# Urban Freeway Traffic Flow Prediction

## Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models

BILLY M. WILLIAMS, PRIYA K. DURVASULA, AND DONALD E. BROWN

The application of seasonal time series models to the single-interval traffic flow forecasting problem for urban freeways is addressed. Seasonal time series approaches have not been used in previous forecasting research. However, time series of traffic flow data are characterized by definite periodic cycles. Seasonal autoregressive integrated moving average (ARIMA) and Winters exponential smoothing models were developed and tested on data sets belonging to two sites: Telegraph Road and the Woodrow Wilson Bridge on the inner and outer loops of the Capital Beltway in northern Virginia. Data were 15-min flow rates and were the same as used in prior forecasting research by B. Smith. Direct comparisons with the Smith report findings were made and it was found that ARIMA $(2, 0, 1)(0, 1, 1)_{96}$ and ARIMA $(1, 0, 1)(0, 1, 1)_{96}$ were the best-fit models for the Telegraph Road and Wilson Bridge sites, respectively. Best-fit Winters exponential smoothing models were also developed for each site. The single-step forecasting results indicate that seasonal ARIMA models outperform the nearest-neighbor, neural network, and historical average models as reported by Smith.

Advances in traffic management and control are coming out of the applied transportation research community at an accelerating pace in response to worsening urban traffic congestion. Many of the traffic control, decision support, and traveler information systems being developed will either require or, at a minimum, be greatly enhanced by reliably accurate traffic flow prediction models. Recent traffic flow prediction efforts have focused on application of neural networks, nonparametric regression (including nearest-neighbor algorithms), and time series analysis techniques. The literature includes some comparative studies involving these approaches. [Smith, (*1*)] Although autoregressive integrated moving average (ARIMA) models have been presented favorably in some papers, we found no references in the literature on the use of seasonal ARIMA models for traffic flow forecasting. The previous ARIMA models have been predominantly $(p, 1, q)$ models—that is, fitting the autoregressive moving average ARMA $(p, q)$ model to the first difference of a traffic flow series. However, given the strongly cyclical nature of traffic flow, the first difference of a traffic flow time series will not yield a stationary series. One of the necessary conditions for time series stationarity is that the expected value of future realizations of the series must be constant for all time intervals. The first difference of many traffic flow series does not satisfy this condition. Specifically, if $\{X_t\}$ is a traffic flow time series and $\{Y_t\} = \{\nabla X_t\}$, then $E[Y_t] > 0$ for all $t$ during times of the day when traffic volumes are expected to be rising to a peak, $E[Y_t] < 0$ for all $t$ during times of the day when traffic volumes are expected to be falling from a peak, and $E[Y_t] \approx 0$ for all $t$ during times of the day when traffic volumes are expected to be in an on-peak or off-peak plateau.

B. M. Williams and P. K. Durvasula, Department of Civil Engineering, University of Virginia, Charlottesville, VA 22903-2442. D. E. Brown, Department of Systems Engineering, University of Virginia, Charlottesville, VA 22903-2442.

## SEASONAL TIME SERIES MODELS

Seasonal ARIMA and exponential smoothing models are briefly described. For a complete discussion the reader is referred to a comprehensive time series analysis text such as that of Brockwell and Davis (*2*).

### Multiplicative Seasonal ARIMA

Simply stated, ARIMA models are linear estimators regressed on past values of the modeled time series (the autoregressive terms) or past prediction errors (the moving average terms). If necessary, the raw time series is made stationary by differencing the series temporally or seasonally.

A time series, $\{X_t\}$, is a seasonal ARIMA $(p, d, q)(P, D, Q)_s$ process with period $s$ if $d$ and $D$ are nonnegative integers and if the differenced series $Y_t = (1 - B)^d(1 - B^s)^D X_t$ is a stationary autoregressive moving average (ARMA) process defined by the expression

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t \tag{1}$$

where $B$ is the backshift operator defined by $B^aW_t = W_{t-a}$ $\phi(z) = 1 - \phi_1 z - \Lambda - \phi_p z^p$, $\Phi(z) = 1 - \Phi_1 z - \Lambda - \Phi_P z^P$, $\theta(z) = 1 - \theta_1 z - \Lambda - \theta_q z^q$, $\Theta(z) = 1 - \Theta_1 z - \Lambda - \Theta_Q z^Q$, $p$ and $P$ are the autoregressive order and seasonal autoregressive order, respectively, $q$ and $Q$ are the moving average order and seasonal moving average order, respectively, and $Z_t$, is a white noise (WN) process identically and normally distributed with mean zero, variance $\sigma^2$, and Cov $(Z_t, Z_{t-k}) = 0 \;\forall\; k \neq 0$, i.e., WN$(0, \sigma^2)$.

Autoregressive processes with no moving average parameters are referred to in shorthand as AR (p) processes, and moving average processes with no autoregressive parameters are referred to as MA (q).

In ARIMA model development, the sample autocorrelation and partial autocorrelation functions (ACF and PACF) play an important role in model identification. For a definition of these functions, the reader is again referred to a comprehensive time series analysis text.

### Seasonal Exponential Smoothing

Exponential smoothing models are linear estimators that place greater weight on recent values. The weight on past values decays exponentially. The rate of this decay is set by means of a smoothing parameter.

Seasonal exponential smoothing models are alternatively known as Holt-Winters or Winters exponential smoothing. The multiplicative form of the Winters model is given by the following equations

$$X_t = M_t S_{t-d} + Z_t \tag{2}$$

$$M_t = L_{t-1} + B_{t-1} \tag{3}$$

$$L_t = \alpha\left(\frac{X_t}{S_{t-d}}\right) + (1-\alpha)(L_{t-1} + B_{t-1}) \tag{4}$$

$$B_t = \gamma(L_t + L_{t-1}) + (1-\gamma)B_{t-1} \tag{5}$$

$$S_t = \delta\left(\frac{X_t}{L_t}\right) + (1-\delta)S_{t-d} \tag{6}$$

where

$\{X_t\}$ = a time series with seasonality and trend,
$L_t$ = the series level (deseasonalized) at time $t$,
$B_t$ = slope at time $t$,
$S_t$ = multiplicative seasonal factor at time $t$,
$Z_t$ = WN(0, $s^2$), and
$\alpha$, $\gamma$, and $\delta$ = smoothing parameters $\epsilon$ (0, 1).

