection Operator (LASSO) regression component and AR models may improve the resultant performance statistics.

Another factor that could improve the model is looking at other variables besides time series data. Factors such as bad weather and accidents have significant impacts on traffic volumes. Adding these major factors as additional exogenous regressors can help in explaining the unpredictable spikes and drops in traffic.

5.3. Practical applications

Accurate short- and long-term traffic volume predictions enable urban planners to anticipate congestion before it occurs, enabling data-driven infrastructure investment and signal optimization strategies. Planners can also use these predictions to better design roads, such as implementing high-occupancy vehicle lanes or bus-priority corridors and improving public transport timing. By reducing congestion, automobile emissions will have a less significant impact on the environment, workers will become more productive because they spend less time on the road, and there will be fewer injuries and deaths due to a reduction in accidents.

5.4. Conclusion

This study showed that a SARIMAX model can effectively forecast traffic volumes when both daily and weekly patterns are captured with Fourier terms. While the AR approach is relatively simple, it still provides a valuable baseline for understanding traffic behavior over time and offers a foundation for more advanced forecasting methods. Overall, this study showed that both hourly and daily traffic data can be predicted with relative accuracy using autoregressive models. The results highlight how time series modeling can help city planners and engineers make smarter, data-driven decisions to reduce congestion and improve urban mobility.

References

[1] B, Loo, “Transport, urban, ” in International Encyclopedia of Human Geography, R. Kitchin and N. Thrift, Eds., Oxford: Elsevier, 2009, pp. 465–469, ISBM: 978-0-08-044910-4. DOI: https: //doi.org/10.1016/ B978-008044910-4.01039-7. [Online]. Available: https: //www. Sciencedirect.com/science/article/pii/.

[2] Federal Highway Administration, C. Systematics, and T. T. Institute, "Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation, " Federal Highway Administration, Tech. Rep. FHWA-HOP-05-064, Sep. 2005. [Online]. Available: https: //ops.fhwa.dot.gov/congestion_report/.

[3] K. Zhang and S. Batterman, "Air pollution and health risks due to vehicle traffic, " Science of the Total Environment, pp. 307-316, 2013. DOI: 10.1016/j.scitotenv.2013.01.074. [Online]. Available: https:

//www.ncbi.nlm.nih.gov/pmc/articles/PMC4243514/.

[4] D. T. Hartgen and M. G. Fields, "Gridlock and growth: The effect of traffic congestion on regional economic performance, " Reason Foundation, Los Angeles, CA, Tech. Rep. Policy Study No. 371, Aug. 2009. [Online]. Available: https: //reason.org/wp-content/uploads/ files/ps371_growth_gridlock_cities_ full_study.pdf.

[5] X. Wang and T. Liu, "The roads one must walk down: Commute and depression for beijing's residents, " Transportation Research Part D: Transport and Environment, vol. 109, p. 103316, 2022. DOI: 10.1016/j.trd.2022.103316. [Online]. Available: https: //www.sciencedirect.com/ science/article/pii/ S1361920922001444.

[6] L. Pun, P. Zhao, and X. Liu, "A multiple regression approach for traffic flow estimation, " IEEE Access, vol. 7, pp. 35998-36 009, 2019. DOI: 10.1109/ACCESS.2019.2904645. [Online]. Available: https: //ieeexplore.ieee.org/document/8666125.

[7] M. Akhtar, S. Moridpour, and M. Nazem, "Traffic forecasting in a freeway corridor using seasonal arima model, " in Australasian Transport Research Forum 2021 Proceedings, Brisbane, Australia, Dec. 2021. [Online]. Available: https: //australasiantransportresearchforum.org.au/wp-content/uploads/2022/05/ ATRF2021_Resubmission_117-1.pdf.

[8] V. N. Katambire, R. Musabe, A. Uwitonze, and D. Mukanyiligira, "Forecasting the traffic flow by using arima and 1stm models: Case of muhima junction, " Surveillance, vol. 5, no. 4, 2023. DOI: 10.3390/2571-9394/ 5(4)/34. [Online]. Available: https: //www.mdpi.com/2571-9394/5/4/34.

[9] O. Bazrafshan, H. Zamani, B. Farokhzadeh, and T. Caloiero, "A spectral analysis-driven SARIMAX framework with fourier terms for monthly dust concentration forecasting, " Earth, vol. 6, no. 4, p. 123, 2025. DOI: 10.3390/earth6040123. [Online]. Available: https: //www. mdpi.com/2673-4834/6/4/123.

[10] R. J. Nau, The logarithm transformation, Web page, Decision 411 - Forecasting course, Duke University, 2007. [Online]. Available: https: //people.duke.edu/~rnau/Decision411_2007/ 411log.htm.

[11] City of Adelaide, Traffic intersection volumes (scats) - hourly counts at traffic-signal intersections, adelaide city council jurisdiction, Dataset, Data.SA - Government of South Australia, 2022. [Online]. Available: https: //data.sa.gov.au/data/dataset/traffic-intersection-count.

[12] R. Pereira et al., "Imputation of data missing not at random: Artificial intelligence and machine learning approaches, " Journal of Computational Science, vol. 66, pp. 101-110, 2024. DOI: 10.1016/j.jocs.2024. 101110. [Online]. Available: https: //www.sciencedirect.com/science/ article/pii/ S0957417424005207.

[13] R. J. Nau, Identifying the order of differencing in arima models, 2007. [Online]. Available: https: //people. duke.edu/~rnau/411arim2.htm.

[14] N. Duan, "Smearing estimate: A nonparametric retransformation method, " Journal of the American Statistical Association, vol. 78, no. 383, pp. 605-610, 1983. DOI: 10.2307/2288126. [Online]. Available: https : //www.jstor.org/stable/2288126.

[15] C. D. Lewis, Industrial and Business Forecasting Methods. London: Butterworths, 1982. [Online]. Available: https: //www.researchgate.net/figure/Criteria-for-accuracy-level-of-MAPE-Lewis1982_tbl2_ 271562818.