Seasonal factors for the first period and slope at the first interval are estimated from a training sample of $\{X_t\}$. These estimates, along with an actual or estimated initial series value, are then applied to the equations above to recursively calculate one-step forecasts. In model training, the smoothing parameter values are set to minimize the mean squared error of the one-step forecasts. Single exponential smoothing is a simplification of the above equations assuming no seasonality and a constant level (no slope). Double exponential smoothing includes slope but no seasonality.

## RELATED RESEARCH

Ahmed and Cooke (*3*) compared the short-term traffic volume forecasts of the ARIMA (0, 1, 3) model to those obtained by double exponential smoothing, simple moving average (with orders of 5, 10, and 20), and exponential smoothing with adaptive response (the value of $\alpha$ is changed for each one-step forecast). The four models were developed and evaluated on 166 data sets obtained from three surveillance systems: Los Angeles (20-sec intervals), Minneapolis (30-sec intervals), and Detroit (60-sec intervals). The authors found that the ARIMA (0, 1, 3) forecasts were the most accurate in terms of the mean absolute error and mean square error.

Levin and Tsao (*4*) compared the performance of the ARIMA (0, 1, 1) model with the Illinois Department of Transportation Traffic Systems Center's ARIMA (0, 1, 0) model [actually an ARIMA (3,1,0) with $\phi_1 = \phi_2 = \phi_3 = 1/3$]. The two models were tested on traffic data from the Dan Ryan Expressway in Chicago. The ARIMA (0, 1, 1) model was superior.

A comparison by Smith (*1*) of neural network, nearest-neighbor, historical average, and ARIMA models showed that an ARIMA (2, 1, 0) model did better than the historical average model and worse than the nearest-neighbor and neural network models. However, the ARIMA model was applied to only one data set and tested for only 2 days of the data set because of embedded missing values.

More recently, a hybrid approach combining ARIMA modeling with Kohonen maps gave encouraging results (*5*). In this study, an ARIMA ($p$, 0, $q$) with $p = 2$ or 3 and $q = 1$ or 2 was applied to clustered outputs from a 15-row by 20-node hexagonal Kohonen map.

## STUDY OVERVIEW

This study addresses the problem of forecasting single-interval traffic volumes at time $t + D$ given the volumes at time $t$, where $D$ is the series interval. We used a value of 15 min for $D$ in keeping with the data sets from Smith (*1*). Traffic flow series intervals in the literature range from 20 sec to 1 h. Fifteen-minute data intervals are appropriate and widely used for highway capacity analysis and traffic control.

Our primary objective was to fit seasonal ARIMA and Winters exponential smoothing models and evaluate their single-interval prediction performance relative to the models developed by Smith (*1*). We used the statistical analysis package SPSS version 7.0 with the Trends module to conduct all the analyses (*6*).

The data are equivalent hourly flow rates converted from 15-min volume counts. Data were available for two sites over the same time period. The outputs were single-interval forecasts generated by the fitted models. The data for each site include a training or development data set and a test data set.

A primary assumption for this study was that accurate forecast models can be developed with current and past flow rates. Statistical analysis by Smith showed that the independent variable, current volume, had the highest correlation with the dependent variable, future volume (*1*). The other predictive variables considered by Smith were speed, occupancy (percentage of time the sensors are detecting the presence of a vehicle—a surrogate for traffic density), and temperature. All these variables showed much less correlation with the future volume than did current volume.

## DATA ACQUISITION

We obtained the raw traffic flow data from Smith (*1*). The data were gathered by the Virginia Department of Transportation's Northern Virginia traffic management system (TMS). The computer system in the TMS facility receives remotely sensed traffic data via hardwire communication from in-pavement loop detectors.

### Study Locations

The two sites studied are approximately 3.22 km (2 mi) apart on the I-95 Capital Beltway south of Washington, DC, as indicated in Figure 1. The sites are referred to as the Telegraph Road site and the Wilson Bridge site. The loop detectors for the Telegraph Road site are located in the inner loop (southbound) lanes under the overpass for the Telegraph Road interchange. The inner loop has four full-width travel lanes at this location.

The loop detectors for the Wilson Bridge site are located in the outer loop (northbound) lanes just east of the Woodrow Wilson Bridge. The outer loop has three full-width travel lanes at this location.

Both locations are subject to extreme traffic congestion. Of the two sites, the Wilson Bridge site experiences greater congestion because it has higher flow rates and three travel lanes instead of four.

### Data Description

As mentioned above, the data are hourly flow rates converted from 15-min volume counts. For each site, data were available for 44 consecutive days in June and July 1993 and for 32 consecutive days in September and October 1993. We used the June/July data for fore-

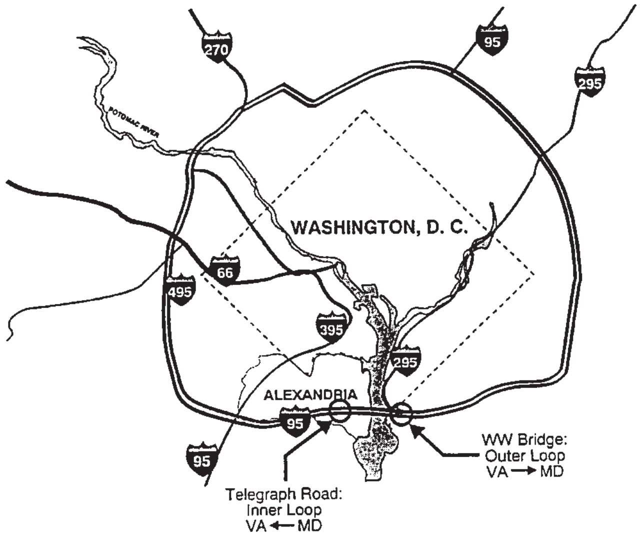


**FIGURE 1 Data locations.**

cast model development and the September/October data for model testing (*1*).

### Missing Value and Outlier Replacement

Approximately 20 percent of the data in the development and test sets were missing. Replacement of the missing values in the test set was necessary because SPSS requires a complete time series for exponential smoothing analysis. There is no consensus in the literature on preferred missing value replacement strategies for traffic flow time series [see Clark et al. (*7*) for discussion]. Therefore, we developed two reasonable estimation methods, one for each development data set. Our intent in using two estimation methods was to compare their impact on subsequent model performance.

However, SPSS allows embedded missing values when fitting ARIMA models. When missing values are present, SPSS uses Kalman filtering to replace the missing values before proceeding with parameter estimation. Kalman filtering theory is based on a state space representation of stochastic processes (*8*). One could in fact represent ARIMA processes in the form of Kalman's recursions.

Therefore, we developed ARIMA models for both sites with the missing values embedded and with the missing values replaced. This allowed us to compare the performance of the alternative missing value replacement methods with Kalman filtering as implemented by SPSS. For the Telegraph Road development set, we developed a missing values estimation method that incorporates both the most recent series value and the historical information for the corresponding time period on the same day of the week. We accomplished this by averaging two preliminary estimates. The first estimate was the previous series value plus a weighted average of the corresponding one-step differences from the previous day and the previous same day of the week. Greater weight was assigned to the corresponding one-step difference for the previous same day of the week on the assumption that it would hold better information on the slope of the series at the missing time period. The second preliminary estimate was based on an average of the two previous same days of the week.

For the Wilson Bridge site development data set, we replaced the missing values by taking an average of the available values for the corresponding 15-min period for the same day of the week. Unlike the Telegraph Road replacement strategy, this method did not include recent series values in the estimation. As mentioned above, our intent was to see how the two methods compared in terms of the forecasting performance of the resulting models.

For a direct comparison with the study of Smith (*1*) we did not replace all the missing values in the test sets. However, to generate a full set of ARIMA forecasts, we had to replace the embedded missing values in the first period of the series—that is, the first 96 values. If there are embedded missing values within the first period of a seasonal time series, SPSS is unable to forecast a value for the corresponding time interval throughout the ARIMA forecast series. Both sites had 32 missing values at the beginning of the test data series. These values had to be "seeded" to provide a full complement of hourly flow rates for the first day.

We calculated the exponential smoothing forecasts with Excel spreadsheets. For these forecasts only the initial value had to be seeded. However, the forecast process used the initial seasonal factors to replace the remaining missing values in the first day. Therefore, the seeding process was essentially the same for both ARIMA and exponential smoothing forecasts.

The forecasting results presented for the ARIMA and exponential smoothing models do not include these estimated values. Our analysis considers errors only where a forecast is paired with an actual data point.

### Outliers

To identify outliers in the development data, we applied a seasonal decomposition to both development sets with a daily periodicity (96 intervals). A more detailed description of how we determined the periodicity of the time series is presented in the data analysis section. We identified potential outliers both by visual inspection of the error plots for the seasonally adjusted series and by identifying error values that were greater than 3 standard deviations from the mean error value. Given the number of values in the development sets, the 3 standard deviation limit defines the range outside of which no values would be expected. Our intent was to replace only true aberrations in the data. After flagging points of extreme error, we checked the raw data to see if we should consider the raw values to be outliers.

We found seven Wilson Bridge development set data points to be outliers. These values were orders of magnitude different from the preceding and succeeding values. For example, one such point had a value of 462, whereas the preceding and succeeding values were 3,677 and 5,290, respectively. Such errors are not uncommon in loop detector data. Supporting data were not available to make a specific determination of the cause of the outliers in the development data sets. We replaced the seven Wilson Bridge development set outliers by using the same estimation technique used for missing values. We also identified and replaced 13 Telegraph Road development set outlying data points.

It is possible that some of the outliers replaced were the actual record of incidents. However, because relatively few values were involved, outlier replacement had only a slight effect on model estimation. We replaced 0.4 and 0.2 percent of the total available values (not including missing values) for the Telegraph Road and Wilson Bridge sites, respectively. To evaluate the replacement effect, we also developed and tested models by using the raw data with no outlier replacement. Although the difference in overall performance was small, these models did not do as well in forecasting the test set values as did the models developed with the outliers replaced.

## DATA ANALYSIS

Before fitting an ARIMA to the time series data, we determined the differencing needed to achieve stationarity. Next, we determined the approximate order of the ARIMA ( *p, d, q*)(*P, D, Q*)*s* models. Finally, we developed ARIMA and seasonal exponential smoothing models and selected the best models for testing.

### Determination of Required Differencing

First, we plotted the development data sets against time. Figure 2 presents a representative week for each site. As expected, these plots indicated a periodic or seasonal component, with a daily cycle (96 intervals) for both sites. Second, we plotted the ACFs of the raw data up to a lag of 768. A lag of 768 corresponds to 8 days. A plot out to an 8-day lag allowed us to check for a weekly cycle. These plots, also presented in Figure 2, show significant correlation with slow decay and a well-defined periodicity of 96 time intervals (1 day). Therefore, the traffic flow process is nonstationary with a strong daily seasonal pattern.

The preceding analysis indicated that a first seasonal difference at a lag of 96 (daily period) was needed to transform the development series into stationary processes. Autocorrelation plots after a first difference and a first seasonal difference confirmed that a first seasonal difference yields a stationary process whereas a first difference still exhibits slowly decaying periodic correlation. Figure 3 presents these ACF plots for the differenced series.

### Preliminary ARIMA Order Estimation

We analyzed plots of the ACF and PACF of the seasonally differenced series up to a lag of 16 for both development sets (Figure 4). Lag 16 is the default SPSS plot option and was sufficient to identify the nonseasonal model order. The ACFs and PACFs strongly suggested a nonseasonal autoregressive (AR) process of order 3 or 4. There was no evidence of a moving average (MA) process in the plots.

Next, we plotted the ACF and the PACF up to a lag of 768, showing the correlations only at daily lags (Figure 5). The plots of the ACF showed a significant spike at the first periodic lag. On the other hand, the plots of the PACF showed a smoother decay. Taken together, the ACF and PACF plots suggested a first order seasonal MA component [MA(1)]. The evidence of a seasonal MA (1) was stronger for the Wilson Bridge data than for the Telegraph Road data. The Telegraph Road ACF and PACF plots also suggested a possible seasonal ARMA (1, 1). In summary, the ACF and the PACF plots suggested a seasonal ARIMA $(4, 0, 0)(0, 1, 1)_{96}$ for the Telegraph Road site and an ARIMA $(3, 0, 0)(0, 1, 1)_{96}$ for the Wilson Bridge site. For completeness and model verification, we fitted several other models to the development data sets for the two sites, including models with mixed seasonal and nonseasonal components and models with higher-order seasonal moving average components.

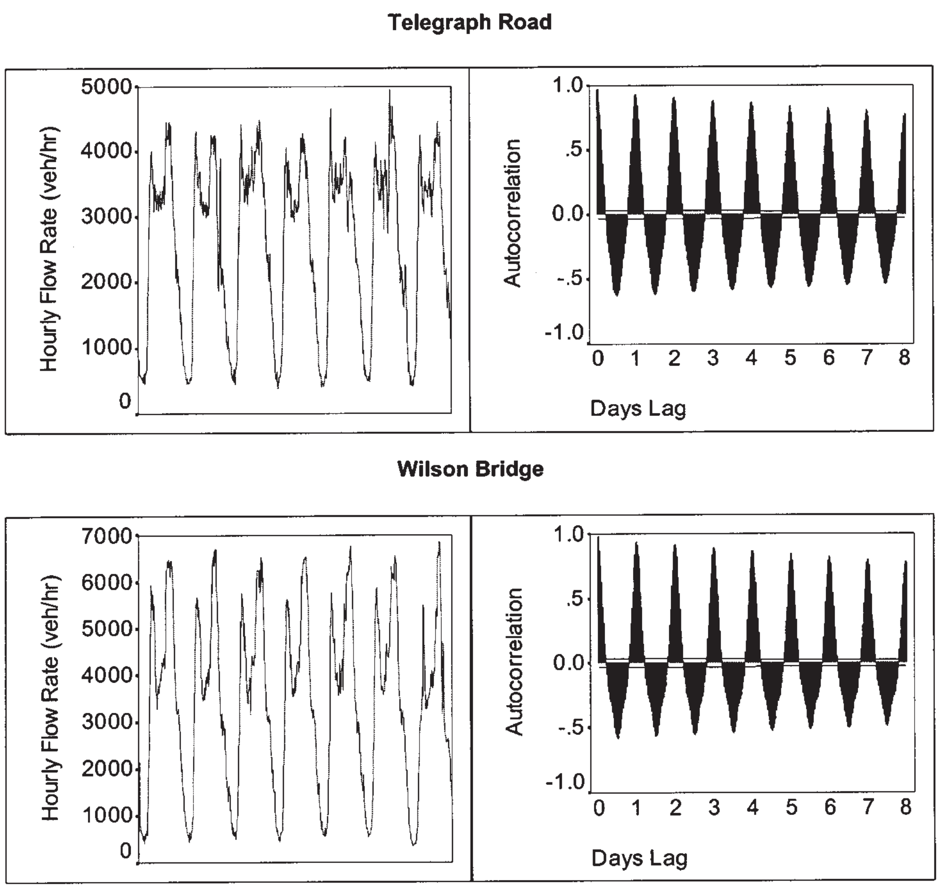


**FIGURE 2 Raw series and autocorrelation plots.**

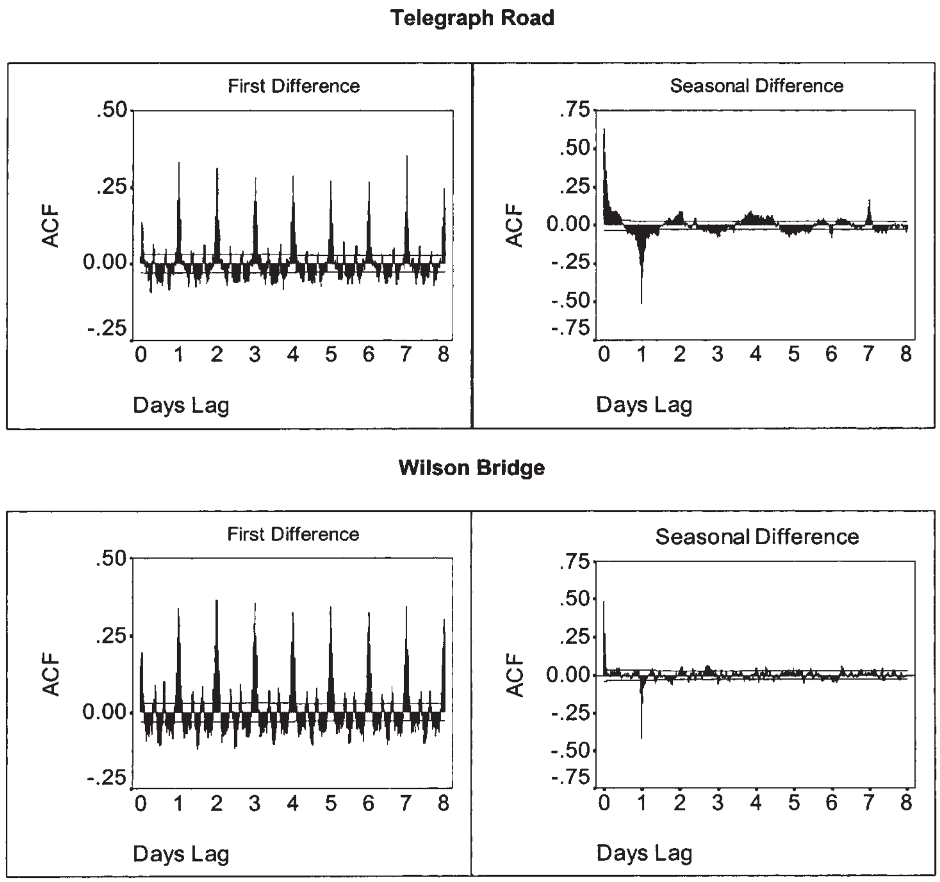


**FIGURE 3 Autocorrelation plots for the differenced series.**

### Exponential Smoothing Model Development

As discussed above, three parameters are estimated when fitting a Winters model: $\alpha$, the level parameter; $\gamma$, the trend parameter; and $\delta$, the seasonal parameter. SPSS performs a grid search to aid in estimation of the Winters smoothing parameters. The user sets ranges and search steps for each parameter. After searching the grid defined by the ranges and steps, SPSS displays the 10 best parameter sets in terms of sum of squared error. We used this SPSS grid search process to estimate the smoothing parameters for the development sets. We present the results in more detail in the next section.

## FORECAST MODELS

### ARIMA Models

Search space definition for ARIMA model development was guided by the preliminary order estimation described in the section on preliminary ARIMA estimation. We tested and verified the preliminary assessment that a pure nonseasonal moving average model was not appropriate. Selection criteria for pure moving average models do not minimize until an order of $q \approx 10$ and the best fits were measurably worse than the pure autoregressive and mixed nonseasonal models. Preliminary model fitting clearly verified the appropriateness of a seasonal MA (1) model component. Seasonal AR parameters degraded model fit in all cases. Also, including MA parameters beyond the first order made a significant improvement to model fit in none of the cases tried, and the estimated higher-order parameters were statistically insignificant in all cases. SPSS calculates the standard error, log likelihood, Akaike information criterion (AIC) statistic, and Swartz Bayesian criterion (SBC) statistic for each estimated ARIMA model. Standard error is given for information but is not specifically a selection criterion because it measures only fit of the model to the development set. The AIC and SBC statistics, on the other hand, are measures designed to predict model fit on future realizations of the modeled process. We emphasized the SBC statistic because studies have indicated that AIC tends to overestimate the AR order (*2*). Also, we considered SBC values close to the minimum to be equivalent, and in these cases we favored a simpler model. It is suggested in the literature that models with SBC or AIC statistics within some constant $c$ ($c = 2$ as a typical value) of the minimum SBC or AIC should be considered competitive (*2*). Based on evaluation of the SBC statistic, we selected the ARIMA models indicated in Table 1 for model testing.

### Exponential Smoothing Models

As mentioned above, we used SPSS's grid search capability to estimate the exponential smoothing parameters. When all parameters were allowed full range from 0 to 1, the 10 models with the lowest sum of square errors all had $\gamma$ and $\delta = 0$ for both training sets. This is not surprising because trend and seasonal factor estimates should not vary significantly over the development set. When $\gamma$ and $\delta = 0$,

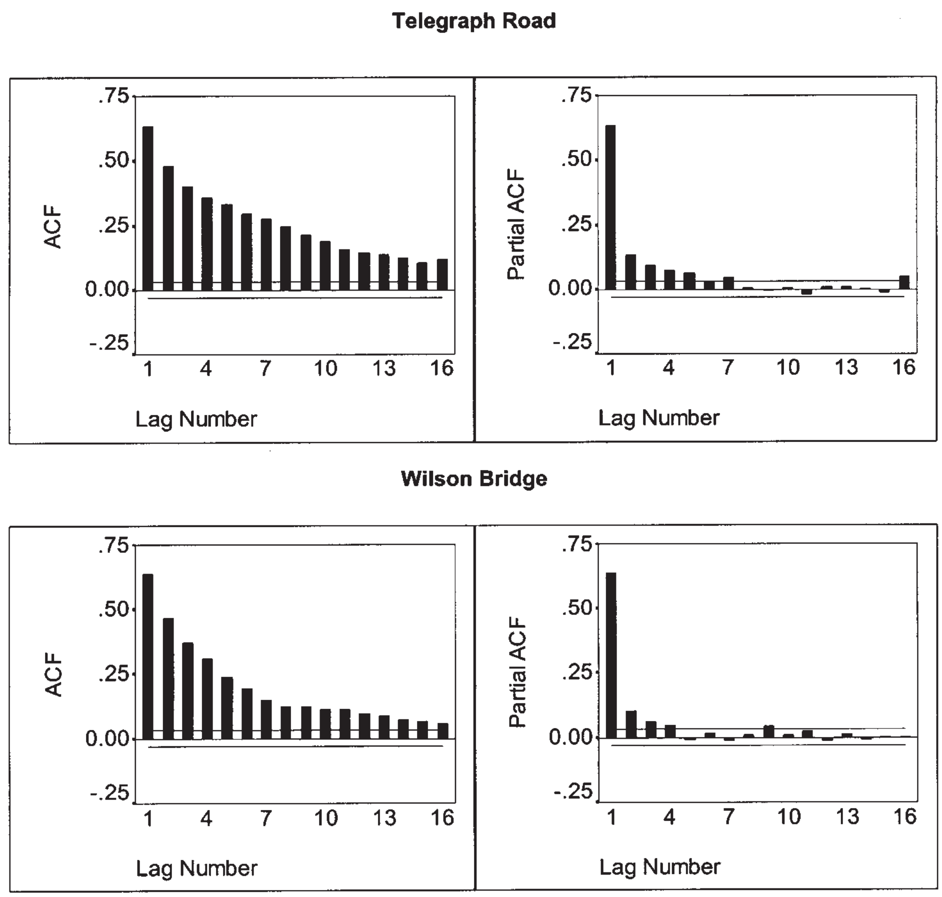


**FIGURE 4 ACF and PACF plots: first 16 lags.**

the model is actually a simple exponential smoothing model on the seasonally decomposed series with a constant trend. We initially hypothesized that a model with a value of 0.01 for $\gamma$ and $\delta$ would provide a more robust model. The fact that both of these smoothing parameters were optimized at zero for the development data suggests that the seasonal factors and trend change slowly. Introducing a minimal smoothing parameter value empowers the models to adapt to long-term changes. The value 0.01 is a generally accepted lower bound for smoothing parameters.

However, we found that the trend term for a 15-min interval traffic flow series is so small (estimated at less than 0.05 vehicle per hour for both sites) compared with expected random fluctuations in the level that even a minimal value for $\gamma$ made the model unstable. We held to our hypothesis that $\delta = 0.01$ would yield a better model and therefore also estimated values for $\alpha$ when $\delta = 0.01$.

The estimated parameters are indicated in Table 2. We applied these parameters to the test sets along with seasonal factors estimated from the development data.

### Analysis of Residuals

We analyzed the residuals for the best models. The ACF plots for the ARIMA and exponential smoothing models showed the residuals to be sufficiently uncorrelated. The ARIMA models also satisfied the Box-Ljung Portmanteau test for independence (*9*). However, the exponential smoothing errors did exhibit statistically significant correlation at the first few lags. This is expected because exponential smoothing models are rigidly restricted to exponentially decaying regression on past values. Properly applied ARIMA models, on the other hand, have complete flexibility in estimating the regression parameters to extract the correlation structure from the modeled series.

## FORECASTING RESULTS

Tables 3 and 4 summarize the model testing results. The minimum values in the tables for each error and bad miss measure are highlighted in boldface type. The presentation format of the test measures follows Smith (*1*) to allow direct comparison with that study's results. The following applies to the Telegraph Road models:

- ARIMA-1: ARIMA (4, 0, 0)(0, 1, 1)$_{96}$ development set with missing values;
- ARIMA-2: ARIMA (2, 0, 1)(0, 1, 1)$_{96}$ development set with missing values;
- ARIMA-3: ARIMA (4, 0, 0)(0, 1, 1)$_{96}$ development set with missing values replaced;
- ARIMA-4: ARIMA (2, 0, 1)(0, 1, 1)$_{96}$ development set with missing values replaced;
- EXSMO-1: Winters exponential smoothing, $\alpha = 0.565$, $\gamma = 0$, $\delta = 0$;

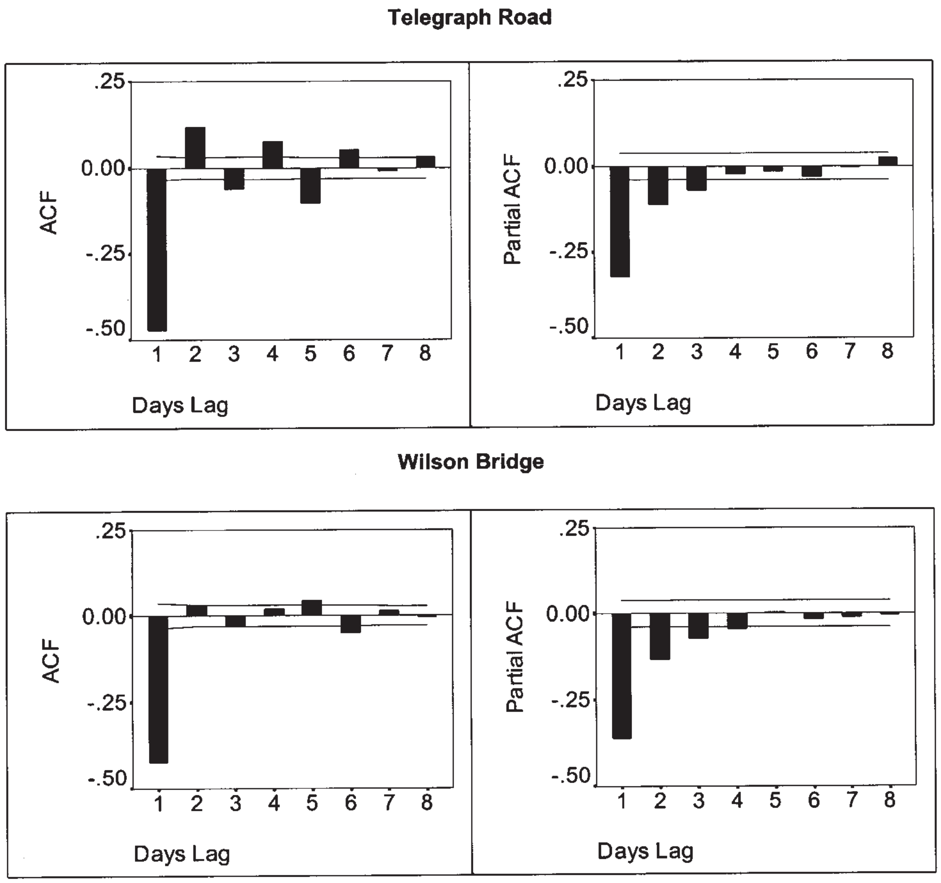


**FIGURE 5 ACF and PACF plots: eight seasonal lags.**

- EXSMO-2: Winters exponential smoothing, $\alpha = 0.58$, $\gamma = 0$, $\delta = 0.01$;
- NNb: nearest-neighbor model (*1*);
- NNw: neural network model (*1*); and
- HA: historical average model (*1*).

The following applies to the Wilson Bridge models:

- ARIMA-1: ARIMA (3, 0, 0)(0, 1, 1)$_{96}$ development set with missing values;
- ARIMA-2: ARIMA (1, 0, 1)(0, 1, 1)$_{96}$ development set with missing values;
- ARIMA-3: ARIMA (3, 0, 0)(0, 1, 1)$_{96}$ development set with missing values replaced;
- ARIMA-4: ARIMA (1, 0, 1)(0, 1, 1)$_{96}$ development set with missing values replaced;
- EXSMO-1: Winters exponential smoothing, $\alpha = 0.68$, $\gamma = 0$, $\delta = 0$
- EXSMO-2: Winters exponential smoothing, $\alpha = 0.682$, $\gamma = 0$, $\delta = 0.01$
- NNb: nearest-neighbor model (*1*);
- NNw: neural network model (*1*); and
- HA: historical average model (*1*).

The results presented in Tables 3 and 4 indicate that for each site the four ARIMA and two exponential smoothing models equal or better the performance of the models presented by Smith (*1*). Furthermore, the best-fit ARIMA models, ARIMA-2 (in terms of mean absolute error and average percent error) for Telegraph Road and ARIMA-3 (in terms of mean absolute error) and ARIMA-4 (in terms of average percentage error) for Wilson Bridge outperform the nearest-neighbor model in terms of mean absolute error, percentage absolute error, and total percentage of cases with greater than 10 and 20 percent estimation error.

Although the neural network model had the lowest percentage of cases with greater than 10 and 20 percent overestimation for the Telegraph Road site, Table 4 also indicates that the neural network model had the largest percentage of total cases with greater than 20 percent estimation error. Furthermore, the neural network model had the highest percentage of cases with greater than 10 and 20 percent estimation error for the Wilson Bridge site and finished a close second to the historical average model for highest percentage of cases with greater than 10 percent estimation error for the Telegraph Road site.

In addition to comparing the summary forecasting statistics described above, we performed related sample statistical tests between the absolute values of the one-step prediction errors for the seasonal ARIMA, Winters, and nearest-neighbor forecasts. The Freidman test was used to test the null hypothesis that the three forecast error samples are from the same population. The Wilcoxon signed-rank test was then performed on the forecast errors, taking two samples at a time. The Wilcoxon signed-rank test results are not valid for making distinctions among the three unless the Freidman test supports the alternative hypothesis—that is, that the three forecast error samples are not

**TABLE 1 ARIMA Model Parameters**

| Model | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\phi_4$ | $\theta_1$ | $\Theta_1$ | Constant |
|---|---|---|---|---|---|---|---|
| **Telegraph Road** | | | | | | | |
| With Missing Values | | | | | | | |
| $(4,0,0)(0,1,1)_{96}$ | 0.541 | 0.075 | 0.048 | 0.077 | NA | 0.895 | 5.87 |
| $(2,0,1)(0,1,1)_{96}$ | 1.206 | -0.281 | NA | NA | 0.670 | 0.888 | 5.76 |
| Without Missing Values | | | | | | | |
| $(4,0,0)(0,1,1)_{96}$ | 0.558 | 0.079 | 0.043 | 0.065 | NA | 0.909 | 5.33 |
| $(2,0,1)(0,1,1)_{96}$ | 1.169 | -0.261 | NA | NA | 0.616 | 0.906 | 5.30 |
| **Wilson Bridge** | | | | | | | |
| With Missing Values | | | | | | | |
| $(3,0,0)(0,1,1)_{96}$ | 0.606 | 0.036 | 0.035 | NA | NA | 0.901 | 10.85 |
| $(1,0,1)(0,1,1)_{96}$ | 0.702 | NA | NA | NA | 0.098 | 0.901 | 10.86 |
| Without Missing Values | | | | | | | |
| $(3,0,0)(0,1,1)_{96}$ | 0.626 | 0.063 | 0.048 | NA | NA | 0.929 | 8.74 |
| $(1,0,1)(0,1,1)_{96}$ | 0.774 | NA | NA | NA | 0.153 | 0.930 | 8.74 |

from the same population. For both sites, the Freidman test supported the alternative hypothesis at the 0.01 level.

For the Telegraph Road forecasts, these tests indicated that the seasonal ARIMA model was most preferred at the 0.01 significance level, but preference of the Winters model over the nearest-neighbor model was not statistically significant. For the Wilson Bridge forecasts, these tests indicated that the nearest-neighbor model was least preferred at the 0.01 significance level, but the preference of the seasonal ARIMA model over the Winters model was not statistically significant.

In summary, the seasonal time series models all performed well relative to the nearest-neighbor model and the best ARIMA models outperformed the nearest-neighbor model.

## CONCLUSIONS AND RECOMMENDATIONS

Our primary finding is that the seasonal time series models performed well for single-interval traffic flow forecasting. The resulting models are simple linear equations that require very few inputs. Once seasonal time series models are developed and tested, they are simple to implement and require insignificant computing time. We present additional conclusions on missing value replacement and intuitive understanding of seasonal time series models.

### Missing Values and Outliers

The ARIMA models developed with each of the replacement strategies performed well. For the Telegraph Road site, the models fitted on the development set with missing values gave lower error measures than the ones fitted on the development set with the missing

**TABLE 2 Exponential Smoothing Parameters**

| Parameter | Optimal Value after Grid Search | Values with $\delta = .01$ |
|---|---|---|
| **Telegraph Road** | | |
| $\alpha$ | 0.57 | 0.58 |
| $\gamma$ | 0.00 | 0.00 |
| $\delta$ | 0.00 | 0.01 |
| **Wilson Bridge** | | |
| $\alpha$ | 0.68 | 0.682 |
| $\gamma$ | 0.00 | 0.00 |
| $\delta$ | 0.00 | 0.01 |

**TABLE 3 Error Measures**

| Model | Mean Absolute Error (vehicles/hour) | Average %Error |
|---|---|---|
| **Telegraph Road** | | |
| ARIMA-1 | 161.17 | 7.42 |
| ARIMA-2 | **160.82** | **7.41** |
| ARIMA-3 | 161.40 | 7.43 |
| ARIMA-4 | 161.11 | 7.42 |
| EXSMO-1 | 167.54 | 7.54 |
| EXSMO-2 | 167.27 | 7.54 |
| NNb | 167.3 | 7.54 |
| NNw | 182.5 | 8.93 |
| HA | 214.6 | 9.57 |
| **Wilson Bridge** | | |
| ARIMA-1 | 221.18 | 7.79 |
| ARIMA-2 | 221.38 | 7.79 |
| ARIMA-3 | **220.74** | 7.73 |
| ARIMA-4 | 220.74 | 7.72 |
| EXSMO-1 | 221.07 | 7.49 |
| EXSMO-2 | 220.87 | **7.49** |
| NNb | 229.3 | 8.07 |
| NNw | 450.3 | 11.0 |
| HA | 300.4 | 9.86 |

**TABLE 4 Bad Miss Measures**

| Model | Percent Cases with Greater Than 10% Error | | | Percent Cases with Greater Than 20% Error | | |
|---|---|---|---|---|---|---|
| | Under-estimate | Over-estimate | Total | Under-estimate | Over-estimate | Total |
| **Telegraph Road** | | | | | | |
| ARIMA-1 | 10.46 | 13.07 | 23.53 | 1.90 | 4.05 | 5.95 |
| ARIMA-2 | 10.67 | 13.03 | 23.7 | **1.74** | 4.14 | **5.88** |
| ARIMA-3 | 10.42 | 12.82 | **23.24** | 1.99 | 4.09 | 6.08 |
| ARIMA-4 | 10.63 | 12.86 | 23.49 | 1.86 | 4.09 | 5.95 |
| EXSMO-1 | **10.38** | 13.23 | 23.61 | 2.36 | 3.93 | 6.29 |
| EXSMO-2 | 10.42 | 13.32 | 23.74 | 2.36 | 4.01 | 6.37 |
| NNb | 13.31 | 10.98 | 24.29 | 1.87 | 4.08 | 5.95 |
| NNw | 24.08 | **7.28** | 31.36 | 10.02 | **1.75** | 11.77 |
| HA | 14.02 | 19.30 | 33.32 | 3.95 | 6.95 | 10.9 |
| **Wilson Bridge** | | | | | | |
| ARIMA-1 | 8.56 | 13.41 | 21.97 | 1.68 | 4.69 | 6.37 |
| ARIMA-2 | **8.41** | 13.64 | 22.05 | 1.60 | 4.89 | 6.49 |
| ARIMA-3 | 8.91 | 12.90 | 21.81 | 1.92 | 4.22 | 6.14 |
| ARIMA-4 | 8.84 | 12.90 | 21.74 | 1.72 | 4.42 | 6.14 |
| EXSMO-1 | 8.62 | 11.98 | 20.6 | 2.10 | **3.39** | 5.49 |
| EXSMO-2 | 8.50 | **11.82** | **20.32** | 1.91 | **3.39** | **5.3** |
| NNb | 11.01 | 14.58 | 25.59 | **1.53** | 4.68 | 6.21 |
| NNw | 32.86 | 12.11 | 44.97 | 14.47 | 5.58 | 20.05 |
| HA | 13.99 | 19.85 | 33.84 | 4.6 | 6.33 | 10.93 |

values replaced. The Wilson Bridge site yielded opposite results. Because the two estimation strategies were not applied to both development sets, we can draw no strong conclusions. However, it is interesting that, although the Wilson Bridge replacement strategy had a greater influence on the estimated parameter values, this strategy also produced better forecasting results. This result appears to indicate that replacement strategies that attempt to minimize influence on the resulting parameter estimates may not always produce the best forecasting results.

There remains much to be investigated about outlier replacement and missing value estimation. Ultimately, the best criterion for judging the relative merit of replacement strategies is their effect on forecasting performance. Better understanding of the underlying structure of traffic flow time series in various situations and on various facility types will be a prerequisite for developing better replacement methods. However, Chen and Liu have developed a rigorous method of handling time series outliers in model estimation (*10*) and forecasting (*11*). Their procedures may provide a solid foundation for developing a specific model estimation and forecasting method for vehicular traffic flow data.

## Intuitive Understanding of Seasonal Time Series Models

The best-fit ARIMA model was ARIMA $(2, 0, 1)(0, 1, 1)_{96}$ for the Telegraph Road site and ARIMA $(1, 0, 1)(0, 1, 1)_{96}$ for the Wilson Bridge site. For the models with a pure nonseasonal autoregressive component the best-fit models were ARIMA $(4, 0, 0)(0, 1, 1)_{96}$ for the Telegraph Road site and ARIMA $(3, 0, 0)(0, 1, 1)_{96}$ for the Wilson Bridge site. In both cases the autoregressive order was one parameter higher for the Telegraph Road site. We believe this can be explained by the fact that the Wilson Bridge site is more heavily congested, subject to more extreme peaks, and more prone to instability and rapid fluctuation in traffic flow. Therefore, the best-fit models need to respond more quickly and be less influenced by past values for the Wilson Bridge site than for the Telegraph Road site.

The exponential smoothing results agree with this explanation. The best-fit exponential smoothing trend parameters, $\alpha$, for Telegraph Road and Wilson Bridge were 0.58 and 0.68, respectively. The higher the $\alpha$ value, the more weight is placed on recent values by the exponential smoothing forecast. These results indicate that in areas of heavier freeway congestion, best-fit models will have shorter memory than in less congested areas. Further research is needed to verify this hypothesis.

## Comparison with Other Models

We conclude that seasonal ARIMA models, with a nonseasonal component modeled as either an autoregressive or an ARMA process, and the Winters exponential smoothing models performed better on the Capital Beltway data set than did the nearest-neighbor, neural network, and historical average models in a previous study (*1*). In the study by Smith (*1*), the nearest-neighbor model is carried forward to multiple-interval (i.e., *k*-step ahead) forecasts. The comparison with Smith (*1*) begun by this study should be continued by using the ARIMA and exponential smoothing models to do multiple-interval forecasts and compare the results with Smith's multiple-interval nearest-neighbor model forecasts.

Our review of the literature indicates that ARIMA models would have performed better in previous studies if seasonal models had been used. In previous research, ARIMA models have been applied to nonstationary forms of traffic data series. We were able to find no use of seasonal models in the literature. We recommend that past comparative studies be revisited to include seasonal time series.

## ACKNOWLEDGMENT

The authors would like to thank Brian Smith for providing the data and forecasts from his research.

## REFERENCES

1. Smith, B. L. *Forecasting Freeway Traffic Flow for Intelligent Transportation Systems Applications.* Ph.D. dissertation. Department of Civil Engineering, University of Virginia, Charlottesville, 1995.
2. Brockwell, P. J., and R. A. Davis. *Introduction to Time Series and Forecasting.* Springer-Verlag New York, 1996.
3. Ahmed, M. S., and A. R. Cooke. Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques. In *Transportation Research Record 772,* TRB, National Research Council, Washington, D.C., 1979, pp.1–9.
4. Levin, M., and Y. Tsao. On Forecasting Freeway Occupancies and Volumes (Abridgment). In *Transportation Research Record 773,* TRB, National Research Council, Washington, D.C., 1980, pp.47–49.
5. Der Voort, M., M. Dougherty, and S. Watson. Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow. *Transportation Research C,* Vol. 4, No. 5, 1996, pp. 307–318.
6. *SPSS Trends Version 6.1,* 1996.
7. Clark, S. D., S. Watson, E. Redfern, and M. R. Tight. *Application of Outlier Detection and Missing Value Estimation Techniques to Various Forms of Traffic Count Data.* ITS Working Paper 384, Institute for Transport Studies, University of Leeds, Jan. 1993.
8. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering,* Vol. 82D, No. 1, 1960, pp. 35–45.
9. Ljung, G. M., and G. E. P. Box. On a Measure of Lack of Fit in Time Series Models. *Biometrika,* Vol. 65, 1978, pp. 297–303.
10. Chen, C., and L-M. Liu. Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association,* Vol. 88, 1993, pp. 284–297.
11. Chen, C., and L-M. Liu. Forecasting Time Series with Outliers. *Journal of Forecasting,* Vol. 12, 1993, pp. 13–35.